# METAGENOMIC AND METATRANSCRIPTOMIC ANALYSIS OF BACTERIAL, ARCHAEAL AND FUNGAL COMMUNITIES WITHIN THE HOT SPRINGS OF LAKE MAGADI IN KENYA

**ANNE KELLY KAMBURA**

**DOCTOR OF PHILOSOPHY**

**(Biotechnology)**

**JOMO KENYATTA UNIVERSITY OF**

**AGRICULTURE AND TECHNOLOGY**

**2016**

# Metagenomic and Metatranscriptomic analysis of Bacterial, Archaeal and Fungal communities within the hot springs of Lake Magadi in Kenya

**Anne Kelly Kambura**

**A Thesis Submitted in Fulfillment for the Degree of Doctor of Philosophy in Biotechnology in the Jomo Kenyatta University of Agriculture and Technology**

**2016**

## DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: ________________________ Date: _______________

**Anne Kelly Kambura**

This thesis has been submitted for examination with our approval as university supervisors

Signature: _______________________ Date: _______________

**Prof. Hamadi Iddi Boga, (PhD)**

**Taita Taveta University (TTU), Kenya**

Signature: ___________________________ Date: _______________

**Dr. Romano Mwirichia, (PhD)**

**University of Embu (UoEm), Kenya**

Signature: ____________________________ Date: _______________

**Dr. Remmy Kasili, (PhD)**

**JKUAT, Kenya**

# DEDICATION

This work is dedicated to my dear family; my husband Edward Nderitu, children; Shawn Karanja, Melissa Wangithi and Shayne Koome. I appreciate the support you have accorded me during the course of my studies. Without your encouragement and support this journey would be longer and tougher.

# ACKNOWLEDGEMENTS

I thank God Almighty for bringing me this far from the commencement of this long and challenging journey.

I wish to thank the following institutions for their assistance in this research. The National Commission for Science, Technology and Innovation (NACOSTI), for a research grant and permit; Higher Education Loans Board (HELB) for a partial scholarship to pursue this degree; Kenya Wildlife Service (KWS) for an access permit to the sampling area in Lake Magadi and Little Magadi; and Tata Chemicals Ltd for providing invaluable guides in this area.

I wish to acknowledge the Director, Institute for Biotechnology Research (IBR), JKUAT; Prof. Bernard Nyende for allowing me to enroll for this degree in the Institute as a staff, and for providing an enabling environment for laboratory and research work for this study.

This work would not have been accomplished without the wise counsel, professional acumen and research mentorship of Prof Hamadi Boga and Dr. Romano Mwirichia in microbial biotechnology. I truly appreciate my supervisors for their patience and for believing in me, even at the time when submission of papers for publication was very traumatic. Thank you for standing by me.

My most sincere gratitude goes to Dr. Remmy Kasili for walking the research walk with me, the distances we covered to collect samples in Lake Magadi and Little Magadi. His presence both in the field and JKUAT made this research possible and the challenges encountered livable.

In addition, no adjectives can describe the appreciation I have for Mr. Josephat Muthanga, for his intellectual ability and ensuring exact sampling techniques were achieved for this research. I am greatly indebted to him.

The latter part of this work would not have been accomplished without the meticulous encouragement from the Dean, Taita Taveta University, School of Agriculture, Earth and Environmental Sciences, Dr. Marianne Maghenda. She burnt the midnight oil with me to make sure that this work was accomplished in time. I appreciate her wisdom.

Particular recognition goes to my colleagues; Odilia Salano, Grace Mungai, Amanuel Menghs, Johnstone Neondo and Richard Wekesa in IBR, JKUAT; and friends (Rosaline Macharia - International Center for Insect Physiology and Ecology (ICIPE), Joyce Njuguna- International Livestock Research Institute (ILRI) and June Mwajuma – Technical University of Mombasa (TUM); Thank you for your assistance particularly in troubleshooting.

Special appreciation is directed to Mr. Nashon Adero for his expertise in Geographical Information Systems (GIS) that has made the maps in this work exemplary; and to my special friends, Mrs. Leah Andisi and Ms. Valentine Malemba who invested their invaluable time and moral patronage in Ngerenyi School.

I would like to register exceptional appreciation to my loving husband Edward Nderitu for his patience and personal sacrifices in providing field support and shouldering the cost of sample analysis in a foreign laboratory. I cannot thank him enough for being there for me.

MAY GOD BLESS YOU ALL!

## TABLE OF CONTENTS

**DECLARATION ........................................ ii**

**DEDICATION ........................................ iii**

**ACKNOWLEDGEMENTS ........................................ iv**

**TABLE OF CONTENTS ........................................ vi**

**LIST OF TABLES ........................................ xi**

**LIST OF FIGURES ........................................ xii**

**LIST OF APPENDICES ........................................ xvi**

**LIST OF ABBREVIATIONS AND ACRONYMS ........................................ xvii**

**ABSTRACT ........................................ xviii**

**CHAPTER ONE ........................................ 1**

**INTRODUCTION ........................................ 1**

1.2 General Introduction ........................................ 1

1.2 Distribution and features of soda lakes ........................................ 1

1.3 The formation of Soda lake environment ........................................ 4

1.4 Microbial Diversity and productivity ........................................ 7

1.5 Statement of the problem ........................................ 8

1.6 Justification ........................................ 8

1.7 Hypotheses ........................................ 9

1.8 Objectives ........................................ 9

1.8.1 General Objective....9
1.8.2 Specific Objectives....9

**CHAPTER TWO ....10**

**GENERAL LITERATURE REVIEW....10**

2.1 Microbial diversity of soda lakes ....10

2.2 Diversity of haloalkaliphiles in Lake Magadi....10

2.3. Molecular microbial ecology ....14

2.4 Restrictions of 16SrRNA approach to diversity studies ....14

2.5 Next Generation sequencing technology ....15

**CHAPTER THREE ....18**

**GENERAL MATERIALS AND METHODS....18**

3.1 Research authorization....18

3.2 Study site....18

3.3 Measurements of physicochemical parameters....19

**3.4 Sample Collection**....21

3.5 Water analysis of samples from the hot springs of Lake Magadi and Little Magadi 22

3.6 Total DNA extraction....23

3.7 Total RNA Extraction ....23

3.8 Synthesis of cDNA from 16S rRNA....24

3.9 Amplicon Library Preparation and Sequencing....25

3.10 Sequence analysis and taxonomic classification....25
3.11 Statistical analysis ....26

**CHAPTER FOUR ........................................................................................... 27**

**BACTERIA AND ARCHAEA DIVERSITY WITHIN THE HOT SPRINGS OF LAKE MAGADI AND LITTLE MAGADI IN KENYA ........................................... 27**

4.1 Introduction ........................................................................................... 27

4.2 Materials and methods ........................................................................................... 28

4.3 Results ........................................................................................... 29

4.3.1 Sampling ........................................................................................... 29

4.3.2 Composition and diversity of the microbial communities ........................................ 31

4.3.3 Total diversity taxonomic composition analysis based on 16S rDNA .................. 33

4.3.4 Taxonomic composition based on 16S rRNA derived from cDNA ...................... 34

4.3.5 Microbial richness and diversity indices ........................................................... 37

4.4 Discussion ........................................................................................... 45

4.5 Conclusion ........................................................................................... 47

**CHAPTER FIVE ........................................................................................... 50**

**DIVERSITY OF FUNGI IN SEDIMENTS AND WATER COLLECTED FROM THE HOT SPRINGS OF LAKE MAGADI ........................................................ 50**

5.1 Introduction ........................................................................................... 50

5.2 Materials and Methods ........................................................................................... 51

5.3 Results ........................................................................................... 52

5.4 Sequence data ........................................................................................... 53

5.5 Composition and diversity of fungal communities ................................................ 54

5.6 Fungal richness and diversity indices ........................................................... 60

5.7 Discussion ........................................................................................... 60

5.8 Conclusion ........................................................................................... 63

**CHAPTER SIX** ..................................................................................................**64**

**METABOLIC POTENTIAL OF THE PROKARYOTIC COMMUNITIES WITHIN THE HOT SPRINGS OF LITTLE MAGADI IN KENYA** .......................**64**

6.1 Introduction ..........................................................................................................64

6.2 Materials and methods ..........................................................................................65

6.2.1 Research Authorization .......................................................................................65

6.2.2 cDNA synthesis from mRNA, library construction and deep sequencing............65

6.2.3 Data analysis ........................................................................................................66

6.3 Results ..................................................................................................................67

6.3.1 Composition and diversity of microbial communities within the three sample types ..........................................................................................................................68

6.3.2 Bacterial taxonomic composition analysis ...........................................................70

6.3.3 Archaeal taxonomic composition .........................................................................73

6.3.4 Viral taxonomic composition ...............................................................................75

6.3.5 Fungal taxonomic composition ............................................................................78

6.3.6 Categories of Clusters of Orthologous Groups (COG) in the three transcriptomes ..........................................................................................................................78

6.3.7 Diversity of genes and pathways between samples ..............................................79

6.3.8 Gene Ontology ......................................................................................................83

6.3.9 Metacyc pathways .................................................................................................86

6.3.10 Functional classification of enzymes and their application in biotechnology .....89

6.4 Discussion ..............................................................................................................91

6.4.1 Active microbial diversity .....................................................................................91

6.4.2 Diversity of genes and pathways ..........................................................................93

6.4.3 Application of various enzymes obtained from the hot spring in biotechnology ..94

6.5 Conclusion .............................................................................................................95

**CHAPTER SEVEN** .......... **97**

**GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS** .......... **97**

7.1 General discussion .......... 97

7.2 Conclusions .......... 100

7.3 Recommendations .......... 101

**REFERENCES** .......... **102**

**APPENDICES** .......... **127**

## LIST OF TABLES

**Table 1.1:** Soda lakes in the world ........................................................................3

**Table 2.1:** Representatives of trophic groups previously identified within Lake Magadi ..........................................................................................................13

**Table 4.1:** Parameters of sampling sites analyzed in this study .....................................30

**Table 4.2:** Diversity indices computed on all OTU-based microbial taxonomic units within DNA and RNA datasets .....................................................................38

**Table 5.1:** Chemical analyses of samples from the Hot Springs Lake Magadi..............53

**Table 5.2:** Diversity indices computed on all OTU-based fungal taxonomic units obtained from samples collected from the hot springs of L. Magadi and Little Magadi.........................................................................................................60

**Table 6.1:** Top ten functional classifications of relatives of bacterial species within the three sample types collected from the hot springs of Little Magadi .............72

**Table 6.2:** Top ten functional classifications of relatives of archaeal species within the three sample types collected from the hot springs of Little Magadi .............74

**Table 6.4:** Absolute abundance of enzymes revealed within samples ............................91

## LIST OF FIGURES

**Figure 3.1:** showing a map of Lake Magadi and Little Magadi study area and locations of the sampled Hot Springs ........................................................ 20

**Figure 3.2:** Other observations at sampling sites. a, b and c shows presence of flamingos and fish at 45.1℃. d and e shows rocks that have been degraded by salt crystals ........................................................ 21

**Figure 3.3:** Samples collected from the three sampling sites. a, b and c represents wet sediments, microbial mats and water samples, respectively. ...................... 22

**Figure 4.1:** Venn diagrams showing the distribution of unique and shared OTUs within various sample types (B) in the three sampling sites (A). The number of OTUs in each hot spring is indicated in the respective circle. Site A, B and C represent hot springs 83.6 ℃, 81 ℃ and 45.1 ℃, respectively. .............. 32

**Figure 4.2:** Relative abundance of the most predominant phyla in various samples collected from the hot springs of L. Magadi. 'Others' represent all taxa that scored a relative abundance of below 1% across all samples in both data sets. ........................................................ 36

**Figure 4.3:** Relative abundance of the most abundant prokaryotic taxa at order level in samples collected from the hot springs of L. Magadi. 'Others' represents all taxa that scored a relative abundance of below 1% across all samples in both data sets. ........................................................ 37

**Figure 4.4:** Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities between microbial compositions of (A) 16S rDNA and (B) 16S rRNA cDNA datasets grouped according to sampling sites. ............... 40

**Figure 4.5:** Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities between microbial compositions of (A) 16S rDNA and (B) 16S rRNA cDNA datasets grouped according to sample types.................41

**Figure 4.6 a:** Hierarchical clustering of 16S rDNA samples collected from the three hot springs of L. Magadi and Little Magadi. ..................................................42

**Figure 4.6 b:** Hierarchical clustering of 16S rRNA cDNA samples collected from the three hot springs of L. Magadi and Little Magadi .....................................43

**Figure 4.7 a:** Comparative analysis (UPGMA similarity tree) of total microbial diversity of various sample types within hot springs of L. Magadi..........................44

**Figure 4.7 b:** Comparative analysis (UPGMA similarity tree) of active microbial diversity of various sample types within hot springs of L. Magadi and Little Magadi .........................................................................................................45

**Figure 5.1:** Venn diagram showing the distribution of unique and shared OTUs within various sample types in the three sampling sites. ......................................55

**Figure 5.2:** Relative abundance of the most predominant fungal species in various samples collected from the hot springs......................................................57

**Figure 5.3:** Hierarchical clustering of DNA samples collected from the three hot springs of L. Magadi and Little Magadi.................................................................59

**Figure 6.1:** Geographic map of 83.6 oC hot spring sampling site in Little Magadi .......67

**Figure 6.2:** Comparison of the most predominant prokaryotic taxa at phylum level within the 16S rDNA, 16S rRNA cDNA and mRNA cDNA....................69

**Figure 6.3:** Taxonomic classification of the most predominant prokaryotic groups at class level within the various samples collected from the hot springs of Little Magadi. 70

**Figure 6.4:** Taxonomic classification of relatives of the most predominant bacterial phyla in various samples collected from the hot springs of Little Magadi 71

**Figure 6.5:** Taxonomic classification of the most predominant relatives of archaeal phyla in various samples collected from the hot springs of Little Magadi. 73

**Figure 6.6:** Taxonomic classification of the most predominant viral groups in various samples collected from the hot springs of Little Magadi. 76

**Figure 6.7:** Sorting of obtained gene clusters into pathways. Genes present in the five regions of the heat map (I-V) were assigned to different KEGG pathways. 80

**Figure 6.8:** A heatmap showing gene clustering of obtained metabolic pathways according to their distribution within the samples. Mat samples were found to harbor most diverse genes 81

**Figure 6.9:** Sorting of sample-specific KEGG pathways Enzyme Commission (EC) between the samples. 83

**Figure 6.10:** Summary of Gene Ontology functional classification of the samples collected from the hot springs of Little Magadi. 84

**Figure 6.11:** Comparison of gene ontology (molecular function level 2) between the samples collected from the hot springs of Little Magadi. 85

**Figure 6.12:** Summary of metacyc pathways in the samples collected from the hot springs of Little Magadi. 86

**Figure 6.13:** Comparison of metacyc pathways between the samples collected from the hot springs of Little Magadi........................................................................88

**Figure 6.14:** Clustering of the six enzyme groups between the samples collected from the hot springs of Little Magadi..................................................................90

## LIST OF APPENDICES

**Appendix 1**: Total diversity environmental data sheet of the samples collected from the hot springs of L. Magadi and Little Magadi ........................................127

**Appendix 2:** Active diversity environmental data sheet of the samples collected from the hot springs of L. Magadi and Little Magadi ....................................128

**Appendix 3:** Absolute abundance of KEGG Pathway (KO) within the samples .........129

**Appendix 4:** Absolute abundance of Metacyclic Pathway within the samples ............130

**Appendix 5:** DNA Extraction Reagents.......................................................................132

**Appendix 6:** RNA Extraction Reagents.......................................................................133

## LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **ANOSIM** | Analysis of Similarity |
| **CDD** | Conserved Domain Database |
| **DNA** | Deoxyribonucleic Acid |
| **GPS** | Global Positioning System |
| **ITS** | Internal Transcribed Spacer |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **MEGAN** | Metagenome Analyzer |
| **NAST** | Nearest Alignment Space Termination |
| **NCBI** | National Centre for Biotechnology Information |
| **NMDS** | Non Metric Dimensional Scaling |
| **OTU's** | Operational Taxonomic Units |
| **PCR** | Polymerase Chain Reaction |
| **PyNAST** | Python Nearest Alignment Space Termination |
| **QIIME** | Quantitative Insights into Microbial Ecology |
| **RDA** | Redundancy analysis |
| **RNA** | Ribonucleic Acid |
| **SDS** | Sodium Dodecyl Sulphate |
| **SSU** | Small Sub Unit |
| **UCLUST** | Universal Clustering |

## ABSTRACT

The earth is a reservoir of genetic and metabolic diversity of microbial cells, mainly represented by uncultured groups. Previously, both culture-dependent and independent methods have been applied to study the diversity of microorganisms in extreme environments such as soda lakes. High throughput methods that involve direct isolation and analyses of nucleic acids from samples have been found more feasible in microbial ecology studies. Although earlier reports point to low microbial diversity, most recent studies indicate enormous diversity close to that found in soil. However, these studies have relied on the analysis of metagenomic DNA and may have a bias, since it incorporates both dead and live cells in the analysis. The main objective of this study was the application of metagenomic (DNA) and metatranscriptomic (RNA) analysis to study the diversity and function of active microbial communities within hot springs of Lake Magadi and Little Magadi in Kenya. Lake Magadi is hypersaline, alkaline lakes situated in the southern part of Kenyan Rift Valley. Solutes are supplied mainly by a series of alkaline hot springs with temperatures as high as 86 °C. Preceding culture independent studies were based on the analysis of 16S rDNA but were done on less saline lakes. In this initial study, illumina sequencing and analysis of amplicons of both total community rDNA and 16S rRNA cDNA; as well as mRNA transcript analysis were combined to determine the diversity and community structure of bacteria, archaea and fungal communities within 3 hot springs of Lake Magadi and Little Magadi.The samples collected included water, wet sediments and microbial mats from the hot springs in Lake Magadi at 45.1 ºC and from Little Magadi ***"Nasikie eng'ida"*** at 81 ºC and 83.6 ºC. Total community DNA and RNA were extracted from samples using phenol-chloroform and Trizol RNA extraction protocols respectively. The 16S rRNA gene variable region (V4 – V7) of the extracted DNA and RNA; and Fungal Internal Transcribed Spacer region, were amplified. Library construction was performed following Illumina sequencing protocol. Sequences were analyzed using QIIME pipeline. The R programming language and Vegan package were used to calculate Bray-Curtis dissimilarities between datasets, hierarchical clustering, Non Metric Dimensional Scaling (NMDS), redundancy analysis (RDA) and diversity indices. The Illumina HiSeq 2000 platform was used to analyze pool of metabolic genes in wet sediments, microbial mats and water samples collected from 83.6ºC hot spring without amplification. Metabolic genes data was analyzed at Computational Bioscience Research Center, King Abdullah University of Science and Technology on Dragon Metagenomic Analysis Platform (DMAP) using the Automatic Annotation of Microbial Genomes (AAMG) pipeline.Results showed that 3502 and 1913 Operational Taxonomic Units (OTU) were recovered from 16S rDNA and 16S

rRNA cDNA respectively. Uncultured diversity accounted for 89.35% 16S rDNA and 87.61% 16S rRNA cDNA reads. The most abundant phyla in both the 16S rDNA and 16S rRNA cDNA datasets included: *Proteobacteria* (8.33 – 50 %), *Firmicutes* 3.52 – 28.92 %, *Bacteroidetes* (3.45 – 26.44 %), *Actinobacteria* (0.98 - 28.57 %) and *Euryarchaeota* (3.55 - 34.48 %) in all samples. The NMDS analyses of taxonomic composition clustered the taxa into three groups according to sample types (i.e. wet sediments, mats and water samples). There was evident overlap of clusters between wet sediments and microbial mats from the three sample types in both DNA and cDNA datasets. The hot spring at 45.1 ºC contained less diverse populations compared to those in Little Magadi at 81-83 $^{o}$C. There were significant differences in microbial community structure at 95 % level of confidence, for both total diversity (P value, 0.009), based on 16S rDNA analysis and active microbial diversity (P value, 0.01) based on 16S rRNA cDNA analysis, within the three hot springs. Using the Fungal ITS dataset, a total of 334, 394 sequence reads were obtained within which, 151 OTUs were realized at 3% genetic distance. Taxonomic analysis revealed that 80.33 % of the OTUs belonged to the Phylum *Ascomycota*, 11.48 % *Basidiomycota* while the remaining consisted of *Chytridiomycota, Glomeromycota* and early diverging fungal lineages. The most abundant *Ascomycota* groups consisted of *Aspergillus* (18.75%)*, Stagonospora* and *Ramularia* (6.25% each) in wet sediment at 83.6 °C. *Penicillium* and *Trichocomaceae (*14.29% each) were dominant in wet sediment at 45.1 °C. The results revealed representatives of thermophilic and alkaliphilic fungi within the hot springs of Lake Magadi and Little Magadi suggesting their ability to adapt to high alkalinity, temperature and salinity. Microbial taxonomic classification of mRNA transcript comprised of cellular organisms and viruses across all samples. Differences in microbial composition and structure were observed as a function of sample type and temperature, with wet sediments harboring the highest diversity. More than 50% of microbial species, KEGG Orthologs, Gene Ontology and Enzymes were found to be unassigned within in the various functional classifications. Consequently, it can be concluded that the alkaline saline hot springs are inhabited diverse microorganisms that can be revealed by application of next generation sequencing technology in microbial ecological studies. Further research should focus on culturing the revealed uncultured groups of bacteria, archaea, fungi and viruses, and possible bioprospecting for their applicability in biotechnology industries.

# CHAPTER ONE

# INTRODUCTION

## 1.2 General Introduction

Extreme alkaline environments include naturally occurring soda lakes, soda deserts, soils and artificial industrial-derived waters. Soda lakes represent the most stable naturally occurring highly alkaline environments on earth. They commonly have pH values greater than 10 (occasionally reaching pH 12) and are characterized by the presence of large amounts of $Na_2CO_3$, (usually as $Na_2CO_3.10H_2O$ or $Na_2CO_3$, $NaHCO_3.2H_2O$) and are depleted of $Mg^{2+}$ and $Ca^{2+}$ because of the insolubility of carbonates under alkaline conditions (Duckworth, 1996; Grant, 2006). Environmental conditions such as pH, temperature and salinity are extremely high or low in these habitats (Horikoshi, 1991). They are often remote from the main centers of human activity and perhaps for this reason, they have been little studied. These environments are home to various groups of microorganisms adapted to these specific conditions. They include alkaliphiles, thermophiles, halophiles and acidophiles, depending on the type of extreme environment in which they inhabit (Horikoshi, 1991).

## 1.2 Distribution and features of soda lakes

Soda lakes are natural alkaline environments that are widely spread in the world (**Table 1.1**). Examples include Mono Lake and Summer Lake in North America (Raymon & Sistrom, 1967), Lake Van in Turkey (Kempe & Ka mierczak, 2002), Lake Malyi Kasytui, Siberia (Bryantseva *et al.*, 1999a) and Lakes; Magadi, Elmenteita, Sonachi, Nakuru and Bogoria in the East African Rift Valley (Baumgarte, 2003). The salinity of East African Rift Valley lakes varies from 5% in the Northern lakes (Bogoria, Nakuru, Sonachi, and Elmenteita) to approximately 30% (saturation) halite/trona saturation (Lake Magadi, Little Magadi and Lake Natron), with

approximately equal proportions of sodium carbonate and sodium chloride as the main salts (Grant, 1992). Lake Magadi is one of the smallest and most saline lakes in the East Rift Valley (Warren, 2006). These lakes are found in hydrologically closed basins, either due to being crater lakes, or due to their position in a natural depression (Warren, 2006). The pH ranges from 9 to 11.5 in the more alkaline depressions (Grant, 1992; Duckworth *et al.*, 1996). Detailed limnological and microbiological studies have been carried out in East African Rift Valley lakes since the early 1930s (Grant *et al.*, 1990; Jones *et al*., 1994; Duckworth *et al.*, 1996; Jones *et al*., 1998; Tindall, 1988; Baumgarte, 2003).

**Table 1.1: Soda lakes in the world**

| Country | Lakes |
|---|---|
| Canada | Lake Manito |
| United States | Albert Lake, Lake Lenore, Soap Lake, Big Soda Lake, Owens Lake, Mono Lake, Searles Lake, Marsh, Harney Lake, Summer Lake, Pyramid Lake, Walker Lake |
| Mexico | Lake Texcoco |
| Hungary | Lake Feher |
| Russia | Tanatar Lakes |
| Turkey | Lake Van |
| India | Lake Loner, Lake Sambhar |
| China | Lake Chahannor; Lake Zabuye; Bange Lake; Lake Baer; Lake Wudunao; Lake Hamatai |
| Libya | Lake Fezzan |
| Egypt | Wadi Natrun |
| Ethiopia | Lake Aranguadi, Lake Kilotes, Lake Abiata, Lake Shala, Lake Chilu, Lake Hertale, Lake Metahara |
| Sudan | Dariba lakes |
| Kenya | Lake Bogoria, Lake Nakuru, Lake Elmenteita, Lake Magadi, Lake Simbi, Lake Sonachi |
| Tanzania | Lake Natron, Lake Embagi, Lake Magad, Lake Manyara, Lake Balangida, Bosotu Crater, Lakes, Lake Kusare, Lake Tulusia, El Kekhooito, Momela Lakes, Lake Lekandiro, Lake Reshitani, Lake Lgarya, Lake Ndutu, Lake Rukwa North |
| Uganda | Lake Katwe, Lake Mahega, Lake Kikorongo, Lake Nyamunuka, Lake Munyanyange, Lake Murumuli, Lake Nunyampaka |
| Chad | Lake Bodu, Lake Rombou, Lake Dijikare, Lake Momboio, Lake Yoan |
| Australia | Lake Corangamite, Red Rock Lake, Lake Werowrap, Lake Chidnup |

Adapted from: Grant and Sorokin, 2011

## 1.3 The formation of Soda lake environment

Soda lakes are characterized by large amounts of sodium carbonate, or carbonate complexes, shaped by evaporative concentration (Grant, 1992; Jones & Grant, 2000). Factors leading to formation of alkaline saline deposits include geographical, geological and climatic conditions (Tindall, 1988). Climatic influence controls the amount of water entering the system as rainfall or surface runoff and the amount leaving by evaporation, hence favoring saline lake formation. Geochemical impacts determine the ions flowing into the system. The carbon dioxide solution forms carbonic acid, which undergoes ion exchange with the surrounding rock thus leaching the minerals. The geochemistry of the region specifically influences the ionic structure of groundwater (Tindall, 1988; Baumgarte, 2003). However, this may be modified further by precipitation of insoluble salts or replenishment of lost carbonate/bicarbonate by additional solution of carbon dioxide (Grant *et al.*, 1990; Behr & Röhricht, 2000; Baumgarte, 2003).

The most important contributing factor in alkaline environment is lack of alkaline earth cations ($Mg^{2+}$ and $Ca^{2+}$) in the surrounding landscape, meaning that there's an absence of rocks of sedimentary origin. Concentration of the ions in groundwater through evaporation leads to a shift in the carbon dioxide/bicarbonate/carbonate equilibrium in favor of carbonate ions. This result in precipitation of insoluble carbonates of calcium and magnesium, removing them from solution and allowing the more soluble carbonates of sodium and potassium to build up (Baumgarte, 2003). The topography allows concentration of these salts in a shallow depression forming a closed drainage basin with a high marginal relief, having adequate rainfall to sustain streams entering the basin to produce a standing body of water (Grant *et al.*, 1990). Generally, there's little or no out flow and water loss is by evaporation. In arid zones with high rates of evaporation exceeding inflow, salts accumulate (Jones *et al.*, 1994; Baumgarte, 2003).

The rift valley lakes are sustained by hot springs that are less concentrated and alkaline. However, gradients of temperature, salinity and pH occur where there is blending, in this way making a scope of diverse soda lake habitats (Duckworth *et al.*, 1996). The chain of lakes occupying the Rift valley floor represents remnants of lakes of the pluvial period that occurred 10,000 years back (Zavarzin *et al.*, 1999). They are situated in a region with an extremely large geothermal gradient and associated with underground aqueous steam vents, for instance in Lake Bogoria and thermal outlets found in Lake Magadi. The vicinity of hot springs in L. Magadi may also realize some dispersion of the salt waters. However, there's no proof that this adds more solutes to the brackish waters or that different wellsprings of sodium carbonate, for example, the fountain of liquid magma Ol Doiny Lengai, which lies south of Lake Magadi, add to the alkalinity and saltiness of this lake, or the neighboring Lake Natron (Baker, 1958; Eugster, 1970; Jones *et al.*, 1977).

**Figure 1.1**: Map of Rift valley, Kenya, showing position of the lakes in relation to main topographical features of the rift. Lake Hannington is now called Lake Bogoria

(Brown, 1973; Baumgarte, 2003).

## 1.4 Microbial Diversity and productivity

Despite the extreme nature of soda lake environments, they are characterized by high productivity probably as a result of the high ambient temperatures, high light intensities, accessibility of phosphates and unlimited access to carbon dioxide (Melack & Kilham, 1974; Grant *et al.*, 1990). They have been found to support a dense and different groups of heterotrophs, organotrophs, halophiles, alkaliphiles, and alkalitolerant representatives of most important bacterial and Archaeal phyla (Duckworth *et al.*, 1996; Jones *et al.*, 1998; Grant *et al.*, 1999; Zavarzin *et al.*, 1999; Mwirichia, 2009). Among these microbial groups, there exists cycling of carbon, sulfur, and nitrogen under aerobic and anaerobic conditions within the lakes.

A marked difference in microbial communities has been observed between the hypersaline, alkaline brines of Lakes Magadi and Natron (Rift Valley), Owens Lake (California) and some of the Wadi Natrun depression (Egypt) with salt concentration approaching saturation (30 % or greater) and the more dilute waters of lakes Elmenteita, Nakuru and Bogoria with about 5 % (w/v) salinity (Jones *et al.*, 1994; Jones *et al.*, 1998). The less alkaline lakes are usually dominated by dense blooms of cyanobacteria while the hypersaline lakes, on occasion (i.e. after extensive rainfall causes dilution of the brines) support blooms of both cyanobacteria and alkaliphilic anoxygenic phototrophs (Jones *et al.*, 1998; Grant *et al.*, 1999).

Although the soda lakes of the Rift Valley are eutrophic, moderately shallow and probably oxygen limited, they maintain dense populations of non-phototrophic aerobic organotrophic microbes that utilize products of photosynthesis as well as products of anaerobic degradation. Viable counts of aerobic organotrophs from diluted lakes range between $10^5$ and $10^6$ cfu ml $^{-1}$ (Grant *et al.*, 1990). Therefore these diverse groups play different roles in the biogeochemical cycles.

## 1.5 Statement of the problem

Kenyan soda lakes are an example of relatively simple but highly productive ecosystems that can serve as models for studying microbial diversity under extreme conditions. The soda lakes are also characterized by hot springs which host both hyperthermophillic and haloalkalithermophilic microorganisms. These represent the organisms at the upper-temperature limits of life (Stetter & Zillig, 1985; Brock, 1986; Stetter, 1996). They grow optimally at temperatures between 80 and 105 °C and are unable to grow below 60 °C. Their adaptation to high pH and elevated temperatures draws consideration not just as a source of industrial enzymes, but also for studying adaptive mechanisms of organisms to extreme environmental parameters. However due to the unculturability of most microbes, very few organisms have been isolated from these extreme environments. Culture independent studies done so far have focused on the diversity and not on the function. Therefore, a better understanding of the structure, functions, and interactions of the different microbial communities is required. This can be done using novel high throughput techniques that explore both the diversity and function of the total microbial community. In this study a combination of metagenomic and metatranscriptomic analysis was used to study the diversity and function of microbial communities within the hot springs of Lake Magadi and Little Magadi in Kenya.

## 1.6 Justification

The use of an integrated approach to understand microbial ecology and function is vital for explaining the biochemical and ecological roles of uncultured microorganisms and their interaction with biotic and abiotic factors. Culture dependent methods are valuable in the detailed investigation of pure cultures to develop a better understanding of microbial physiology and genetics. Culture dependent approach cannot be used for objective analysis of populations within natural communities because, less than 1% of the observed diversity can be cultured in the laboratory. It is therefore doubtful that such would help in the accurate description of microorganisms as they occur within natural

habitats. Through the application of new sequencing technologies, it is possible to observe the occurrence patterns of individual populations upon which the structure of microbial communities is based. Earlier culture independent studies on Rift valley soda lakes, based on the analysis of 16S rDNA were conducted on less saline (5%), alkaline lakes with lower temperatures. Moreover, such information on Little Magadi has not been documented. This study involved application of metagenomic (DNA) and metatranscriptomic (RNA) analysis to determine the diversity and function of microbial communities within the hot springs of Lake Magadi.

## 1.7 Hypotheses

1. The hot springs of Lake Magadi are rich in microbial diversity.
2. The hot springs of Lake Magadi harbors active microbial diversity.
3. The hot springs of Lake Magadi are rich in diverse metabolic genes.

## 1.8 Objectives

### 1.8.1 General Objective

To study the functional microbial diversity within the hot springs of Lake Magadi and Little Magadi, using metagenomic and metatranscriptomic analysis.

### 1.8.2 Specific Objectives

1. To assess the total microbial diversity in the hot springs of Lake Magadi
2. To assess the active microbial community in the hot springs using Illumina sequencing of cDNA libraries generated from rRNA.
3. To assess the diversity of metabolic genes expressed in the hot springs using Illumina sequencing of cDNA libraries generated from mRNA.

# CHAPTER TWO

# GENERAL LITERATURE REVIEW

## 2.1 Microbial diversity of soda lakes

The East African Rift valley soda lakes have been shown to support a dense and diverse population of autotrophic, heterotrophic, organotrophic, halophilic, alkaliphilic and alkalitolerant representatives of major bacterial, archaeal and fungal phyla (Duckworth *et al.*, 1996; Jones *et al.*, 1998; Grant *et al.*, 1999; Zavarzin *et al.*, 1999). The lakes are dominated by anaerobic breakdown of organic matter, with relatively limited obligate alkaliphilic anaerobes, mostly of the clostridium group having been isolated (Grant *et al.*, 1999).

## 2.2 Diversity of haloalkaliphiles in Lake Magadi

Despite the elevated saline conditions coupled with high pH and low ammonium ion concentrations in the soda lake environments (McGenity & Oren, 2012), studies have demonstrated that this niche is home to a specialized groups of organisms. Although species diversity is often lower in extreme environments (Kristjánsson and Hreggvidsson, 1995) such as Lake Magadi (Jones *et al.,* 1994), the degree of specialization of microbes in this carbonate rich environment is thought to have arisen from long and complex evolutionary processes (Zavarzin *et al.,* 1999; Bell, 2012). It is hypothesized that the relationship between salt tolerance and efficient energy metabolism in organisms as the partial reason for the decline in community composition (Oren 1999). The study of total mineralization; rather than high alkalinity is the main limiting factor of the whole trophic system present in soda lakes (Zhilina *et al*., 2012).

The hypersaline brines of Lake Magadi harbor a unique population of prokaryotes as the dominant microbial biomass. Many of these microorganisms have been reported from previous studies as being obligately alkaliphilic or alkalitolerant (Grant & Sorokin,

2011) which represent separate alkaliphilic lineages within pre-established taxa (Zavarzin *et al.,* 1999). These microbes are extremely prolific due to the high ambient temperatures, high light intensities, availability of phosphates and an unlimited supply of carbon dioxide from carbonates in solution (Melack & Kilham, 1974; Grant & Tindall, 1986).

Seasonally, Lake Magadi exhibits the pinkish-red coloration of the haloalkaliphilic archaea due to the synthesis of $C_{50}$ carotenoids by these microorganisms. In addition the colorful blooms (green, pink or red) of cyanobacteria and alkaliphilic anoxygenic phototrophs are also visible periodically due to dilution of the brines (Rodriguez-Valera *et al.,* 1981; Mwatha & Grant, 1993). Gram-negative bacteria have been reported to be the major populations in this soda lake environment. The variabilities in conductivity, alkalinity, phosphate and nitrogen levels determine species dominance during a particular season (Jones *et al*., 1994).

Studies on microbial species from Lake Magadi have been an interesting area of research and many haloalkaliphilic strains have been retrieved using isolation media enriched with appropriate concentrations of sodium carbonate and sodium chloride (Jones & Grant, 1999; Baumgarte, 2003; Kambura *et al*., 2012). Biochemically reactive microorganisms that are able to produce various extracellular hydrolytic enzymes such as lipases, proteinases and cellulases have been isolated from Lake Magadi (Kambura *et al*., 2012). Some of the groups obtained through culture dependent techniques have been assigned to existing taxonomic groups while most fall into new microbial assemblages. The 16S rRNA and 18SrRNA gene based studies of the organisms within soda lakes have revealed novel taxa yet to be cultivated and characterized (Salano, 2011; Kambura *et al*., 2012).

The presence of representatives of all the main trophic groups involved in the active nutrient cycling (carbon, sulfur and nitrogen) under both aerobic and anaerobic conditions in Lake Magadi, has been reported (Grant and Sorokin, 2011). The basis of primary production within Lake Magadi is however unclear, since the trona beds are

often dominated by organotrophic archaea (Grant, 2006). Grant and Tindall (1986) suggest that primary productivity is probably driven by the genus *Halorhodospira*, a typical soda lake component. A representative of this genus the sulfur oxidizing *Halorhodospira halophila* has been isolated from Lake Magadi. The obligately halophilic and alkaliphilic archaea present at densities of $10^6$ -$10^7$ $ml^{-1}$ are presumed to be vital for secondary productivity. The representatives of major trophic groups identified under aerobic and anaerobic conditions within Lake Magadi are summarized on **Table 2.1**.

**Table 2.1: Representatives of trophic groups previously identified within Lake Magadi**

| Trophic level | Examples of aerobes |
|---|---|
| Phototrophs | *Arthrospira platensis* (Grant, 2006) *Oscillatoria limnetica* and *Synechocytis salina* (Dubinin *et al.,*1995) |
| Eukaryotic microalgae | *Chlorella minutissima* (Gerasimenko *et al.,* 1999) |
| Sulfur-oxidizers | *Halomonas magadiensis* (Duckworth *et al.,* 2000) and *Thioalkalivibrio nitratis* (Sorokin *et al.,* 2011) |
| Haloalkaliphilic archaea | *Natrialba magadii; Natronobacterium pharaonis* and *Natronobacterium gregoryi* (Tindall *et al.,* 1984) |
| | **Examples of anaerobic alkaliphiles** |
| Anoxygenic phototrophs | *Tindallia magadiensis* (Kevbrin *et al.,* 1998) |
| Sulphate-reducing bacterium | *Halorhodospira halophila* (Grant and Tindall, 1986) |
| Saccharolytic anaerobes | *Desulfonatronovibrio hydrogenovorans* (Zhilina *et al.,* 1997 b) |
| Saccharolytic spirochetes | *Halonatronum saccharophilum* (Zhilina *et al.,* 2001a) *Amphibacillus fermentum* and *Amphibacillus tropicus* (Zhilina *et al.,* 2001b) |
| Acetogenic organotrophs | *Spirochaeta alkalica* and *Spirochaeta africana* (Zhilina *et al.,* 1996a) |
| Methanotrophic methanogen | *Natroniella acetigena* (Zhilina *et al.,* 1996b) *Methanohalophilus zhilinae* strain Z7936 (Zhilina and Zavarzin, 1994; Kevbrin *et al.,* 1997) |

**Adapted from:** Mwatha and Grant, 1993; Jones and Grant, 1999; Zavarzin *et al.,* 1999.

## 2.3. Molecular microbial ecology

Molecular methods provide tools for analyzing the entire microbial community, including those that have not been cultured. Such methods became essential in microbial ecology in the early 90s (Pickup, 1991; Stackebrandt *et al.*, 1993; Amann *et al.*, 1995; Holben & Harris, 1995). DNA analysis has been applied to analyze whole communities, individual isolates, and clones of particular genes. Low resolution and broad scale analysis of community DNA like DNA reassociation, allow assessment of total diversity of microbial communities (Torsvik *et al.*, 1996).

The use of 16S rRNA gene sequences has been the most common genetic marker in studying the taxonomy and phylogeny of prokaryotes. This could be attributed to its occurrence in almost all bacteria and archaea; commonly existing as a multigene family or operons (Patel *et al.*, 2000). Previously, several studies to describe new species are based upon small subunit (SSU) sequences or other polyphasic information. As indicated by Stackebrandt and Goebel, 1994, the emergence of SSU sequence technology and its potential usefulness was summarized in the definition of a species. Although 97% similarity of 16S rRNA gene sequence data on an individual strain, with a nearest neighbor has been shown to represent a novel species, its significance is not clear, since it can demonstrate grouping inside of a formerly characterized taxon (Petti, 2007).

## 2.4 Restrictions of 16SrRNA approach to diversity studies

The earth is a pool of genetic and metabolic diversity of prokaryotic cells, mainly represented by uncultured groups (Sleator *et al.*, 2008). This reservoir has been studied through isolation, and identification of microorganisms as well as sequencing of cloned libraries (Grant *et al.*, 1990; Duckworth *et al.*, 1996; Cardenas & Tiedje, 2008; Lorrayne *et al.*, 2012). Microbial conventional cultivation methods are restricted to the minority of culturable species (Chistoserdova, 2010) and therefore most of the microbial diversity is mainly represented by uncultured groups. The cultured diversity

has helped shape understanding of physiology and metabolic functions of diverse groups of organisms. However, this is laborious, time consuming, selective and biased for growth of specific microorganisms.

Libraries of PCR - amplified 16S rRNA genes may not represent an accurate picture of prokaryotic diversity within a given community. Some of the molecular fingerprinting approaches (RFLP, ARDRA) are too sensitive; giving too high resolution gives dependable genotypic description at community level. Total DNA from complex microbial communities contain too much information to be analyzed directly by high resolution methods (Torsvik *et al.*, 1996). The species diversity is so great that sequencing may be biased because of unequal amplification of species' 16S rRNA genes.

Riesenfeld *et al*. (2004), reported that regardless of the fact that all presently accessible sequences were joined, they would not constitute a complete statistics of the considerable number of 16S rRNA qualities on earth. There may be predispositions of the impacts of different prokaryotic groups to libraries. The efficiencies of nucleic acid extraction may also vary for different bacteria and the number of copies of 16S rRNA genes. There may be preferential amplification of some sequence types in respect to others (Frostegard *et al.*, 1993; Embley & Stackebrandt, 1997; Von Wintzingerode *et al.*, 1997) and some sequences may arise from contaminating DNA, hence may not represent the genuine microorganisms present in the sample being concentrated on (Tanner *et al.*, 1998).

## 2.5 Next Generation sequencing technology

Molecular based methods involving direct isolation and analyses of nucleic acids from samples have been revealed to help overcome some of the biases experienced in culture dependent studies. These include metagenomics and metatranscriptomics (Hugenholtz, 2002; Handelsman, 2004; Xu, 2006; Warnecke & Hess, 2009;

Chistoserdova, 2010), and they assist in exploration of mixed microbial communities existing in various natural environments (Shi *et al.*, 2009; Gifford *et al.*, 2010). Their approaches involve sequencing of random DNA or cDNA profiles, determining taxonomic diversity and prospective genes related to environmental responses (Handelsman, 2004).

Metagenomics enables discovery of interactions between microorganisms and the environment, and assignment of ecosystem functions to various communities (Hugenholtz, 2002; Handelsman, 2004; Lopez-Garcia & Moreira, 2008). Functional genes of uncultured organisms can be linked to phylogenetic groups by cloning and sequencing of large genomic DNA fragments (Sjöling & Cowan, 2008; Carola & Rolf, 2009). This enables assessment of dominant biosynthetic pathways and primary energy sources (Biddle *et al.*, 2008; Frias-Lopez *et al.*, 2008; Carola & Rolf, 2009). A research on microbial assemblages from surface water at the Hawaiian Ocean Time-Series revealed community-wide metabolic activities and day-night patterns of differential gene expression (Poretsky *et al.*, 2009). The transcript pools composition was found to be consistent with various models of prokaryotic gene expression (Poretsky *et al.*, 2009; Lorrayne *et al.*, 2012).

Metatranscriptomics allow monitoring of microbial gene expression profiles in natural environments by studying global transcription of genes by random sequencing of mRNA transcripts pooled from microbial communities at a particular time and place (Moran, 2009). RNA extracted from environmental samples provides more valuable information than DNA in revealing active microbial communities versus dormant microbial communities (Torsvik & Øvreås, 2002). Several genes, e.g., ammonia oxidation, nitrogen fixation, denitrification and sulfate reduction, have been amplified from DNA/RNA isolated from microbial communities to obtain insights into key microbial processes (Hansel *et al.*, 2008). These techniques have been applied to water samples from Mono Lake, an alkaline, hyper saline lake in California (Poretsky *et al.*, 2005), a salt marsh off the coast of Georgia (Linhoss *et al.,* 2011) and the North Pacific

Ocean (Poretsky *et al.*, 2005). Combining metatranscriptomic approaches with new sequencing methods has been demonstrated as a powerful approach in the study of marine microbes (Frias-Lopez *et al.*, 2008; Segata *et al.*, 2013).

These novel methods have been invigorated by the introduction of next generation sequencing technologies whereby more data can be practically generated in a reasonably short time and in cost-effective way (Elahi & Ronaghi, 2004; Kozarewa *et al.*, 2009; Creer *et al.*, 2010). They allow direct sequencing of DNA or cDNA, hence avoid possible cloning bias leading to large-scale studies (Adams *et al.*, 2009). Advances in throughput and cost-reduction of sequencing technologies have also increased the number and size of metagenomic sequencing projects. The data obtained helps in the exploration of biodiversity and performance of various organisms in diverse ecosystems (Carola & Rolf, 2009).

# CHAPTER THREE

# GENERAL MATERIALS AND METHODS

## 3.1 Research authorization

Research authorization was obtained from the National Commission for Science, Technology and Innovation (NACOSTI) on 30$^{th}$ August 2013 in Kenya, (Research Permit Number NCST/RCD/12B/103/61) and permission to conduct research in Lake Magadi (Reference Number KWS/BRM/5001) was obtained from the Kenya Wildlife Service (KWS) on 24$^{th}$ September 2013.

## 3.2 Study site

Lake Magadi is a hypersaline lake that lies in the southern part of the Kenyan Rift Valley close to the Tanzanian border, between Lake Natron in the south and fresh water Lake Naivasha to the north. It is approximately 2° S and 36° E of the Equator at an elevation of about 600 m above sea level, and lies in the lowest part of the trough in a naturally formed closed lake basin. The lake covers an area of 90 km$^2$ and is one of the smaller Rift Valley lakes (Behr & Röhricht, 2000).

Evaporation is intense during the dry season (3500 mm per annum), the range of temperature being between 22 °C and 34 °C. The Loita hills and the Mau escarpment to the west shield the valley floor from rainfall resulting in an annual total of approximately 500 mm of rainfall in the two rainy seasons (Behr & Röhricht, 2000). The lake water lies below the surface and surface water is usually only found around the edges of the crystalline deposits where thermal springs feed the lake (Tindall, 1980).
There are no permanent rivers entering Lake Magadi basin and solutes are supplied mainly by a series of alkaline springs with temperatures as high as 86 °C. The springs are located around the perimeter of the lake. Where the salinity is low enough and the temperatures are not too high, these peripheral lagoons support a thriving colony of fish

*Tilapia grahami*, which can tolerate a pH of 10.5 and temperature of 39 °C. In some saline lagoons, probably in the absence of Tilapia, mass accumulations of microorganisms may be observed (Tindall, 1988). The crystalline trona deposits of the lake itself are colored off-white, red/orange, or red/purple. Closer examination of the surface trona deposits show that under appropriate conditions, a visible microbial stratification occurs which resembles stromatolitic formations found in other benthic saline environments (Tindall, 1980). The lower layer of the lake is a region of degrading organic matter rendered black by sulfate reduction (Behr & Röhricht, 2000).

In summary, in terms of water chemistry and mineralogy, Lake Magadi is an example of a typical alkaline saline lake at the stage of maximum evaporite productivity. It is located in the rain shadow of mountains with a large catchment area. Samples analyzed in this study were collected from 3 hot springs: one hot spring within the main Lake Magadi (02º 00 3.7 S 36º 14 32 E) at 45.1 ºC and pH 9.8; and two hot springs within Little Magadi "***Nasikie eng'ida***" (01º 43 28 S 36º 16 21 E), and (01º 43 56 S 36º 17 11 E) at elevations of 611 m and 616 m, temperatures of 81 ºC and 83.6 ºC and pH range of 9.2 and 9.4, respectively.

### 3.3 Measurements of physicochemical parameters

A Garmin eTrex 20, a hand-held GPS receiver with multi-constellation ability assuring an absolute positioning accuracy of 3 meters, was used to sample the points. The GPS was set up to use the local datum approved by the Survey of Kenya, which is Arc 1960. To accurately show the sampled hot springs to scale, maps were constructed in ArcGIS 10.3 using the GPS coordinates captured from the lake during fieldwork. The layers for towns, rivers, lakes and roads were added from ArcGIS Online database to enrich the thematic maps, as shown in **Figure 3.1A** and **B.** The right cartographic standards were adopted to ensure adequate visualization and scale for the two maps.

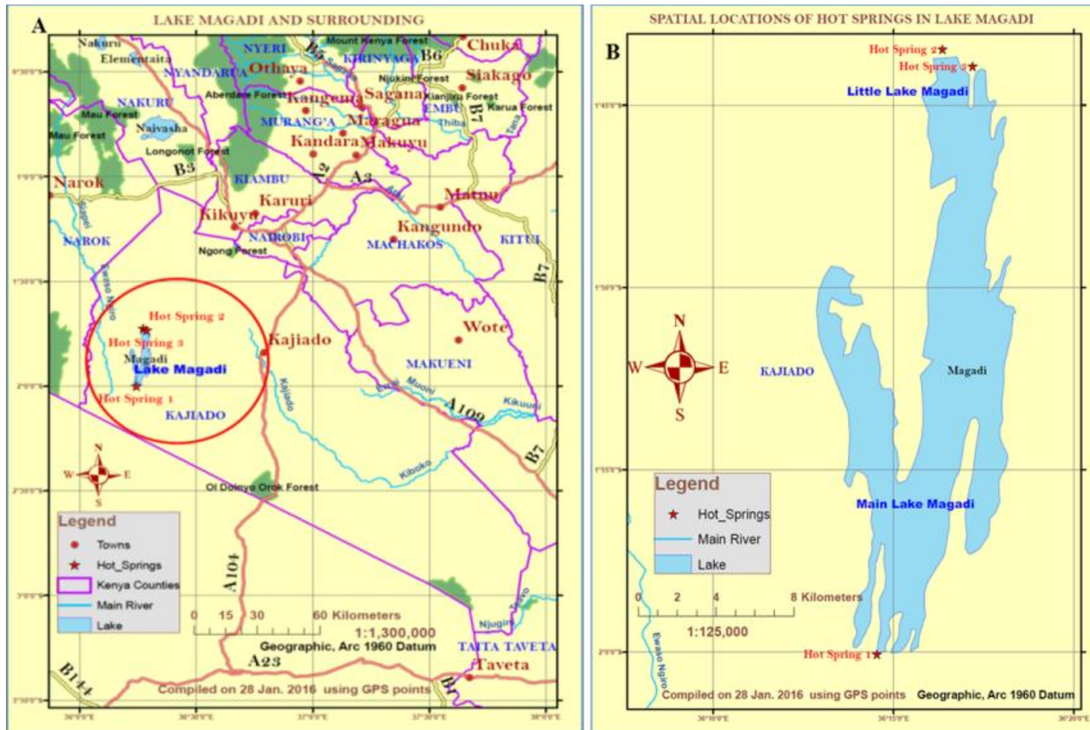


**Figure 3.1 showing a map of Lake Magadi and Little Magadi study area and locations of the sampled Hot Springs**

The pH for each sampling point was measured with a portable pH-meter (Oakton pH 110, Eutech Instruments Pty. Ltd) and confirmed with indicator strips (Merck, range 5–10). Temperature, Electrical Conductivity (EC), Total Dissolved Solids (TDS) and dissolved oxygen (DO) were measured on site using Electrical Chemical Analyzer (Jenway - 3405) during sampling. *In situ* temperature was recorded once for each study site and assigned to all the three sample types for that site. Other physical features at sampling site included presence of salt lake Tilapia fish in large numbers, heavy presence of algae and flamingos as well as rocks that had been degraded by salt crystals (**Figure 3.2).**

**Figure 3.2: Other observations at sampling sites. a, b and c shows presence of flamingos and fish at 45.1ºC. d and e shows rocks that have been degraded by salt crystals**

### 3.4 Sample Collection

All samples were collected randomly in triplicates from each hot spring. Water samples were collected using sterile 500 ml plastic containers that had been cleaned with 20 % sodium hypochlorite and UV-sterilized for one hour. Wet sediments were collected by scooping with sterilized hand shovel into sterile 50 ml Falcon tubes. The upper 5 mm from each microbial mat developing on the hot spring water margins was collected into sterile 500 ml plastic jam jars (**Figure 3.3**). All samples were transported on dry ice to the laboratory at Jomo Kenyatta University of Agriculture and Technology. Water for

nucleic acid extraction (500 ml) was filtered through a 0.22 µM filter papers (Whatman) and all filter papers containing samples were stored at -80 ℃. Pellets were obtained from water samples by re-suspending the filter papers in phosphate buffer solution, and centrifuging 5ml of the suspension at 13000 rounds per minute for 10 minutes. These were used for nucleic acid extraction.

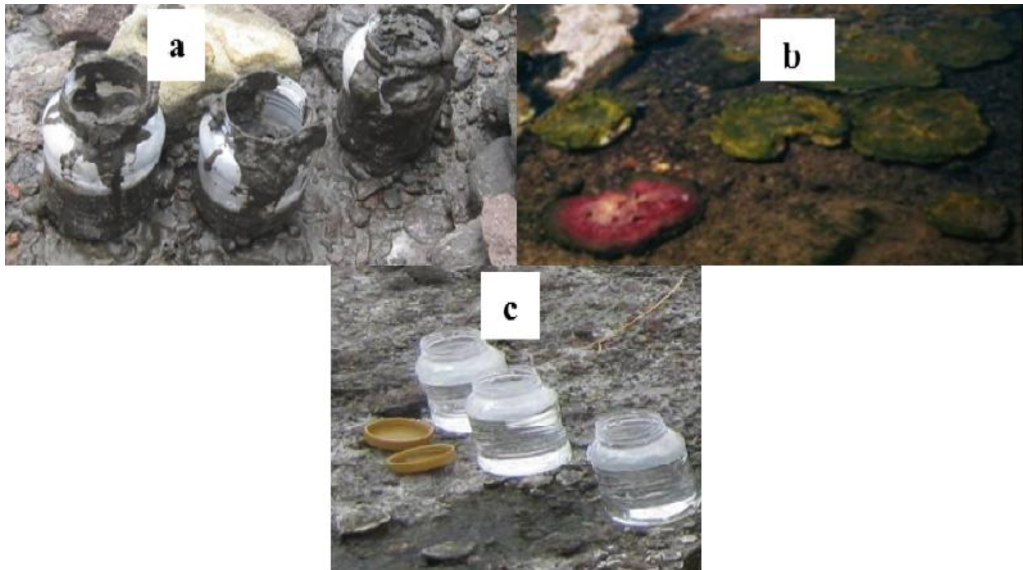


**Figure 3.3: Samples collected from the three sampling sites. a, b and c represents wet sediments, microbial mats and water samples, respectively.**

### 3.5 Water analysis of samples from the hot springs of Lake Magadi and Little Magadi

In order to identify and quantify the chemical components and properties of the hot spring waters, the samples were subjected to water chemistry analyses for pH, major cations, anions and trace elements. Water analysis of samples from Lake Magadi and Little Magadi was carried out at Crop Nutrition Laboratory Services (CNLS), Coopers, Nairobi. Cations such as Ca, Mg, K, Na, Mn, Fe, Cu, Mo, B, Zn, S were analyzed using Atomic Absorption Spectrometry (AAS) while anion analysis was carried out using Mass spectrometry.

### 3.6 Total DNA extraction

The sediments, microbial mats and the pellets were obtained from water samples (described in section 3.4 above) were thawn from -80 ºC. 0.2 g of sediment samples and 0.4 g of microbial mat samples were weighed into separate sterile eppendorf tubes. DNA extraction was carried out as described by Sambrook et al., (1989). 500 µl of solution A (50 mM Tris pH 8.5, 50 mM EDTA pH 8.0 and 25 % sucrose solution) was added to each tube and mixed by gently inverting several times and centrifuged at 13000 revolutions per minute for one minute to remove salts and exopolysaccharides from sediment samples. The supernatant was discarded and the sample re-suspended in 200 µl of solution A. To this were added 5 µl of Lysozyme (20 mg/ml) and 5 µl of RNAse A (20 mg/ml), gently mixed and incubated at 37 $^{o}$C for one hour. Following incubation, 600 µl of solution B (10 mM Tris pH 8.5, 5 mM EDTA pH 8.0 and 1 % SDS) was added and contents were mixed by inverting the eppendorf several times. 10 µl of Proteinase K (20 mg/ml) was added, mixed gently and incubation at 50 $^{o}$C for 1 hour.

DNA was extracted by adding equal volumes of phenol: chloroform, mixed by inverting several times and centrifuged for 15 minutes at 13000 revolutions per minute. This step was repeated once and a second extraction was done with equal volume of Chloroform: Isoamyl alcohol (24:1) mixed by inverting several times and then centrifuged for 15 minutes at 13000 revolutions per minute. The aqueous phase was carefully transferred into a new tube. This step was repeated once to remove all the phenol from the sample and DNA precipitation was done overnight at -80 ºC, by adding an equal volume of isopropanol and 0.1 volumes of 3M NaCl. DNA pellets were washed twice by use of 70% Ethanol, air dried at room temperatures for 20 minutes and stored at -20 ºC.

### 3.7 Total RNA Extraction

Total RNA was extracted from 0.25 g of sediment and mat samples, and pellets obtained from the water samples (described in section 3.4 above), in triplicates using Trizol RNA extraction protocol (Chomczynski & Sacchi, 1987). Briefly, 750 µl of Trizol LS and 250

µl of each sample were added to a 2 ml eppendorf tube and vortexed for 5 seconds. The samples were incubated at room temperature for 10 minutes to allow complete lysis of the cells; and then centrifuged for 30 seconds to get the liquid down the tube. 200 µl of chloroform (Molecular grade) was added to the sample supernatant and vortexed in order to get the phases mixed. The samples were incubated at room temperature for 10 minutes and then centrifuged at 12000 revolutions per minute for 10 minutes at 4 $^{o}$C. The aqueous phase (500-550 µl) was transferred to a new 1.5 ml eppendorf tube. 1 µl of glycogen and 500 µl of isopropanol (Molecular grade) were added, vortexed for 30 seconds and then centrifuged at 12000 revolutions per minute for 10 minutes at 4 $^{o}$C. At this stage, RNA precipitate formed a gel-like pellet on side/bottom of the tube. The supernatant was removed and discarded. 500 µl of 75 % ethanol was added to the RNA precipitate and the tube was inverted gently, centrifuged for 2 minutes at 12000 rounds per minute at 4 $^{o}$C and the supernatant was removed and discarded. The RNA pellet was air dried at room temperature for 10 minutes and stored at –80 $^{o}$C awaiting cDNA synthesis.

### 3.8 Synthesis of cDNA from 16S rRNA

cDNA synthesis, amplification and sequencing were performed at Molecular Research DNA Lab (www.mrdnalab.com, Shallowater, TX, USA). The quality of total RNA was assessed using gel electrophoresis. The extracted RNA was dissolved in RNase-free water and subsequently treated to remove DNA contaminants using the Amplification Grade DNase I Kit (Sigma, MO) according to manufacturer's instructions. cDNA first-strand and second-strand synthesis was done using the Superscript III First-Strand Synthesis SuperMix (Invitrogen, CA) and the Second-strand cDNA Synthesis Kit (BeyoTime, Jiangsu, China), respectively, following manufacturer's instructions. Single-strand reverse transcription was done to provide template for amplicon libraries using Superscript III (Invitrogen) according to the manufacturer's protocol, random hexamer primed and with subsequent RNAse H digestion. The Double stranded cDNA synthesis was carried out as described by Urich *et al.* (2008).

## 3.9 Amplicon Library Preparation and Sequencing

PCR amplification of the 16S rRNA gene V4 variable region was carried out from extracted DNA and cDNA generated from rRNA, using bacteria/archaeal primers 515F (GTGCCAGCMGCCGCGGTAA) that had barcode and 806R (GGACTACHVGGGTWTCTAAT) according to Caporaso et al. (2012). Amplification proceeded in a 30 cycle PCR using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) with initial denaturation heating at 94 °C for 3 minutes, followed by 28 cycles of denaturation at 94 °C for 30 seconds, annealing at 53 °C for 40 seconds and extension at 72 °C for 1 minute, and a final elongation at 72 °C for 5 minutes. The quality of PCR products was assessed on 2 % agarose gel to determine the success of amplification and the relative intensity of bands. Multiple samples, tagged with different barcodes, were pooled in equimolar ratios based on their DNA concentrations from the gel images. Pooled samples were purified using calibrated Ampure XP beads (Beckman Coulter) for use in library preparation. The pooled and purified PCR products were used to prepare 16S rDNA and cDNA library by following Illumina TruSeq DNA library preparation protocol (Yu & Zhang, 2012). Sequencing was performed at MR DNA (www.mrdnalab.com, Shallowater, TX, USA) on MiSeq 2 x 300 base pairs Version 3 following the manufacturer's guidelines.

## 3.10 Sequence analysis and taxonomic classification

Sequences obtained from the Illumina sequencing platform were depleted of barcodes and primers using a proprietary pipeline (www.mrdnalab.com, MR DNA, Shallowater, TX) developed at the service provider's laboratory. Low quality sequences were identified by denoising and filtered out of the dataset (Reeder & Knight, 2010). Sequences which were < 200 base pairs after phred20- based quality trimming, sequences with ambiguous base calls, and those with homopolymer runs exceeding 6bp were removed. Sequences were analyzed by a script optimized for high-throughput data to identify potential chimeras in the sequence files, and all definite chimeras were

depleted as described previously by Caporaso *et al.* (2010a). *De novo* OTU clustering was done with standard UCLUST method using the default settings as implemented in QIIME pipeline Version 1.8.0 at 97 % similarity level. Taxonomy was assigned to each OTU using BLASTn against SILVA SSU Reference 119 database at default e-value threshold of 0.001 in QIIME (Oksanen *et al.*, 2012).

### 3.11 Statistical analysis

Diversity indices (Shannon, Simpson and Evenness) for each sample were calculated using R software version 3.1.3 (DeLong *et al.*, 2006). Community and Environmental distances were compared using analysis of similarity (ANOSIM) test, based upon Bray-Curtis distance measurements with 999 permutations. Significance was determined at 95 % confidence interval ($p = 0.05$). Calculation of Bray-Curtis dissimilarities between datasets, hierarchical clustering, Non Metric Dimensional Scaling (NMDS), redundancy analysis (RDA) and parameter correlation were carried out using the R programming language (DeLong *et al.*, 2006) and the Vegan package (Roesch *et al.*, 2007). To support OTU-based analysis, taxonomic groups were derived from the number of reads assigned to each taxon at all ranks from domain to genus using the taxa_summary.txt output from QIIME pipeline Version 1.8.0.

# CHAPTER FOUR

# BACTERIA AND ARCHAEA DIVERSITY WITHIN THE HOT SPRINGS OF LAKE MAGADI AND LITTLE MAGADI IN KENYA

## 4.1 Introduction

Extreme environment refers to any setting that exhibits life conditions detrimental to living organisms with respect to its physicochemical properties such as pH, temperature, pressure, nutrient and saline concentration (Thiel, 2011). Extreme physicochemical parameters include acidity (pH<5), alkalinity (pH>9), hyper salinity (salinity >35 %), pressure (>0.1 MPa), high temperature (>40 ℃), low temperature (<5 ℃), water stress (aw<0.80), and high-radiation environments (Takai *et al*., 2008). The extreme environments are inhabited by organisms referred to as extremophiles that are so well-adapted that they readily grow and multiply (Rothschild & Mancinelli, 2001). In Kenya, the haloalkaline soda lakes are characterized by exceptionally rich productivity rates presumably because of the high ambient temperatures, high light intensities, availability of phosphates and unlimited access to $CO_2$ in these carbonate rich waters (Melack & Kilham, 1974; Grant *et al*., 1990). Salinity levels can be as high as 30% to saturation in Lake Magadi, whereas the pH ranges between 9 and 11.5 (Warren, 2006). In Lake Magadi, solutes are supplied mainly by a series of alkaline springs with temperatures varying from 33 °C to 86 °C (Warren, 2006; Baumgarte 2003).

Previous culture dependent and culture independent studies on Lake Magadi have revealed a dense and diverse population of aerobic, organotropic, halophilic, alkaliphilic, and haloalkaliphilic and alkalitolerant representatives of major bacterial phyla (Tindall *et al*., 1980, 1984; Mwatha & Grant, 1993; Duckworth *et al.,* 1996; Jones *et al.,* 1998; Grant *et al.,* 1999; Zavarzin *et al.,* 1999; Kambura *et al.,* 2012).

Although conventional microbial cultivation methods have helped shape understanding of physiology and metabolic functions of diverse organisms, they are laborious, time

consuming, selective and biased for specific microbial growth. On the other hand, culture - independent studies done on soda lakes in Kenya have been based on the analysis of clone libraries of PCR amplified rDNA. This may not represent an accurate picture of prokaryotic diversity within a given community due to low speed and coverage of a cloning and Sanger-sequencing based approach, which gives a lower number of amplicon sequences compared to the millions of generated by High Throughput Sequencing technologies such as Illumina Sequencing (Pawlowski *et al.,* 2014).

This is the first culture independent study of the microbial community within the hot springs located around the hypersaline Lakes Magadi and Little Magadi. This study employed Illumina Sequencing of PCR products of both 16S rDNA and 16S rRNA cDNA to obtain a less biased estimation of microbial community within the hot springs' ecosystem. The main objective of this study was to analyze the targeted total community rDNA and cDNA generated from rRNA so as to compare the total versus active microbial communities within the hot springs of Lake Magadi and Little Magadi in Kenya.

## 4.2 Materials and methods

Research authorization, Study site, Measurements of physicochemical parameters, Sample collection, and Nucleic acid extraction, cDNA Synthesis, Amplicon library preparation, amplicon sequencing, sequence analysis and statistical analysis are described in **chapter 3**. Initial **s**equence quality control was done using a proprietary pipeline (www.mrdnalab.com, MR DNA, Shallowater, TX) developed at the service provider's laboratory. *De novo* OTU clustering was done with standard UCLUST method using the default settings as implemented in QIIME pipeline Version 1.8.0 at 97% similarity level. Taxonomic classification was done using BLASTn against SILVA SSU Reference 119 database at default e-value threshold of 0.001 in QIIME (Caporaso *et al*., 2010a and Quast *et al*., 2013). The obtained sequences were submitted to the

NCBI Sequence Read Archive with SRP# Study accessions: SRP061805. These included SRX1124606: RNA-Seq of Prokaryotes: Alkaline Hot springs and SRX1124607: DNA-Seq of Prokaryotes: Alkaline Hot springs (**Appendix 1 - 2**).

## 4.3 Results

### 4.3.1 Sampling

Three hot springs of Lake Magadi and Little Magadi were selected based on different temperature and pH levels. Temperatures ranged from 45.1 ℃ to 83.6 ℃ while pH ranged from 9.2 to 9.8. The Total Dissolved Solids (TDS) was above the measurement range for the Electrical Chemical Analyzer; hence all the readings appeared as one (1). The metadata collected before sampling is summarized in **Table 4.1**.

**Table 4.1: Parameters of sampling sites analyzed in this study**

| Sampling sites | Parameter | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Latitude ºS | Longitude ºE | Elevation (m) | Temperature ºC | pH | EC (mS/cm) | TDS (mg/l) | Dissolved Oxygen (mg/l) |
| **Hot spring 1** | 02º 00 3.7 | 36º 14 32 | 603 | 45.1 | 9.8 | 0.03 | 1 | 12.4 |
| **Hot spring 2** | 01º 43 28 | 36º 16 21 | 611 | 83.6 | 9.4 | 1 | 1 | 0.04 |
| **Hot spring 3** | 01º 43 56 | 36º 17 11 | 616 | 81 | 9.2 | 1 | 1 | 0.71 |

### 4.3.2 Composition and diversity of the microbial communities

After denoising and demultiplexing, a total of 271,345 and 214,663 sequence reads were generated from 16S rDNA and 16S rRNA cDNA data respectively. Total OTU richness at 3% distance amounted to 3502 and 1915 Operational Taxonomic Units (OTUs) respectively. Eighty five (85) and sixty two (62) OTUs were shared across all hot springs while 82 and 45 OTUs were shared across all sample types in the two data sets respectively. **Figure 4.1A** and **B** shows the distribution of OTUs across hot springs and sample types. For 16S rDNA data, hot spring 83.6 ºC and 45.1 ºC showed a relatively larger overlap (142 OTUs) than other hot springs, while 45.1 ºC harbored most OTUs unique to one hot spring in both datasets.

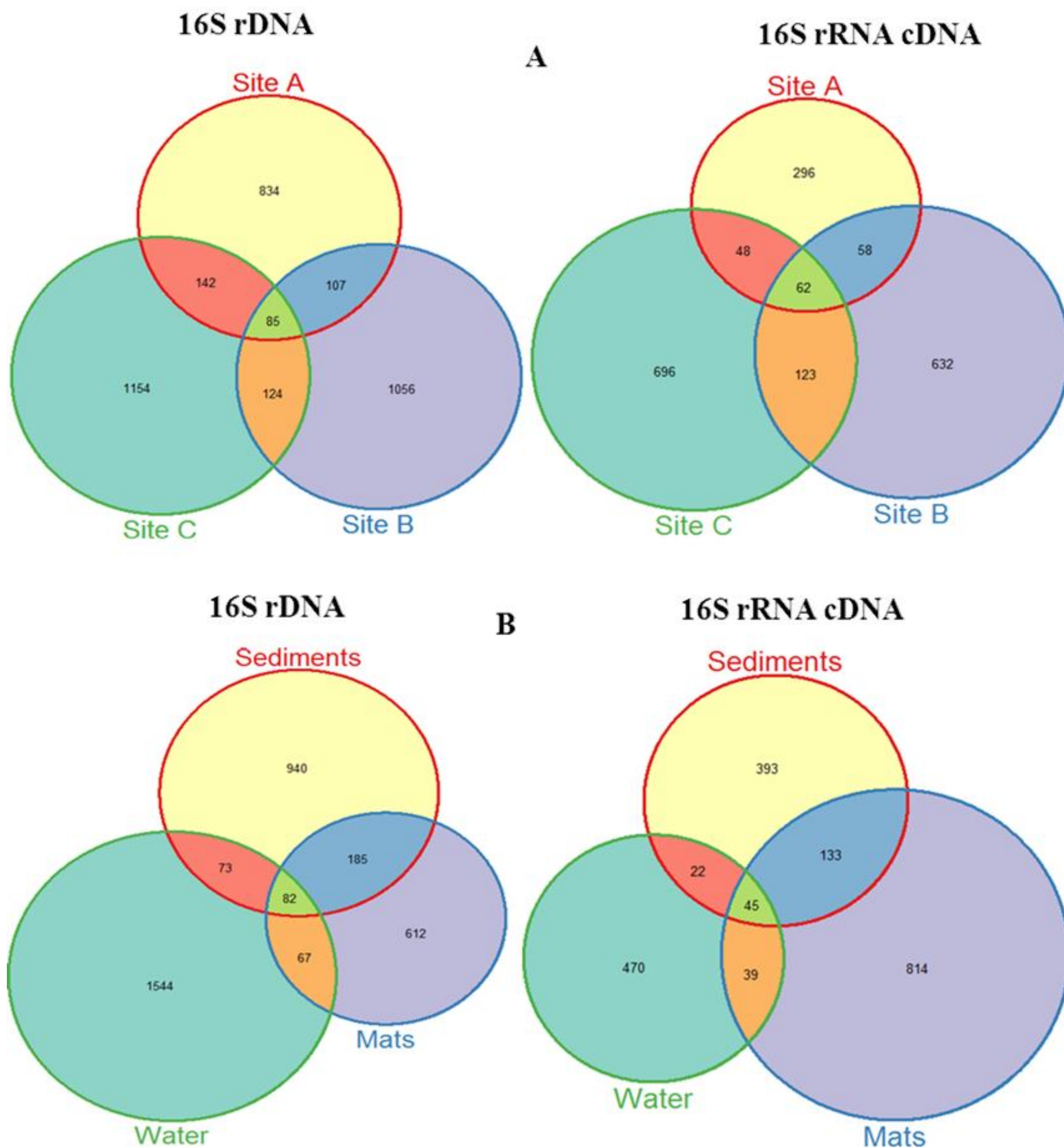


**Figure 4.1: Venn diagrams showing the distribution of unique and shared OTUs within various sample types (B) in the three sampling sites (A). The number of OTUs in each hot spring is indicated in the respective circle. Site A, B and C represent hot springs 83.6 ºC, 81 ºC and 45.1 ºC, respectively.**

### 4.3.3 Total diversity taxonomic composition analysis based on 16S rDNA

#### 4.3.3.1 Bacterial taxonomic composition

16S rDNA OTUs comprised of Bacteria (85.55%) and Archaea (11.27%) while 16S rRNA cDNA OTUs comprised of Bacteria (83.22 %) and Archaea (13.95 %). The groups with highest relative abundances at phylum level belonged to members of *Proteobacteria* (16 – 31.4%), *Firmicutes* (2.8 – 28.9 %), *Bacteroidetes* (3.4 – 23.8 %), *Actinobacteria* (1.0 – 14.4 %), *Cyanobacteria* (1.1 – 8.3 %, *Chloroflexi* (0.3 – 7.2 %), and *Deinococcus-thermus* (0.3 – 5.3 %) (**Fig. 4.2**). Other groups scoring high relative abundance in some samples include *Gemmatimonadetes* (24.1 %) in water samples at 81 ℃, *Planctomycetes* (15.9 %) in water samples at 83.6 ℃, *Lentisphaerae* 5.3% in mat samples at 81 ℃, *Spirochaetae* 4.5 % in wet sediment samples at 45.1 ℃, *Thermotogae* (1.7 %) in wet sediment samples at 83.6 ℃ and *Verumicrobia* (3.9 %) in mat samples at 45.1 ℃.

At family level, OTUs were distributed in 292 bacterial families with the most abundant (9.43%) belonging to *Rhodobacteraceae* in wet sediment samples at 45.1 ℃, *Cyanobacteria* Family I (5.89%), *Desulfonatronaceae* (4.55%), *Thermaceae* (4.01%), *Ectothiorhodospiraceae* (3.71%), *Spirochaetaceae* (3.34%), *Nitriliruptoraceae* (3.19%), *Anaerolineaceae* (3.03%), *Peptococcaceae* (3.03%) and *uncultured gamma proteobacterium* (3.03%) in various samples.

#### 4.3.3.2 Archaeal taxonomic composition

The OTUs were distributed among three Archaeal phyla; *Euryarchaeota* (1.76 - 34.48 %) across all samples, *Crenarchaeota* (up to 2.46 %) and *Thaumarchaeota* (up to 2.78 %) in wet sediment samples at 81 ℃ (**Figure 4.2 and 4.3**). At the family level, OTUs were distributed in 24 families with the most abundant belonging to *Halobacteriaceae* (27.06%) in water samples at 81 ℃, *Deep Sea Hydrothermal Vent Gp 6* (*DHVEG-6*) (0.35 - 5.15%) across all samples, *South African Goldmine Gp* (*SAGMEG*); *Uncultured Archaeon* (4.01%) and *Methanocaldococcaceae* (2.47%) in wet sediment samples at 81

°C. *Crenarchaeota* phyla members identified belonged to the families *Desulfurococcaceae* (0.93%), *Thermoproteaceae* (0.93%), *Thermofilaceae* (0.61%) in wet sediment samples at 81 °C and *Sulfolobaceae* (0.42%) in wet sediment samples at 83.6 °C, while *Thaumarchaeota* were mainly assigned to *uncultured archaeon* with up to 1.68% relative abundance in wet sediment samples at 83.6 °C.

### 4.3.4 Taxonomic composition based on 16S rRNA derived from cDNA

#### 4.3.4.1 Bacterial taxonomic composition

The 1913 OTUs of 16S rRNA cDNA data set were dominated by bacterial phyla comprising *Proteobacteria* (8.3 – 50 %) and *Actinobacteria* (2.0 – 28.6 %) across all samples. The data confirm the predominance of previously defined groups belonging to *Proteobacteria* (Tekere *et al*., 2011). Other groups scoring high relative abundance in some samples include *Bacteroidetes* (26.4% in wet sediment samples at 45.1 °C), *Chloroflexi* (5.3% in water samples at 45.1 °C), *Firmicutes* (23.9% in mat samples at 45.1 °C), *Cyanobacteria* (8% in mat samples at 45.1 °C), *Gemmatimonadetes* (25% in water samples at 81 °C), *Lentisphaerae* (7.1% in wet sediment samples at 83.6 °C) and *Planctomycetes* (6.6% in mat samples at 83.6 °C) (**Fig. 4.2**).

Unlike 16S rDNA dataset, the most abundant bacterial families found within cDNA dataset belonged to *Methylophilaceae;* scoring up to 8.3 % relative abundance in some samples, *Corynebacteriaceae*, *Dermatophilaceae*, *Micrococcaceae, Nocardioidaceae, Bacteroidaceae; ML635J-40 aquatic group, Sphingomonadaceae*, *Burkholderiaceae*, and *Myxococcales; 0319-6G20, Alcanivoracaceae, Pseudomonadaceae* (7.14%), *Desulfuromonadales; GR-WP33-58 (5.11%), Cyanobacteria*; Subsection III; Family I, *Syntrophomonadaceae* (*4.7%*), *Spirochaetaceae* (3.7%), *Anaerolineaceae* (3.44%), *Gemmatimonadetes*; *BD2-11* terrestrial group; uncultured bacterium 3.3%, *Rhodospirillaceae* (3.3%) and *Lentisphaerae* (*ML1228J-2; uncultured bacterium*) (2.6%).

#### 4.3.4.2 Archaeal taxonomic composition

16S rRNA cDNA OTUs were distributed among three Archaeal phyla, similar to those obtained from the 16S rDNA dataset. These were *Euryarchaeota* (6.6 - 28.4 %) across all samples, *Crenarchaeota* (up to 1.38 %) in water samples at 81 ℃ and *Thaumarchaeota* (up to 1.15 %) in wet sediment samples at 45.1 ℃ (**Fig. 4.3**). The most abundant families belonged to *Halobacteriaceae* (21.31%), *Deep Sea Hydrothermal Vent Gp 6* (DHVEG-6) (8.59%), and *Marine Benthic Group D* and *DHVEG-1* (2.29%). *Crenarchaeota* phyla members identified belonged to the families *Desulfurococcaceae* (1.11%) in wet sediment samples at 81 ℃ and *Sulfolobaceae* (0.28%) in wet water samples at 45.1 ℃, while *Thaumarchaeota* were mainly assigned to *uncultured archaeon* with up to 1.15% relative abundance in wet sediment samples at 45.1 ℃.

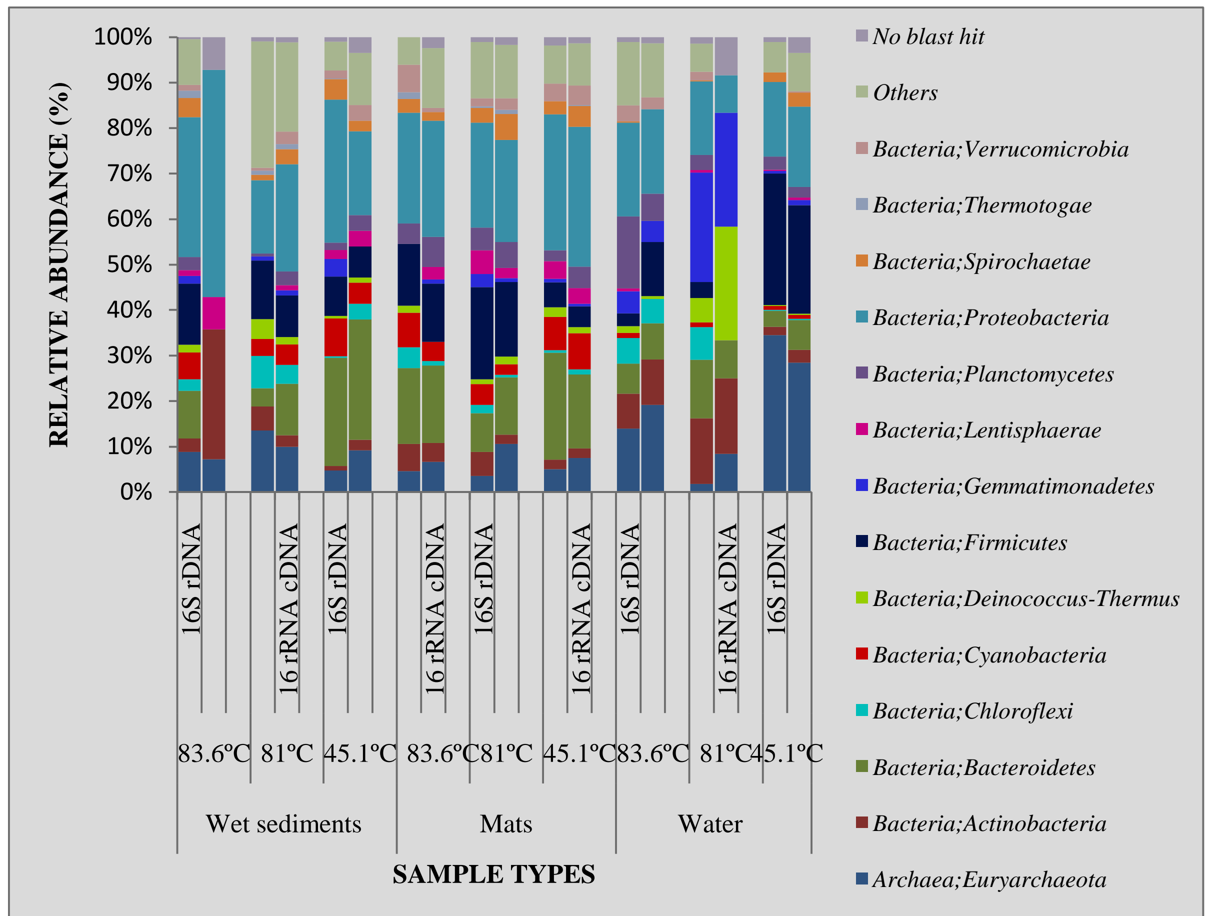


**Figure 4.2: Relative abundance of the most predominant phyla in various samples collected from the hot springs of L. Magadi. 'Others' represent all taxa that scored a relative abundance of below 1% across all samples in both data sets.**

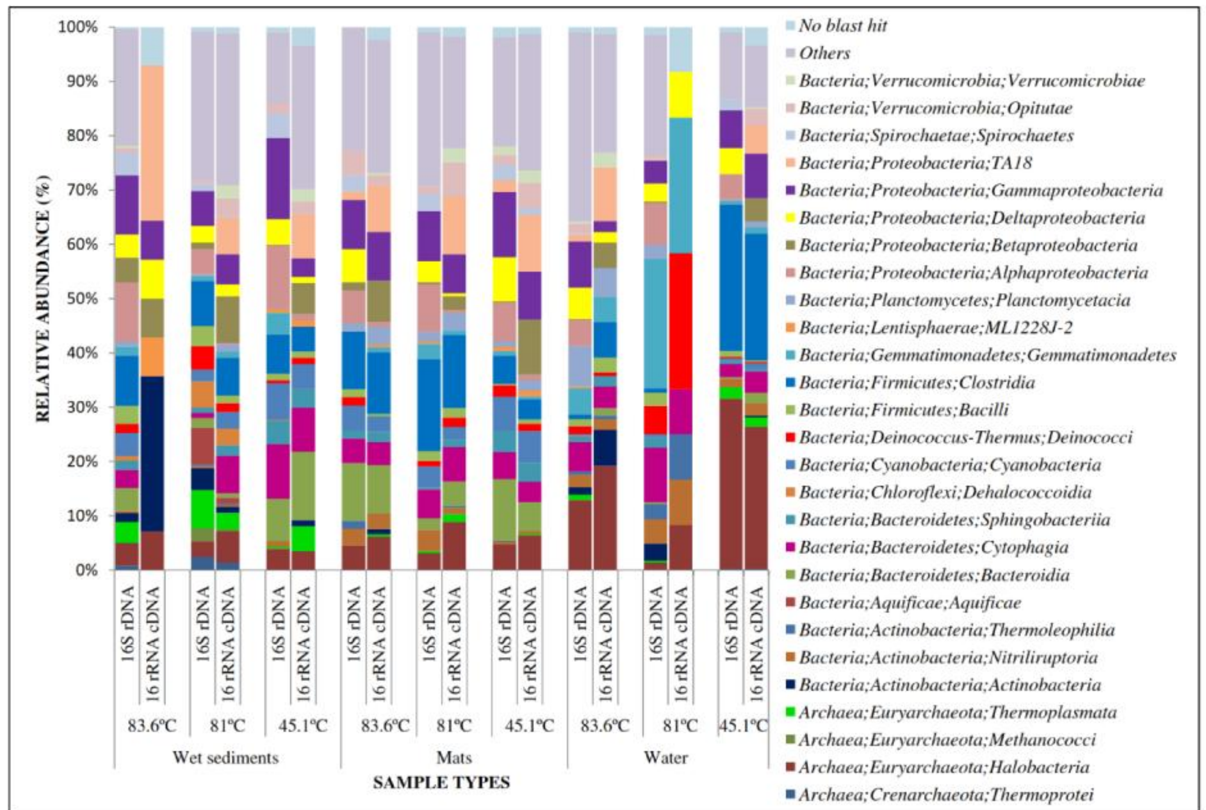


**Figure 4.3: Relative abundance of the most abundant prokaryotic taxa at order level in samples collected from the hot springs of L. Magadi. 'Others' represents all taxa that scored a relative abundance of below 1% across all samples in both data sets.**

### 4.3.5 Microbial richness and diversity indices

Using rarefaction, the same number of sequences from each sample was used in comparison of community alpha and beta diversity measures. Paired t-tests at class taxonomic level of both 16S rDNA and 16 rRNA cDNA indicated that significant differences between samples based on alpha diversity indices whose values obtained were as follows: Shannon diversity index (H'); wet sediment 83.6 ºC (7.9 vs 3.8), water

81 ºC (9.1 vs 3.6) and Simpson (1/D); wet sediment 83.6 ºC (34.54 vs 12.2); water 81ºC (7.76 vs 6) and water 45.1ºC (28.53 vs 10.3), respectively (**Table 4.2**).

**Table 4.2: Diversity indices computed on all OTU-based microbial taxonomic units within DNA and RNA datasets**

| Sample ID | Dataset | No. of sequences | No. of OTUs | Shannon (H') | Simpson (1/D) | Evenness |
|---|---|---|---|---|---|---|
| **wet sediment (83.6ºC)** | 16S rDNA | 38753 | 238 | 5.5 | 34.54 | 0.658 |
| | **16S rRNA cDNA** | 1625 | 14 | 2.6 | 12.2 | 0.975 |
| **Mats (83.6ºC)** | 16S rDNA | 7575 | 66 | 4.2 | 27.41 | 0.475 |
| | **16S rRNA cDNA** | 21451 | 212 | 5.3 | 32 | 0.632 |
| **water (83.6ºC)** | 16S rDNA | 19186 | 680 | 6.5 | 14.88 | 0.356 |
| | **16S rRNA cDNA** | 5124 | 151 | 5.0 | 18.1 | 0.594 |
| **wet sediment (81ºC)** | 16S rDNA | 36047 | 324 | 5.8 | 22.57 | 0.427 |
| | **16S rRNA cDNA** | 67235 | 361 | 5.9 | 53.6 | 0.636 |
| **Mats (81ºC)** | 16S rDNA | 24252 | 282 | 5.6 | 34.55 | 0.61 |
| | **16S rRNA cDNA** | 33751 | 349 | 5.8 | 34.3 | 0.538 |
| **Water(81ºC)** | 16S rDNA | 36126 | 568 | 6.3 | 7.76 | 0.278 |
| | **16S rRNA cDNA** | 638 | 12 | 2.5 | 6 | 0.866 |
| **wet sediment (45.1ºC)** | 16S rDNA | 52115 | 509 | 6.2 | 35.09 | 0.577 |
| | **16S rRNA cDNA** | 11471 | 87 | 4.4 | 27.7 | 0.784 |
| **Mats (45.1ºC)** | 16S rDNA | 37446 | 382 | 6.0 | 22.45 | 0.785 |
| | **16S rRNA cDNA** | 48578 | 375 | 6.0 | 37.4 | 0.572 |
| **water (45.1ºC)** | 16S rDNA | 19789 | 377 | 6.0 | 28.53 | 0.461 |
| | **16S rRNA cDNA** | 24790 | 352 | 5.9 | 10.3 | 0.346 |

Total microbial diversity based on 16S rDNA and 16S rRNA cDNA ANOSIM at order level showed that there were significant differences in microbial community structure in the samples at 95% level of confidence (P value, 0.009), and 0.383 R statistic value while active microbial diversity based on 16S rRNA cDNA had (P value, 0.01), and 0.333 R statistic value. Samples from 45.1 ºC harbored more closely related populations because it had less extreme conditions as compared to the two other hot springs.

Distance based redundancy analysis showed that the microbial community evenness significantly differed from each site for both 16S rDNA and 16S rRNA cDNA. 16S rDNA ordination of the three sample types showed a significance of 0.017 and while 16S rRNA cDNA dataset showed a significance of 0. 011. Samples from each site clustered close to each other in separate quadrants indicating high beta diversity between the three sampling sites.

NMDS analyses supported by OTU and taxonomic composition, divide the datasets into three ellipses: one for each hot spring (**Figure 4.4**). The taxa were also observed to cluster according to sample types (i.e. wet sediments, microbial mats and water samples). There was an overlap of taxonomic clusters between wet sediments and microbial mats from the three sample types in both 16S rDNA and 16S rRNA cDNA derived datasets**.** However, water samples formed separate clusters from the other two sample types in both 16S rDNA and 16S rRNA cDNA datasets **(Figure 4.4** and **4.5)**.

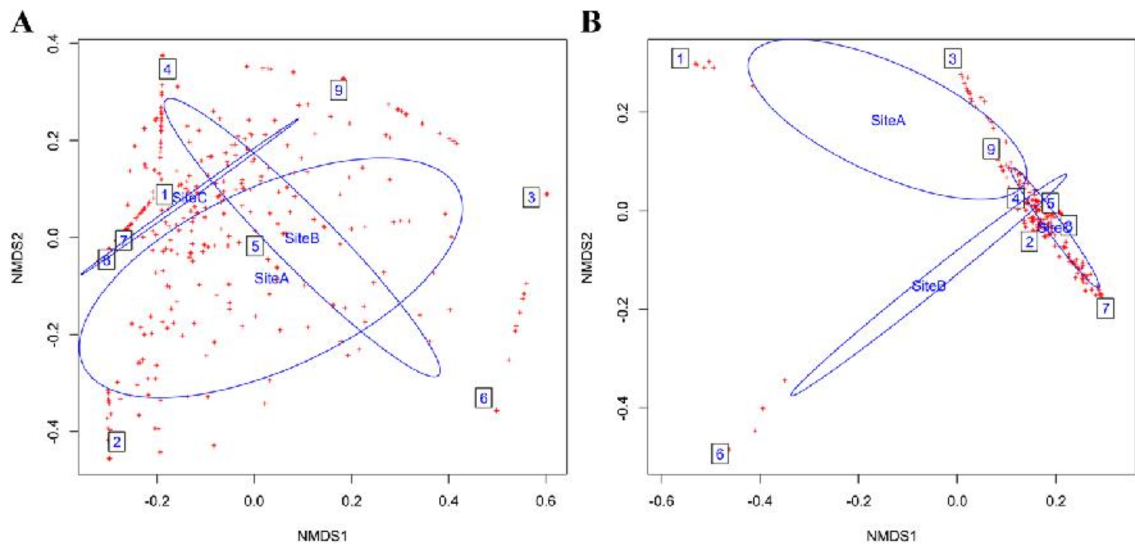


**Figure 4.4: Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities between microbial compositions of (A) 16S rDNA and (B) 16S rRNA cDNA datasets grouped according to sampling sites.**

Site A, B and C represent hot springs 83.6 ºC, 81 ºC and 45.1 ºC respectively. The boxes 1 - 3 represent wet sediments, mats and water samples from 83.6 ºC, 4 - 6 represent wet sediments, mats and water samples from 81 ºC and 7 - 9 represent wet sediments, mats and water samples from 45.1 ºC.

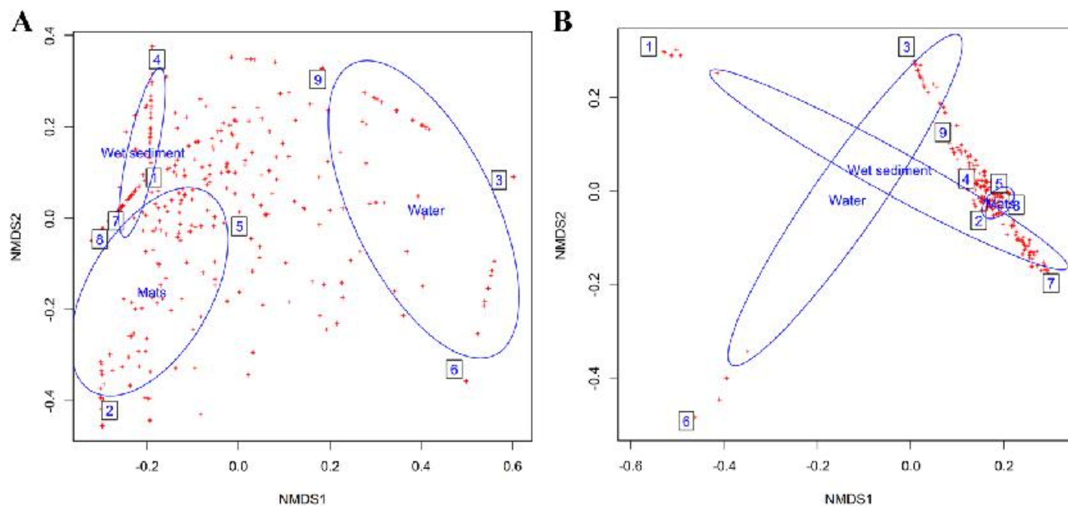


**Figure 4.5: Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities between microbial compositions of (A) 16S rDNA and (B) 16S rRNA cDNA datasets grouped according to sample types.**

Each ellipse represents a set of the three sample types collected from each hot spring. The boxes 1 - 3 represent wet sediments, mats and water samples from 83.6 ºC, 4 - 6 represent wet sediments, mats and water samples from 81 ºC and 7 – 9 represent wet sediments, mats and water samples from 45.1 ºC.

Hierarchical clustering between samples collected from lake Magadi and Little Magadi revealed samples from the two hot springs in Little Magadi "***Nasikie eng'ida"*** were closer than samples from the hot spring in the main lake (**Figure 4.6** **(a** and **b)** and (**Figure 4.7 (a and b**)).

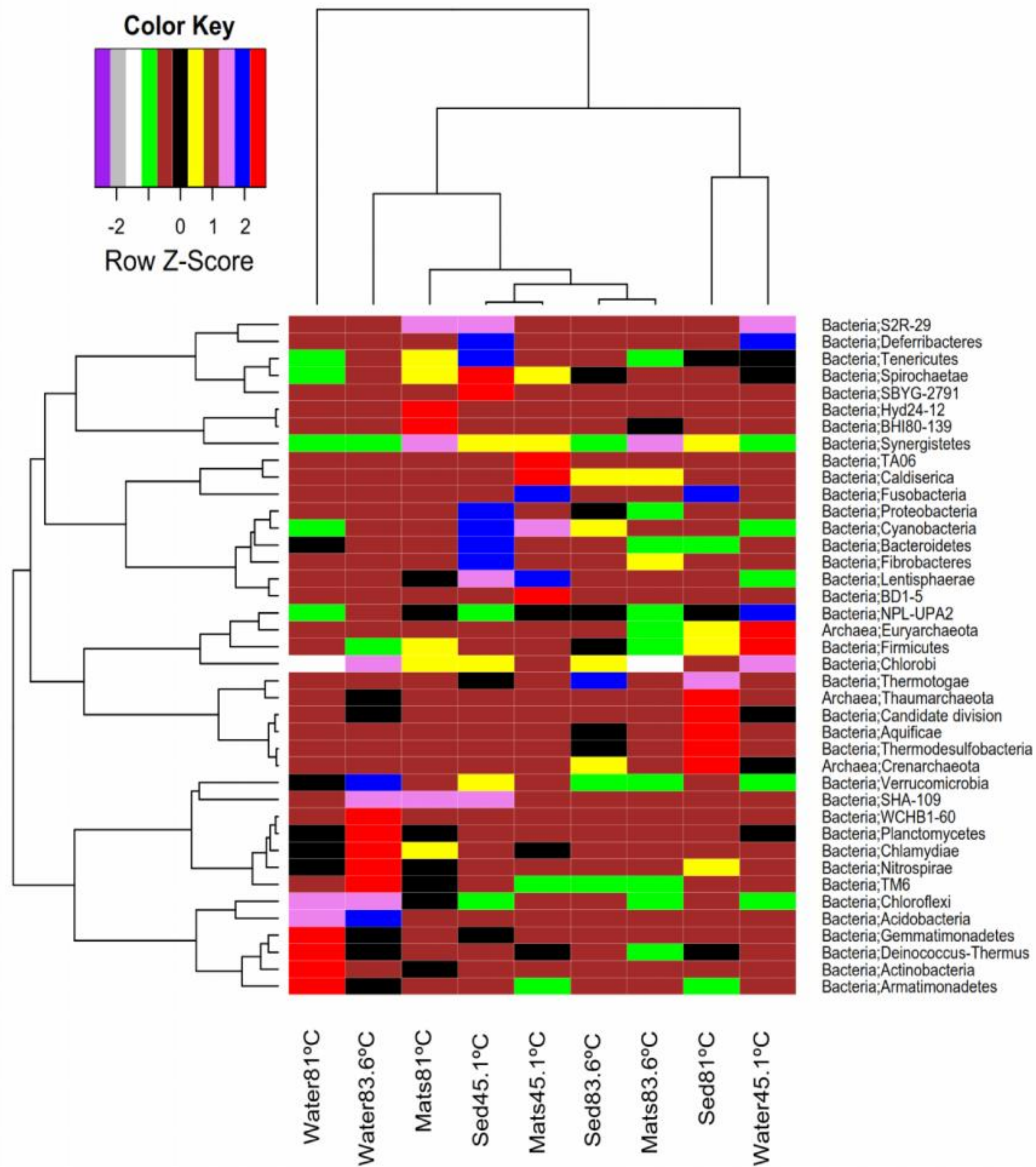


**Figure 4.6 a: Hierarchical clustering of 16S rDNA samples collected from the three hot springs of L. Magadi.** Phylum level was chosen to be used in hierarchical clustering to assess the relationships between samples and taxa. “Sed” represent wet sediment samples from respective temperature.

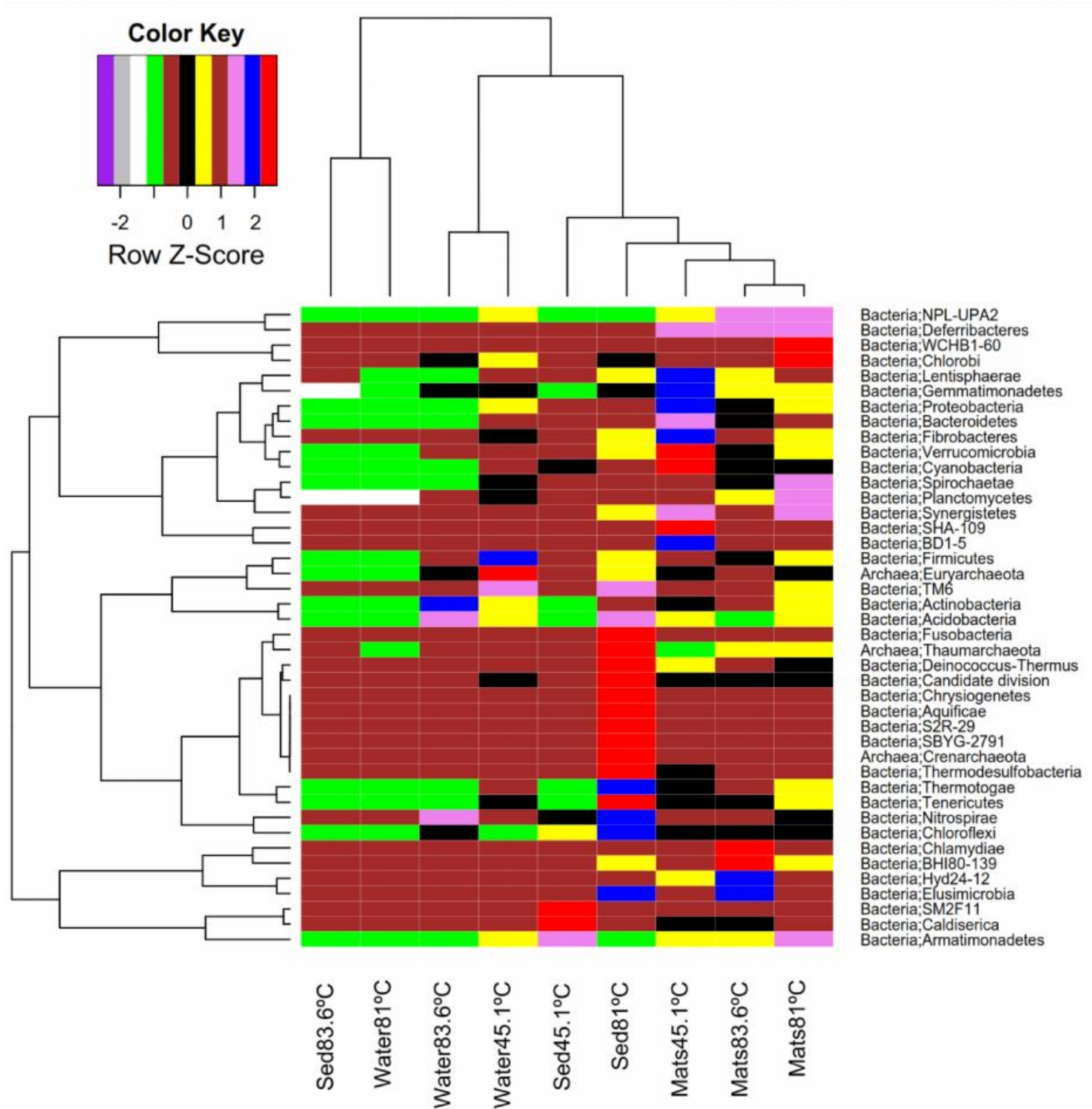


**Figure 4.6 b: Hierarchical clustering of 16S rRNA cDNA samples collected from the three hot springs of L. Magadi.** Phylum level was chosen to be used in hierarchical clustering to assess the relationships between samples and taxa. "Sed" represent wet sediment samples from respective temperature.

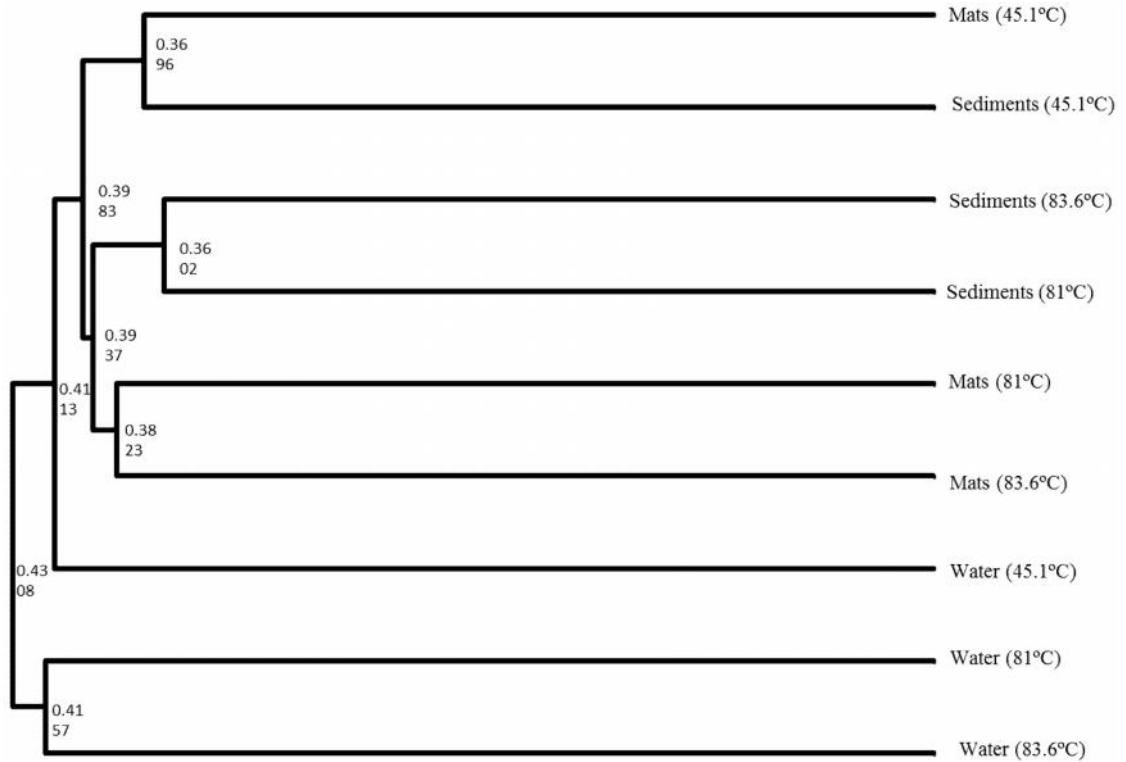


**Figure 4.7 a: Comparative analysis (UPGMA similarity tree) of total microbial diversity of various sample types within hot springs of L. Magadi.**

The numbers at the nodes denote bootstrap values inferred on distances after 999 permutations.

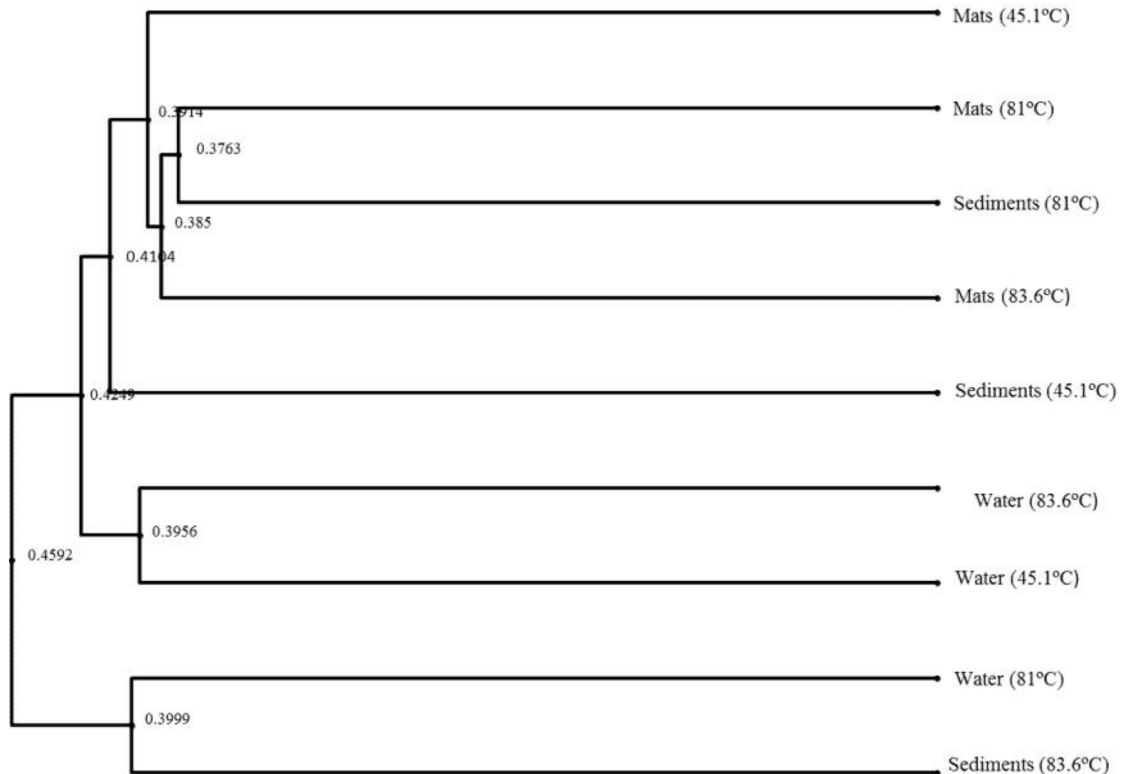


**Figure 4.7 b: Comparative analysis (UPGMA similarity tree) of active microbial diversity of various sample types within hot springs of L. Magadi**. The numbers at the nodes denote bootstrap values inferred on distances after 999 permutations.

## 4.4 Discussion

Differences in relative abundance were seen as a function of sample type and temperature, with wet sediments harboring the highest taxa. Active microbial diversity richness based on cDNA of 16S rRNA was comparable to that of 16S rDNA - derived dataset in most samples, indicating that the observed diversity is real. Although rRNA richness cannot theoretically be higher than that of available DNA, richness estimates suggested this in microbial mats samples from hot springs at 81ºC, 83ºC and wet sediments samples from hot springs at 81ºC samples. The difference in abundance of different groups may have been due to several factors such as possible DNA

contamination in the RNA that could have caused false positives in the final PCR, challenges encountered during sampling leading to low quality RNA and PCR bias that might have favored amplification of the RNA (cDNA) templates more than the rDNA template which is relatively longer, as reported by Lanze´n et al. (2013).

The dominant taxa corresponded with those reported in previous studies conducted on deep sea and marine sediment community composition (DeLong *et al*., 2006; Brown *et al*., 2009; Liao *et al*., 2011). For example, a review by Brown *et al.* (2009) on microbial life in extreme environments that compared metagenome analyses of different high thermal habitats, observed that microbes adapted to these habitats, are different with respect to species abundance and community structure. However, some bacterial taxa such as *Thermotoga, Deinococcus-Thermus* and *Proteobacteria*, were common within the samples under review (Brown *et al*., 2009). These bacterial taxa were also found within the samples from the hot springs of Lakes Magadi and Little Magadi. The deep-sea hydrothermal vent chimneys have been found to harbor *Proteobacteria* (Brazelton *et al*., 2011; Nakai *et al*., 2011), *Bacteroidetes* and *Planctomycetes* (Xie *et al*., 2011).

Several 16S rRNA gene sequences related to novel Archaea (Euryarchaeota) were previously retrieved from the alkaline saltern at Lake Magadi (Grant *et al*., 1999). Haloalkaliphilic Archaea related to *Natronomonas*, *Natrialba*, *Natronolimnobius* and *Halorubrum spp*. have also been isolated from Lake Magadi and Inner Mongolian soda lakes (Grant & Sorokin, 2011). Other previous studies on thermal groundwater in a thermal field in Russia showed that Archaea is dominated by a novel division in the phylum *Euryarchaeota* related to the order *Thermoplasmatales* (39% of all archaea) and by another abundant group (33% of all archaea) related to the phylum *Crenarchaeota.* Both groups are widely spread in hot springs all over the world (Mardanov *et al*., 2011). Some Archaeal taxa such as *Methanococcus, Thermoprotei* and *Thermococcus* were also common within the samples under review by Brown et al. (2009). These are similar to the classes obtained in this study, indicating that Archaea

are well adapted to extreme conditions and could be responsible for various functional processes within the ecosystem.

Non Metric Dimensional Scaling (NMDS) analysis indicated some microbial taxa to be shared between habitats in both 16S rDNA and 16S rRNA cDNA derived datasets. This scenario was more pronounced in 16S rDNA - derived dataset, indicating that DNA pool contained a ''seed bank'' of inactive and sporulating organisms (Jones and Lennon 2010), while fewer taxa were active within the ecosystem as shown in the 16S rRNA cDNA derived dataset. NMDS analyses supported by OTU and taxonomic composition, divide the datasets into three ellipses: one for each hot spring (**Figure 4.4**). Similar results were observed in a study on Ethiopian soda lakes where NMDS analyses supported both by OTU and taxonomic composition; divided the datasets into six well-separated habitats with relatively few OTUs that were shared between more than one or two habitats (Lanze´n *et al*., 2013). The significant differences in microbial community structure in the samples at 95% level of confidence could be attributed to differences in temperature and pH of the specific sites during sampling.

## 4.5 Conclusion

The combined findings of this study show that estimated diversity and richness within the hot spring samples were found to be as high as those found in other environments such as soil and deep-sea hydrothermal environments (Zavarzin *et al*., 2000). The results confirm that different groups of microorganisms have the capacity to adapt and thrive even in the most hostile environments. Some of these groups (*Acidobacteria; Blastocatella, Bryobacter* and *Telmatobacter* genera, *Bacteroidetes; Bacteroidales, Rhodothermaceae, Flavobacteriaceae, Sphingobacteriales* and *Chloroflexi; Dehalococcoidales and Thermomicrobia*) also obtained from previous similar studies were reported to have a fermentative ability (Lanze´n *et al*., 2013). It was observed that from the cDNA dataset, photosynthetic taxa were represented by Cyanobacterial genera *Leptolyngbya* and *Lyngbya*, among other uncultured groups. Primary production within

the hot springs is probably supported by some groups of non-sulfur purple bacteria from the family *Rhodobacteraceae* (specifically the genera *Roseobacter* that scored 1.1 % relative abundance), and purple sulfur bacteria from the family *Ectothiorhodospiraceae* present across samples at different relative abundance. The presence of *Planctomycetes* within samples could be an indicator that anaerobic ammonium oxidation may be another metabolic pathway supporting primary production in the low-oxygen, saline environment, since the dissolved oxygen concentration of the sampling sites ranged between 0.04 – 12.4 mg/l. *Actinobacteria* and *Firmicutes* are believed to have adaptive advantage under low-nutrient conditions of the highly alkaline, saline hot springs hence their high relative abundance levels. The presence of sulfate reducers in the family *Desulfohalobiaceae* (mainly *Desulfonatronovibrio*), suggested an internal sulfur cycle within the lake, as previously suggested for Khadin, Tuva, Russia and Natron soda lakes (Zavarzin *et al*., 2000; Sorokin *et al*., 2011). Taxa typical for highly specialized metabolisms that were encountered in this study include *Nitriliiruptor*, known for their ability to catabolize nitriles or cyanides (Bengtsson *et al*., 2010) and heterotrophic *Oceanospirillaceae* (*Marinospirillum*). Other functional taxa encountered include aerobic heterotrophs (e.g. *Bacteroidetes*, *Marinicella*) and fermentative anaerobes such as *Thermoplasmatales* among other uncultured groups. *Euryarchaeota* members were clustered into the classes, *Halobacteria*, *Methanobacteria*, *Methanomicrobia*, *Methanococci*, *Thermococci* and *Thermoplasmata* while *Crenarchaeota* phyla comprised *Thermoprotei* class. Previously, *Methanobacteria* and *Methanomicrobia* have been reported in oilfields, while *Halobacteria* and *Thermoprotei* have been reported in petroleum reservoirs (Lenchi *et al*., 2013). Some *Halobacteria* members are important in organic fertilizer production industry as Lignin decomposers (Nura *et al*., 2015). However, most of the genera identified in this study are known to be heterotrophs responsible for the primary degradation of organic matter (Grant & Sorokin, 2011). The actual function of microbial taxa reported in this study could further be explored and established using culture dependent methods as well as mRNA transcripts.

This study presented microbial diversity analysis of samples collected from the hot springs of L. Magadi and Little Magadi based on both DNA and RNA, using Illumina Sequencing Technology. The results showed comparable profiles of microbial community using 16S rDNA and 16S rRNA cDNA derived datasets, hence indicating that the observed diversity is real. The findings showed a broad microbial distribution with water from the spring at 83.6°C found to be the richest sample, constituting 680 observed species. Despite the fact that the sampling environment is multi-extreme due to high pH, temperature, and salinity, this study shows that there are stable and active microbial communities that have adapted to this environment.

# CHAPTER FIVE

# DIVERSITY OF FUNGI IN SEDIMENTS AND WATER COLLECTED FROM THE HOT SPRINGS OF LAKE MAGADI

## 5.1 Introduction

Fungi have colonized diverse habitats such as tropical regions (Hawksworth, 1991), extreme environments such as deserts, areas with high salt concentrations (Vaupotic *et al.,* 2008), ionizing radiation (Dadachova *et al.,* 2007), deep sea sediments (Raghukumar & Raghukumar, 1998) and ocean hydrothermal areas (Le Calvez *et al*., 2009). Most fungi grow in terrestrial environments, though several species live partly or solely in aquatic habitats, such as the chytrid fungus *Batrachochytrium dendrobatidis*, a parasite that has been responsible for a worldwide decline in amphibian populations (Brem & Lips, 2008).

In most ecosystems, fungi are the major decomposers, playing an essential role in nutrient cycling as saprotrophs and symbionts that degrade organic matter into inorganic molecules (Barea *et al.,* 2005, Lindahl *et al.,* 2007 and Gadd, 2007). While there are well-known examples of bacteria that can grow in a variety of natural environments including hot springs and geysers where temperatures can reach 100  C, eukaryotes are much more sensitive because, above 65  C, their membranes become irreparably damaged (Magan & Aldred, 2007). However, mesophilic thermo-tolerant fungi exist. For example, some *Deuteromycetes* isolated from thermal springs have maximum growth temperature of 61.5  C (Magan, 2006). The presence of fungi in extreme alkaline saline environments has been recognized by culture-dependent methods, with the majority showing similarity to terrestrial species (Mueller & Schmit, 2006; Salano, 2011, Ndwigah *et al.,* 2015). Culture-independent methods have revealed highly novel fungal phylotypes such as *Chytridiomycota* and unknown ancient fungal groups (Yuriko & Takahiko, 2012). pH tolerance in fungi has been attributed to efficient control of

proton movement into and out of the cells, and is able to meet necessary energy requirements (Magan & Aldred, 2007). The exact diversity and function of fungi in extreme environments is still poorly understood. The aim of this study was to explore the fungal diversity within the hot springs of Lake Magadi and Little Magadi in Kenya using metagenomic analysis.

## 5.2 Materials and Methods

Research authorization, study site, measurements of physicochemical parameters, Sample collection, and DNA extraction are described in **chapter 3**. PCR amplification of ITS region was done using ITS1 (TCCGTAGGTGAACCTGCGG) and lTS4 (TCCTCCGCTTATTGATATGC) primers with barcode according to (White *et al*., 1990). Amplification proceeded in a 30 cycle PCR using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) with initial heating at 94 °C for 3 minutes, followed by 28 cycles of denaturation at 94 °C for 30 seconds, annealing at 53 °C for 40 seconds and extension at 72 °C for 1 minute, after which a final elongation step at 72 °C for 5 minutes was performed. PCR products were visualized on a 2% agarose gel to determine the success of amplification and the relative intensity of bands. Multiple samples were pooled together in equal proportions based on their DNA concentrations. Pooled samples were purified using calibrated Ampure XP beads (Agencourt Bioscience Corporation, MA, USA). The pooled and purified PCR product was used to prepare DNA library by following Illumina sequencing protocol (Yu & Zhang, 2012). Sequencing was performed at Molecular Research DNA (www.mrdnalab.com, Shallowater, TX, USA) on a MiSeq platform following the manufacturer's guidelines. Sequence analysis, taxonomic classification and data Submission are described in chapter 3. The obtained sequences were submitted to the NCBI Sequence Read Archive with SRP# Study accessions: DNA-Seq of Fungal ITS: Alkaline Hot springs: SRP061806.

## 5.3 Results

Wet sediment and water samples were randomly collected at three different locations in hot springs of Lake Magadi and Little Magadi. Physico-chemical parameters of sampling stations in Lake Magadi and Little Magadi measured before sampling. The physio-chemical parameters of sampling stations in Lake Magadi and Little Magadi measured before sampling are summarized in **Table 4.1** shown in **chapter 4**.

Temperature measurement showed a gradient from hot spring in the main Lake Magadi, with the springs at Little Magadi measuring between 81 °C and 83.6 °C. Cation analysis of the water samples showed that the levels of calcium range between 0.33-0.62 ppm, iron (<0.01- 0.014 ppm) and magnesium (<0.02 - 0.026 ppm). Sodium levels were very high (11,300 ppm, 17,300 ppm and 17,700 ppm) and potassium levels were 225 ppm, 458 ppm and 287 ppm. Anion analysis showed that phosphorus range from 2.72 ppm - 6.31 ppm. Chloride levels were high in all samples ranging from 4000 ppm – 4640 ppm (**Table 5.1**).

**Table 5.1: Chemical analyses of samples from the Hot Springs Lake Magadi**

| Parameter | Sampling sites | | |
|---|---|---|---|
| Chemical properties | Hot springs 45.1ºC | Hot springs 81ºC | Hot springs 83.6ºC |
| pH | 9.61 | 9.2 | 9.41 |
| EC (mS cm-1) | 30.3 | 30.5 | 29.9 |
| Ammonium (ppm) | 0.94 | 2.66 | 2.57 |
| Calcium (ppm) | 0.62 | 0.53 | 0.33 |
| Magnesium (ppm) | <0.02 | 0.026 | <0.02 |
| Potassium (ppm) | 287 | 458 | 225 |
| Phosphorus (ppm) | 6.31 | 4.17 | 2.72 |
| Nitrate N (ppm) | 0.53 | 0.67 | 0.67 |
| Nitrates (ppm) | 2.35 | 2.98 | 2.97 |
| Sulphur (ppm) | 129 | 107 | 58.9 |
| Sulphates (ppm) | 387 | 322 | 176 |
| Iron (ppm) | <0.01 | 0.012 | 0.014 |
| Manganese (ppm) | 0.016 | 0.012 | <0.01 |
| Zinc (ppm) | <0.01 | <0.01 | <0.01 |
| Boron (ppm) | 9.3 | 15.5 | 8.06 |
| Copper (ppm) | 0.043 | <0.01 | <0.01 |
| Molybdenum (ppm) | 0.14 | 0.12 | 0.071 |
| Sodium (ppm) | 17700 | 17300 | 11300 |
| Chlorides (ppm) | 4000 | 4640 | 4220 |
| Bicarbonates (ppm) | 14200 | 17500 | 17100 |
| Silicon (ppm) | 28.1 | 51.1 | 23.8 |

## 5.4 Sequence data

Four (4) out of the nine (9) sequenced hot spring samples ( three sediment and one water sample) generated a total of 548,639 sequences, of which 334, 394 sequences were retained after removing sequences with different tags at each end for quality filtering and denoising. All mat samples and two water samples did not generate sufficient

sequence data for further analysis. After removing singletons, chimeric sequences and OTUs of non-fungal organisms (< 200 base pairs after phred20- based quality trimming, sequences with ambiguous base calls, and those with homopolymer runs exceeding 6bp), a total of 151 fungal OTUs recovered at 3% genetic distance, were included in the final analysis.

### 5.5 Composition and diversity of fungal communities

Based on BLASTn searches in SILVA SSU Reference 119 database, 151 OTUs were identified, most of which had their best matches against accessions in SILVA database. These 151 OTUs spanned 5 phyla namely; *Ascomycota*, *Basidiomycota, Fungi unspecified phylum*, *Chytridiomycota* and *Glomeromycota*. Sediment samples collected from 81 $^{o}$C had the highest number of OTUs (135 OTUs) while 88 OTUs were shared among all sample types (**Figure 5.1**).

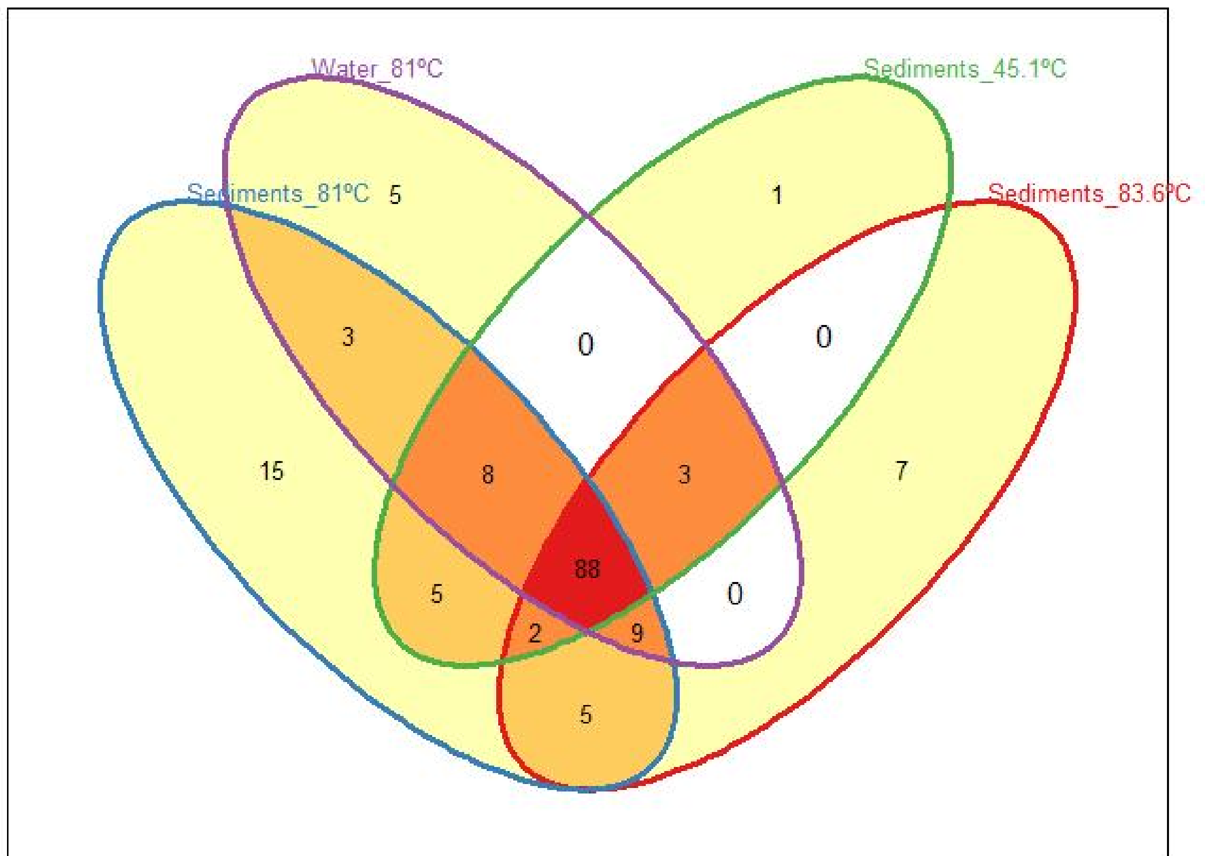


**Figure 5.1: Venn diagram showing the distribution of unique and shared OTUs within various sample types in the three sampling sites. The number of OTUs in each hot spring is indicated in the respective circle.**

The shared OTUs were distributed among the phyla; *Ascomycota* (up to 42.9% relative abundance in sediment sample at 45.1 $^{o}$C), *Fungi unspecified phylum* (up to 6.2% relative abundance in sediment sample at 83.6 $^{o}$C), *Basidiomycota* (up to 3.3% relative abundance in sediment sample at 81 $^{o}$C), *Chytridiomycota* and *Glomeromycota* (up to 1.5% relative abundance each in water sample at 81 $^{o}$C. OTUs belonging to the Phylum *Ascomycota* were the most abundant and were represented by the most genera as shown in Figure 2. In sediments at 45.1$^{o}$C the OTUs were affiliated to the genus *Aspergillus, Penicillium* and *Trichocomaceae*. *Aspergillus oryzae* was the most abundant species with a relative abundance of 42.86%. Other species present were *Penicillium sp*. 5/975 and *Trichocomaceae sp*. lm65 with 14.29% relative abundance (**Figure 5.2**). However in the sediment sample at 83.6 $^{o}$C genera represented were *Aspergillus*, *Aureobasidium*, *Cladosporium*, *Malassezia*, *Neurospora*, *Penicillium*, *Phaeosphaeria*, *Pleosporales*, *Radulidium* and *Trichocomaceae*. *Basidiomycota* phylum comprised *Rhodotorula* and *Termitomyces* species (**Figure 5.2**). *Malassezia* sp., *Neurospora* sp., *Ascomycota* sp. lm221, *Aspergillus aculeatus, Aureobasidium pullulans* and *Cladosporium cladosporioides* are unique to sediment samples at 83.6 ºC. In the sediment sample collected at 81 $^{o}$C, *Aspergillus* sp. (*terreus, oryzae, flavus* and *Aspergillus* sp. bf7) within the phylum *Ascomycota* were the most abundant with *Aspergillus oryzae* scoring 18.75%. Other taxa represented in the sample included *Ascomycota* sp. ar_2010, *Ramularia eucalypti,* and *Stagonospora* sp. vegae284 each scoring a relative abundance of 6.25%. *Antarctic fungal* sp. gi944 of *Fungi unspecified phylum* also scored a relative abundance of 6.25% (**Figure 5.2**). The water sample collected at 81 ºC was found to harbor a higher diversity of fungi with low species richness (**Figure 5.2**).

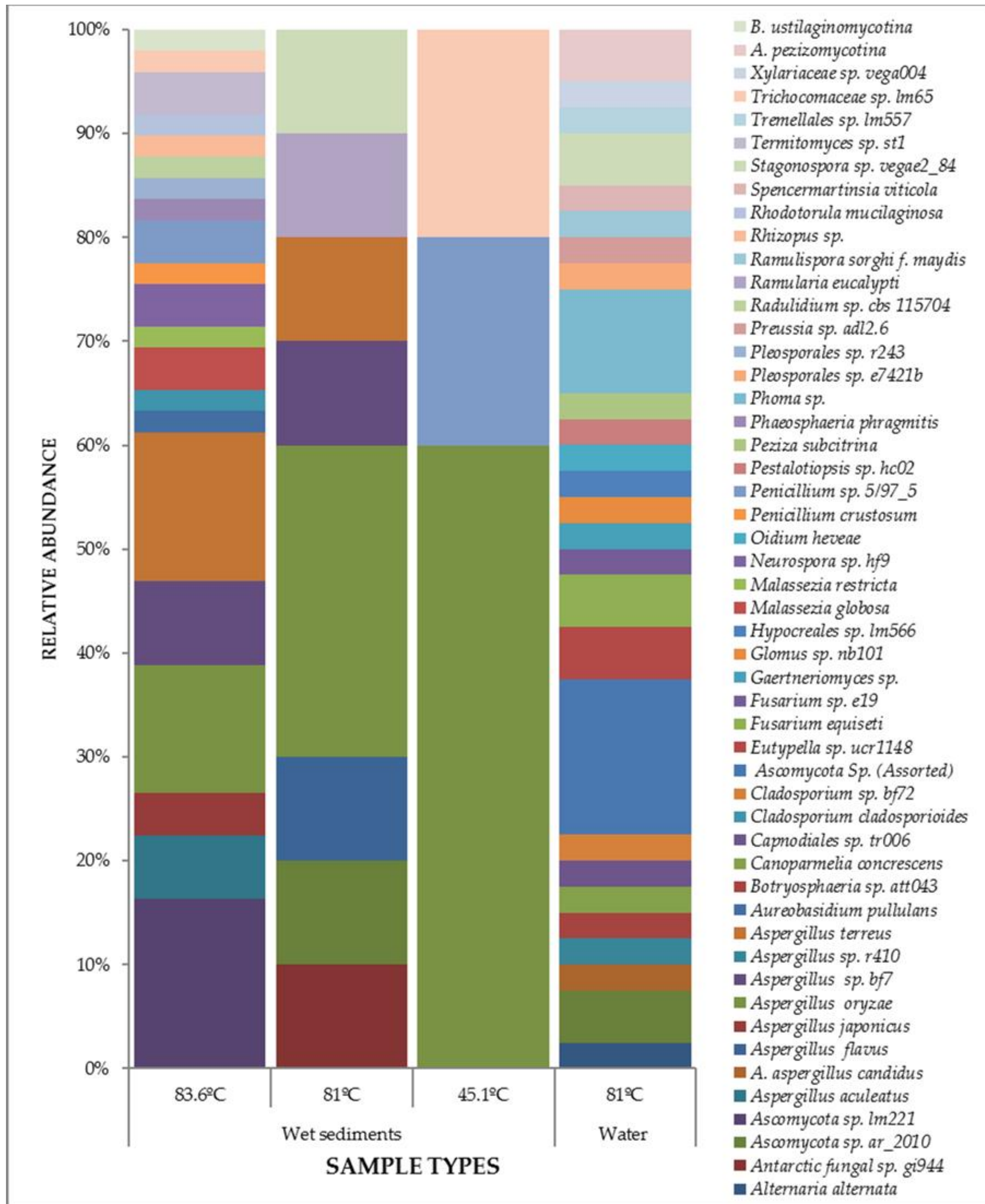


**Figure 5.2: Relative abundance of the most predominant fungal species in various samples collected from the hot springs.** Major species belong to *Ascomycota, Basidiomycota, Chytridiomycota, Glomeromycota* and *unspecified fungal* phyla.

Hierarchical clustering between samples collected from Lake Magadi revealed sediment samples from the two hot springs in Little Magadi "***Nasikie eng'ida"*** to be closer than the sample from the hot spring in the main lake. Majority of the groups at species level included *Aspergillus*, *Ascomycota*, *Penicillium*, *Neurospora*, *Termitomyces*, *Malassezia*, *Trichocomaceae*, *Stagonospora*, *Ramularia* and *Hypocreales* (**Figure 5.3**). The dendogram shows relationship between the four samples.

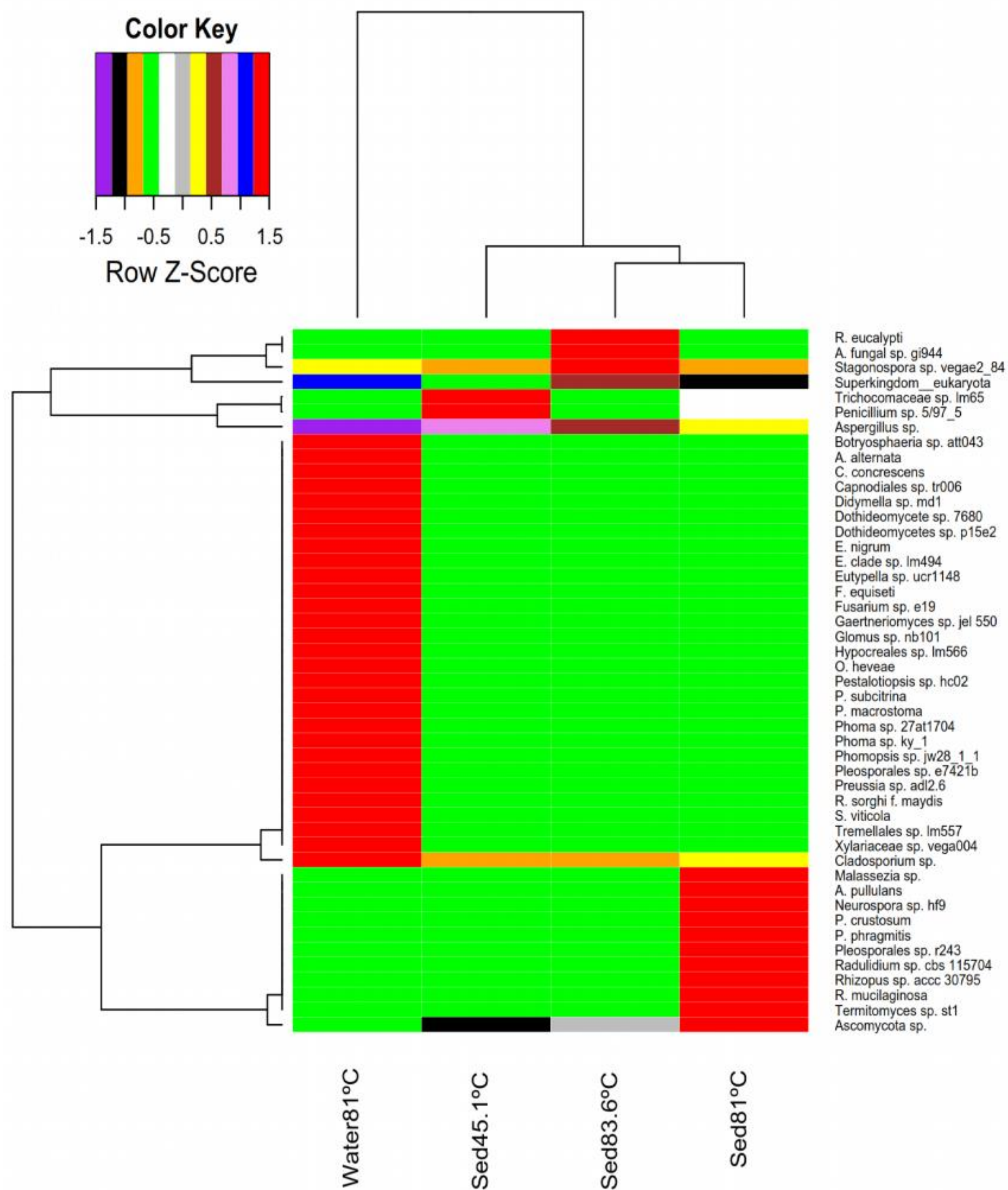


**Figure 5.3: Hierarchical clustering of DNA samples collected from the three hot springs of L. Magadi and Little Magadi**. Species level was chosen to be used in hierarchical clustering to assess the relationships between samples and taxa.

## 5.6 Fungal richness and diversity indices

Richness (S) estimated the water sample (81°C) to be the richest site, constituting 35 taxa. Sediment samples from the three sites had Evenness (J') scores close to 1 (0.663 – 0.897), hence showing evenness in their number of taxa members than the water sample (81°C)**.** Simpson (1/D) also indicated the sediment 45.1°C to harbor the most diverse taxa (9.98). The Shannon's index (H' = 1.28–2.63) indicated low variation in the level of diversity among the sediment and water samples (**Table 5.2**).

**Table 5.2: Diversity indices computed on all OTU-based fungal taxonomic units obtained from samples collected from the hot springs of L. Magadi and Little Magadi**

| Sample | No. of sequences after filtering | No. of OTUs | Richness (S) | Shannon(H') | Simpson (1/D) | Evenness (J') |
|---|---|---|---|---|---|---|
| Wet sediment (81ºC) | 112, 262 | 61 | 9 | 1.90 | 4.92 | 0.739 |
| Wet sediment (45.1ºC) | 59,138 | 7 | 21 | 2.63 | 9.98 | 0.663 |
| Wet sediment (83.6ºC) | 80,702 | 16 | 4 | 1.28 | 3.27 | 0.897 |
| Water (81ºC) | 82,292 | 67 | 35 | 2.66 | 5.06 | 0.410 |
| **Total** | **334, 394** | **151** | | | | |

Analysis of similarity and distance based redundancy analysis at genus level showed connectivity of distance matrix with threshold dissimilarity of 1 indicating that data of the four samples are connected ([1] 1 1 1 1), hence there were no significant differences in community structure in the samples at 95% level of confidence (P value, 0.05).

## 5.7 Discussion

The significance of fungal communities in the hot springs of hypersaline lakes is unclear, mainly because data on the fungal species in these habitats is limited. Using

traditional culture-based methods, researchers reported relatively low levels of diversity for fungal communities in extreme environments (Salano, 2011). In this study high-throughput sequencing technology was used in order to comprehensively analyze fungal communities within the hot springs. The high sensitivity of Illumina sequencing enables the detection of rare species, thus providing more detailed information on fungal diversity in these habitats. Members of *Ascomycota* were more frequently identified in the hot springs than those of *Basidiomycota*, whereas members of *Chytridiomycota* and *Glomeromycota* represented only a small proportion of the hot spring fungal communities. The abundance of *Ascomycota* is similar to the abundance of fungi determined in the previous study on soda soils at the edge of several lake basins, where filamentous fungi that could grow at high ambient pH values were isolated (Alexey *et al*., 2015). The results in that study revealed 100 strains of fungi with various degrees of alkali tolerance and taxonomic affinity within *Ascomycota* (Alexey *et al*., 2015). Additionally, 6.2% of the fungi detected in wet sediment at 83.6ºC were unspecified Phylum. These may be undiscovered and possibly indigenous species in the hot springs. Some of the groups in this study are similar to those recovered from a previous culture dependent study conducted on the Hot spring in main Lake Magadi (Salano, 2011). The isolates recovered in that study belonged to *Aspergillus, Penicillium*, *Neurospora, Polyozellus multiplex, Pycnoporus* sp., *Teratosphaeria, Acremonium, Talaromyces*, *Sagenomella, Paecilomyces* and *Aphanoascus* genera (Salano, 2011). Filamentous fungi like *Aspergillus* and *Penicillium* are attractive organisms for production of useful protein and biological active secondary metabolites. They have high secretion capacity and are effective hosts for the production of foreign proteins (Tsukagoshi *et al*., 2001).

*Penicillium* genus were found in wet sediments 45.1 ºC and 81 ºC with relative abundance of 14.29 % and 3.28 % respectively. This is similar to previous studies in hypersaline water of salterns that revealed different species of *Aspergillus, Penicillium* and diverse non-melanised yeasts (Gunde-Cimerman *et al*., 2005). Another study that used morphological and molecular techniques to identify a series of halotolerant fungi from hypersaline environments of solar salterns revealed 86 isolates of 26 species from

salt ponds, which were identified as *Cladosporium cladosporioides*, nine *Aspergillus* sp., five *Penicillium* sp. and the black yeast *Hortaea werneckii* (Cantrell *et al*., 2006). *Rhodotorula mucilaginosa,* a yeast species and *Rhizopus* sp. 30795, a *Zygomycota* were found in wet sediment at 81 ºC while unclassified *Antarctic fungal* sp. gi944 dominated wet sediment at 83.6 ºC. Other plant pathogenic fungi recovered included *Fusarium* sp., *Cladosporium cladosporioides, Aspergillus flavus, Aspergillus japonicas* and *Aspergillus oryzae.* Most of these organisms may have found their way to the hot springs through various dispersal mechanisms or may be adapted in these extreme environments.

According to Frontier (1985), harsh environments experiencing one or more extreme conditions tend to harbor fewer species. In contrast, wet sediments at hot spring 45.1 ºC were found to have the least OTUs (107 OTUs) as compared to higher temperature samples, distributed within *Aspergillus oryzae* (42.86 %), *Penicillium* sp. 5/97_5 (42.28 %) and *Trichocomaceae* sp. lm65 (42.28 %). Although water samples at 81 ºC were found to harbor a higher diversity of fungi with lower species richness, wet sediments showed a lower diversity with high abundance of present groups. This could be due to high abundance of organic matter and lower oxygen levels which favored decomposition processes; hence the groups present have sufficient carbon sources (Neira *et al*., 2001; Buee *et al*., 2009). The widespread fungal groups within the wet sediments may therefore be degraders of organic matter (Edgcomb *et al*., 2011a; Nagahama *et al*., 2011; Burgaud *et al*., 2013; Coolen *et al*., 2013).

This study revealed presence of moderate and weak alkalitolerant fungi such as *Alternaria alternata, Penicillium* sp., *Cladosporium* sp. and *Fusarium* sp. previously reported to grow optimally at neutral or below neutral pH values. These species have previously appeared in existing reports on the alkalitolerant and halotolerant fungi (Kladwang *et al*., 2003; Gunde-Cimerman *et al*., 2009). They have therefore been considered as transition species in the alkaline environments, since they are also known to inhabit neutral soils worldwide. *Hypocreales* and *Pleosporaceae* have been reported

as strong alkalitolerants and effective alkaliphiles inhabiting soda soils at the edge of lake basins (Alexey *et al*., 2015). In this study, *Hypocreales* sp. lm 566 was identified in water samples at 81 ºC while *Pleosporales* sp. was found in wet sediment and water samples at 83.6 ºC and 81ºC respectively. Other interesting groups recovered from this study include *Pestalotiopsis* sp., *Neurospora* sp., and *Xylariaceae* sp. These have been reported to have various applications in Biotechnology industries (Russell *et al*., 2011, Roche *et al*., 2014; Healy *et al*., 2004; Posada *et al*., 2007).

## 5.8 Conclusion

This study presented fungal diversity analysis of samples collected from the hot springs of L. Magadi and Little Magadi, using Illumina sequencing technology. The results revealed representatives of thermophilic and alkaliphilic fungi within the hot springs, suggesting their ability to adapt to a multi-extreme sampling environment due to high pH, temperature, and salinity. Culture dependent and independent studies in future will help us unravel the survival mechanisms used by these polyextremophilic fungi.

# CHAPTER SIX

# METABOLIC POTENTIAL OF THE PROKARYOTIC COMMUNITIES WITHIN THE HOT SPRINGS OF LITTLE MAGADI IN KENYA

## 6.1 Introduction

The East African Rift Valley exhibits active volcanism with numerous hot springs on the shores of some lakes (Baumgarte, 2003). These hot springs are less alkaline and saline than the lakes they feed. However where mixing occurs complex pH, temperature and salinity gradients occur thus creating a range of soda lake habitats (Duckworth *et al.*, 1996). The chain of lakes occupying the rift floor depressions represent remnants of lakes of the pluvial period that occurred 10,000 years ago (Zavarzin *et al.*, 1999). The lakes are located in an area with an extremely large geothermal gradient and are connected with underground hydrotherms forming gas-steam vents, as in Lake Bogoria or thermal outlets in Lake Magadi. These environments have been shown to support a dense and diverse population of aerobic, organotropic, halophilic, alkaliphilic, and alkalitolerant representatives of major microbial phyla (Duckworth *et al.*, 1996), (Jones *et al.*, 1998), (Grant *et al.*, 1999), (Zavarzin *et al.*, 1999). Between these groups, there is cycling of carbon, sulfur, and nitrogen under aerobic and anaerobic conditions present in the lakes (Baumgarte, 2003).

Next-generation sequencing of PCR-amplified regions of 16S rRNA gene and reverse transcribed mRNA have been used to investigate the genes and gene expression levels of microbial communities in Ethiopian soda lakes (Lanzen *et al*., 2013). However, there's no evidence of similar work that has been conducted on microbial communities of Kenyan soda lakes and only a few limited high-throughput sequencing metagenomic studies have been reported (Akanga *et al*., 2015). A recent amplicon-based Illumina sequencing of 16S rRNA genes has provided an indication of the microbes present within the hot springs of Lake Magadi and Little Magadi and focused on the total versus active diversity (Kambura *et al*., 2016). However, a comprehensive study of

microbial activities and their function in the ecosystem has not been conducted. In the present study, the pool of metabolic genes in wet sediments, microbial mats and water samples collected from the hot spring (83.6 ℃) were analyzed. Analysis of the gene profiles and genetic diversity from samples collected in the hot springs could provide insight into the microbial ability within these extreme environments. In the present study, the pool of metabolic genes in wet sediments, microbial mats and water samples collected from 83.6 ℃ hot spring were analyzed.

## 6.2 Materials and methods

### 6.2.1 Research Authorization

Research authorization, Study site, measurements of physicochemical parameters sample collection and total RNA extraction were described in **chapter 3**.

### 6.2.2 cDNA synthesis from mRNA, library construction and deep sequencing

cDNA synthesis, library construction and short gun sequencing were performed at Molecular Research DNA Lab (www.mrdnalab.com, Shallowater, TX, USA). The quality of total RNA was assessed using gel electrophoresis. The extracted RNA was dissolved in RNase-free water and subsequently treated to remove DNA contaminants using the Amplification Grade DNase I Kit (Sigma, MO) according to the manufacturer’s instructions. cDNA first-strand and second-strand synthesis was done using the Superscript III First-Strand Synthesis SuperMix (Invitrogen, CA) and the Second-strand cDNA Synthesis Kit (BeyoTime, Jiangsu, China), respectively, following the manufacturer’s instructions. Single-strand reverse transcription was done to provide template for amplicon libraries using Superscript III (Invitrogen) according to the manufacturer’s protocol, random hexamer primed and subsequent RNAse H digestion. The double stranded cDNA synthesis was carried out as described by Urich et al. (2008). PCR enrichment was done to add the barcode to the cDNA library using phusion high fidelity taq polymerase enzyme (NEB, USA). The library concentration was determined

using Quant-iT dsDNA HS assay kit and a Qubit fluorometer (Invitrogen GmbH, Karlsruhe, Germany) as recommended by the manufacturer. The library was then concentrated using Ampure beads, eluted in 10µl TE buffer and used as template for Illumina Sequencing following manufacturer's instructions (Yu & Zhang, 2012). Sequencing was done at the Molecular Research DNA Lab on an Illumina HiSeq 2000 platform without amplification.

### 6.2.3 Data analysis

Data analysis was done at Computational Bioscience Research Center, King Abdullah University of Science and Technology in Saudi Arabia, on Dragon Metagenomic Analysis Platform (DMAP) using the Automatic Annotation of Microbial Genomes (AAMG) pipeline (Alam *et al*., 2013). Assemblies of the hot spring transcriptomes were generated using AAMG pipeline (Alam *et al*., 2013). After quality control, genome annotation for each of the transcriptomes was carried out as described by Alam *et al.* (2013). Prediction of Open Reading Frames (ORF) was done using Prodigal (Hyatt *et al*., 2010) and MetaGeneAnnotator (Noguchi *et al*., 2008). A series of similarity searches were performed to select optimal gene annotations using UniProt, NCBI's NR, NCBI's Conserved Domain Database (CDD), Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Interproscan. Filtered shotgun reads were subjected to taxonomic classification using BLASTn against SILVA SSU Reference 119 database (Quast *et al*., 2013). Assembly full-length rRNA contigs was carried out independently using shotgun reads from taxonomic groups as described previously (Radax *et al*., 2012).

## 6.3 Results

Based on temperature, the hot spring 83.6 ºC of Little Magadi was selected for metatranscriptome analysis as shown on **Figure 6.1**. The site pH was 9.4 while Dissolved Oxygen was 0.04 mg/l. The Total Dissolved Solids (mg/l) and Electrical Conductivity (mS/cm) were beyond the range of Electrical Chemical Analyzer that was used during sampling.

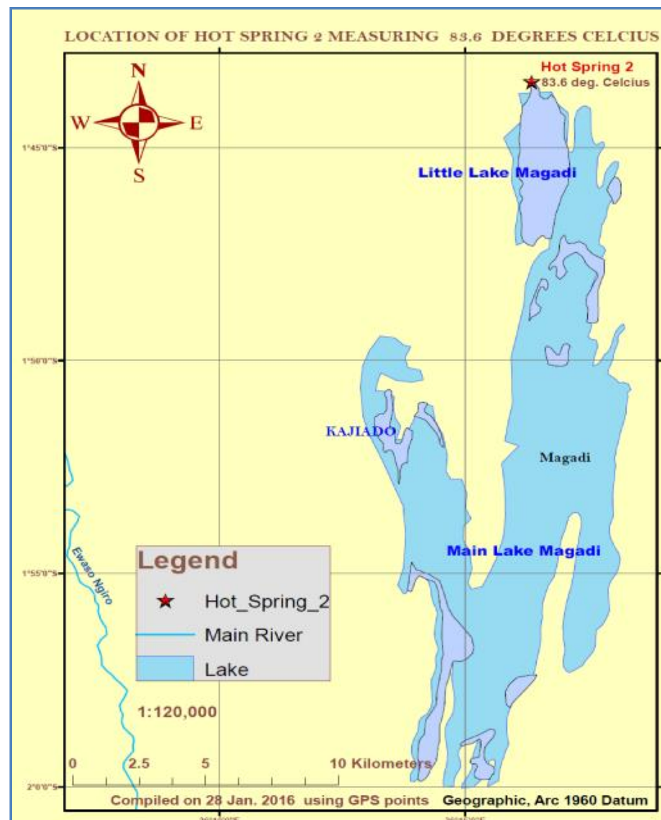


**Figure 6.1: Geographic map of 83.6 oC hot spring sampling site in Little Magadi**. The map was generated from the GPS coordinates captured from the lake during fieldwork using ArcGIS 10.3

### 6.3.1 Composition and diversity of microbial communities within the three sample types

The 16S rRNA gene fragments were extracted from the hot spring transcriptomes to obtain an initial overview of the microbial community based on BLASTn. Microbial taxonomic classification comprised cellular organisms, viruses and fungi across all samples. Cellular organisms included the super kingdom Bacteria and Archaea while viruses included unclassified phages, ssDNA viruses, dsDNA viruses, Retro-transcribing viruses, dsRNA viruses, environmental samples and ssRNA viruses. In sediments, mats and water samples, Archaea accounted for about 11 %, 1% and 6 % of their 16S rRNA fragments, respectively. The sediments and water samples were dominated by *Gammaproteobacteria* scoring 47 % and 37 %, respectively. The dominant class in mat samples was *Alphaproteobacteria* having scored 51 %. At genus level, *Halorhodospira halophila* (23 % in sediments), *Dinoroseobacter shibae* (13 % in mats) and; T*hioalkalivibrio sulfidophilus* and *Halorhodospira halophila* (16 % each in water) were the major communities in the hot spring transcriptomes.

The comparison of the three datasets within this study; 16S rDNA, 16S rRNA cDNA and mRNA cDNA results also indicated that *Proteobacteria* was the most dominant phyla across all samples (**Figure 6.2**). The highest relative abundance was 42.2 % in wet sediment within the mRNA dataset. There were representatives of major bacterial and archaeal phyla across the samples indicating that the prokaryotic diversity is real (**Figure 6.2**).

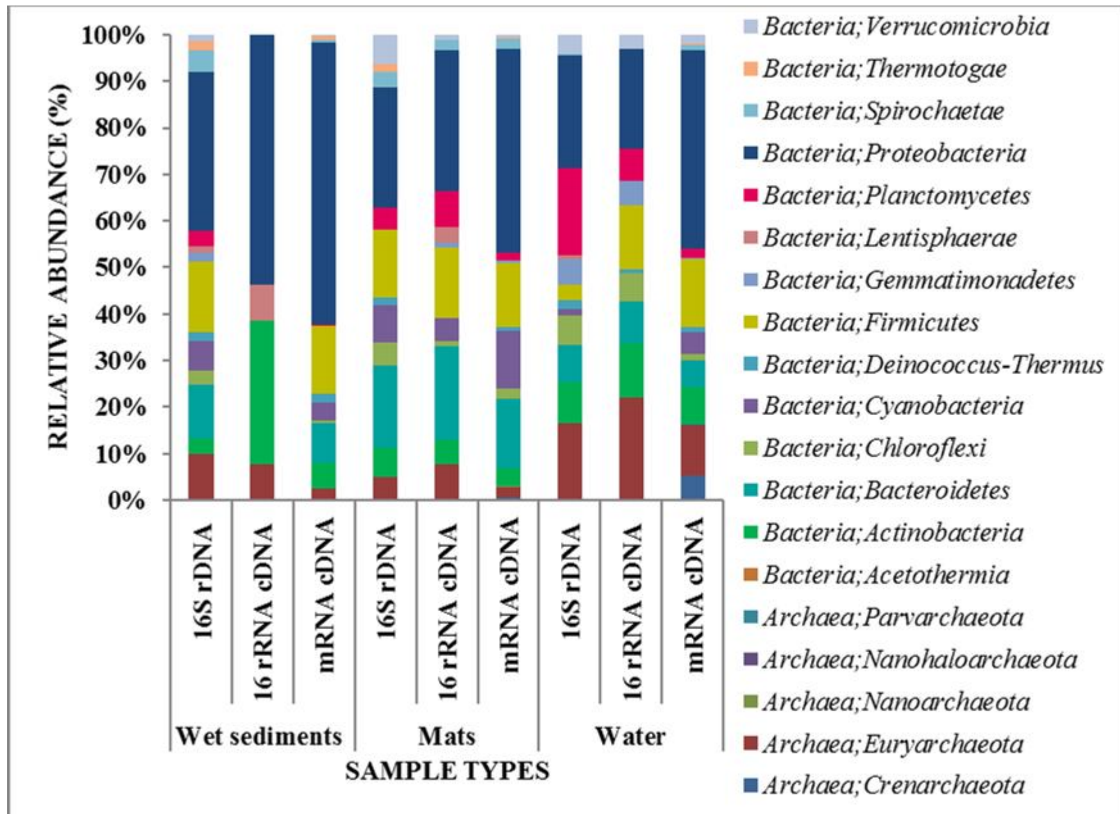


**Figure 6.2: Comparison of the most predominant prokaryotic taxa at phylum level within the 16S rDNA, 16S rRNA cDNA and mRNA cDNA.**

The most abundant phyla include: *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, *Chloroflexi*, *Deinococcus-thermus*, *Euryarchaeota, Nanoarchaeota* and *Crenarchaeota*.

Other than the dominant *proteobacteria* phylum, there was a significant manifestation of *Firmicutes*, *Actinobacteria* and *Bacteroidetes* that could be the major functional groups within the ecosystem. The comparison of prokaryotic taxa within the three sample types at class level (**Figure 6.3**).

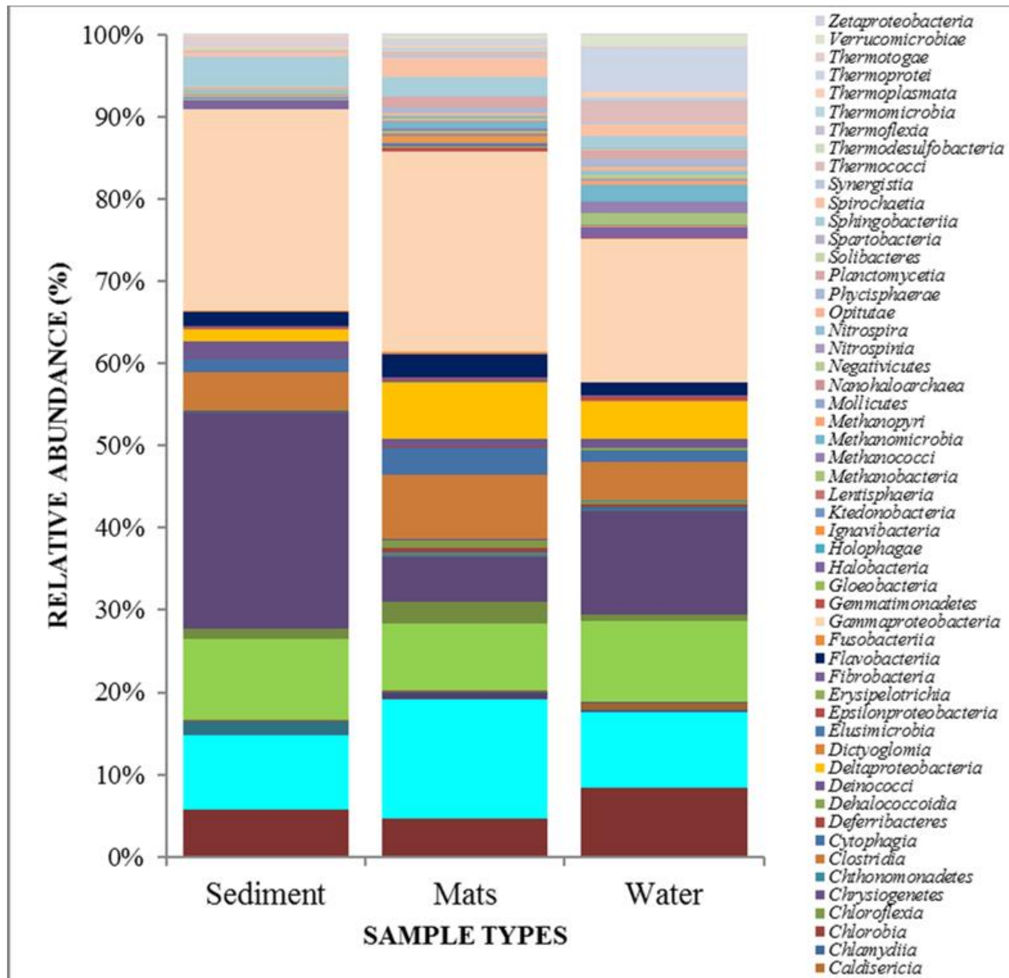


**Figure 6.3: Taxonomic classification of the most predominant prokaryotic groups at class level within the various samples collected from the hot springs of Little Magadi.** *Gammaproteobacteria* and *Alphaproteobacteria* are among the most predominant Proteobacteria phyla.

### 6.3.2 Bacterial taxonomic composition analysis

*Proteobacteria* was the most dominant bacterial phylum within the wet sediment, mats and water samples from the hot springs of Little Magadi. The other major bacterial phyla included *Firmicutes*, *Bacteroidetes*, *Cyanobacteria, Actinobacteria, Planctomycetes*, *Thermotogae*, *Deinococcus-thermus*, *Gemmatimonadetes* and *Aquificae* among others (**Figure 6.4**).

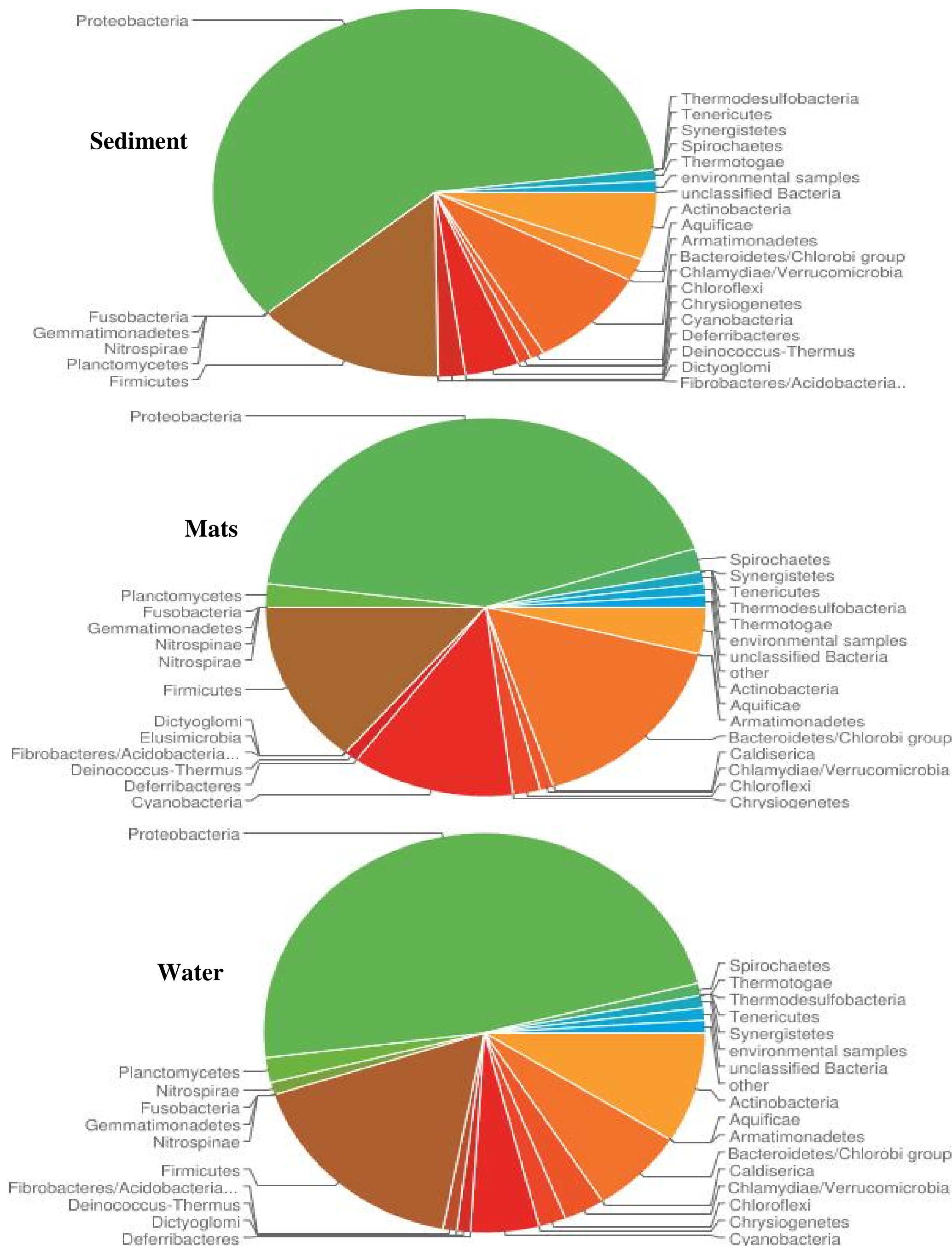


**Figure 6.4: Taxonomic classification of relatives of the most predominant bacterial phyla in various samples collected from the hot springs of Little Magadi**. Most abundant phyla included: *Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria, Cyanobacteria, Chloroflexi* and *Deinococcus-thermus*.

At species level, *Rhodothermus marinus, Bacillus azotoformans, Xanthomonas campestris*, *Thioalkalivibrio sulfidophilus, Halothece sp. PCC 7418* and *Dactylococcopsis salina* were the predominant of relatives of bacterial groups within the samples. The top ten functional classifications of relatives of bacterial species within the three sample types are shown in **Table 6.1**.

**Table 6.1: Top ten functional classifications of relatives of bacterial species within the three sample types collected from the hot springs of Little Magadi**

| **Sediment** | | **Mats** | | **Water** | |
|---|---|---|---|---|---|
| **Species (Blast)** | **Absolute counts** | **Species (Blast)** | **Absolute counts** | **Species (Blast)** | **Absolute counts** |
| *Xanthomonas campestris* | 435 | *Rhodothermus marinus* | 484 | *Xanthomonas campestris* | 254 |
| *Escherichia coli* | 139 | *Bacillus azotoformans* | 342 | *Bacillus azotoformans* | 204 |
| *Burkholderiales bacterium JOSHI_001* | 113 | *Thioalkalivibrio sulfidophilus* | 302 | *Propionibacterium acnes* | 135 |
| *Propionibacterium acnes* | 110 | *Halothece sp. PCC 7418* | 286 | *Escherichia coli* | 118 |
| *Bacillus cereus* | 102 | *Dactylococcopsis salina* | 240 | *Bacillus cereus* | 101 |
| *Methylibium petroleiphilum* | 101 | *Thiorhodovibrio sp. 970* | 181 | *Microcystis aeruginosa* | 89 |
| *Leptothrix cholodnii* | 81 | *Salinibacter ruber* | 178 | *Burkholderiales bacterium JOSHI_001* | 74 |
| *Meiothermus silvanus* | 76 | *uncultured bacterium* | 141 | *uncultured bacterium* | 64 |
| *Agrobacterium tumefaciens* | 69 | *Thiocapsa marina* | 134 | *Phycisphaera mikurensis* | 61 |
| *Hydrogenobacter thermophilus* | 55 | *Thioflavicoccus mobilis* | 130 | *Leptothrix cholodnii* | 60 |

### 6.3.3 Archaeal taxonomic composition

In this study, the most predominant archaeal phylum within the three sample types was found to be *Euryarchaeota*. The other major phyla included *Crenarchaeota*, *Thaumarchaeota*, *Korarchaeota*, *Nanoarchaeota*, *Nanohaloarchaeota*, *Parvarchaeota* and *unclassified archaea* (**Figure 6.5**).

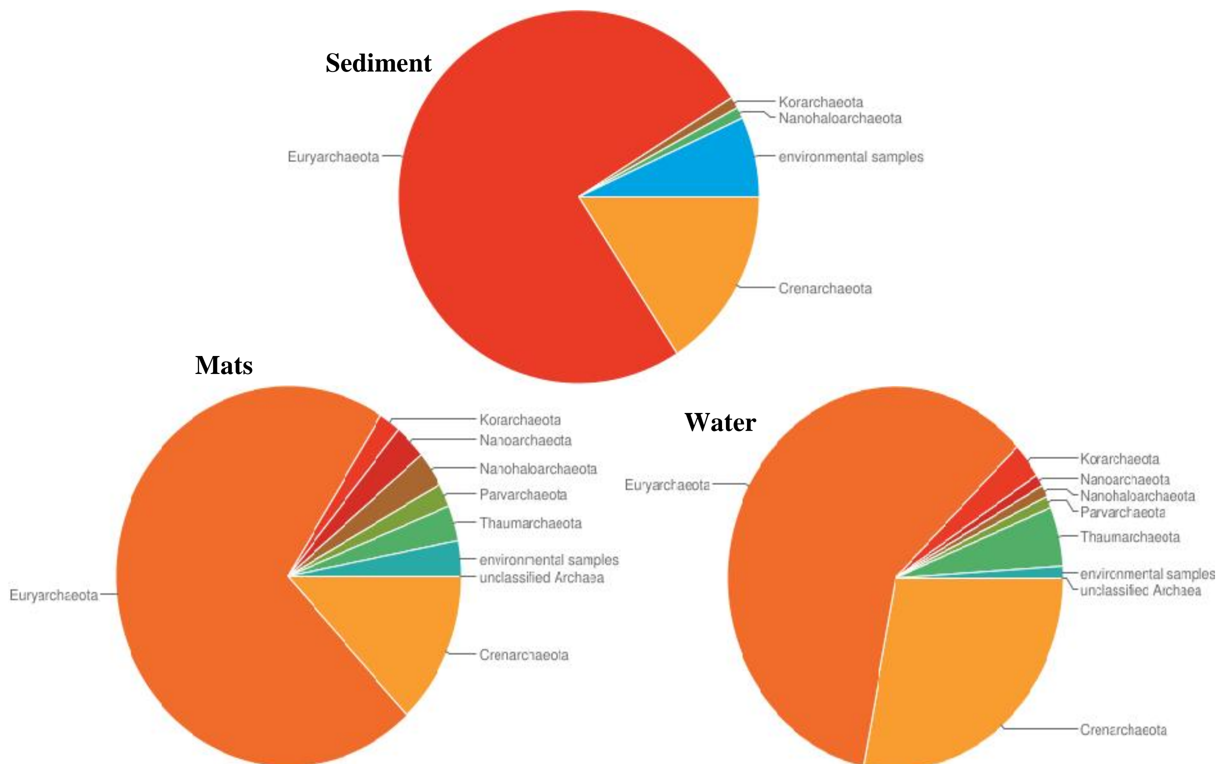


**Figure 6.5: Taxonomic classification of the most predominant relatives of archaeal phyla in various samples collected from the hot springs of Little Magadi.** Most abundant phyla include: *Euryarchaeota*, *Crenarchaeota*, *Thaumarchaeota*, *Korarchaeota* and *Nanoarchaeota.*

At species level, sediment samples were dominated by relatives of *uncultured archaeon, Halorhabdus utahensis* and *Methanopyrus kandleri. Candidatus Korarchaeum cryptofilum, Thermofilum pendens* and *Methanothermus fervidus* were the dominant archaea groups in water samples while mat samples were dominated by *Thermoplasmatales archaeon SCGC AB-540-F20, Thermoplasmatales archaeon SCGC AB-539-N05* and *Candidatus Nanosalinarum sp. J07AB56* among others. The top ten functional classifications of relatives of archaeal species within the three sample types (**Table 6.2**).

**Table 6.2: Top ten functional classifications of relatives of archaeal species within the three sample types collected from the hot springs of Little Magadi**

| **Sediment** | | **Mats** | | **Water** | |
|---|---|---|---|---|---|
| **Species (Blast)** | **Absolute Count** | **Species (Blast)** | **Absolute Count** | **Species (Blast)** | **Absolute Count** |
| *uncultured archaeon* | 8 | *Thermoplasmatales archaeon SCGC AB-540-F20* | 14 | *Candidatus Korarchaeum cryptofilum* | 52 |
| *Halorhabdus utahensis* | 7 | *Thermoplasmatales archaeon SCGC AB-539-N05* | 9 | *Thermofilum pendens* | 42 |
| *Methanopyrus kandleri* | 4 | *Candidatus Nanosalinarum sp. J07AB56* | 8 | *Methanothermus fervidus* | 41 |
| *Methanothermobacter thermautotrophicus* | 4 | *Thermococcus barophilus* | 8 | *Methanopyrus kandleri* | 35 |
| *Thermococcus litoralis* | 4 | *nanoarchaeote Nst1* | 8 | *Pyrococcus horikoshii* | 30 |
| *Thermococcus sibiricus* | 4 | *uncultured archaeon* | 8 | *Hyperthermus butylicus* | 29 |
| *Thermofilum pendens* | 4 | *Candidatus Korarchaeum cryptofilum* | 7 | *Candidatus Nitrososphaera gargensis* | 25 |
| *Archaeoglobus fulgidus* | 3 | *Methanopyrus kandleri* | 7 | *Pyrococcus furiosus* | 25 |
| *Halalkalicoccus jeotgali* | 3 | *Methanothermus fervidus* | 7 | *Pyrolobus fumarii* | 25 |
| *Halobacterium salinarum* | 3 | *Methanotorris igneus* | 7 | *Archaeoglobus fulgidus* | 24 |

### 6.3.4 Viral taxonomic composition

In this study, some lineages of viruses were obtained from the mRNA transcripts in various samples collected from the hot springs of Little Magadi. These include *ssDNA viruses*, *ssRNA viruses*, *dsDNA viruses*, *Retro-transcribing viruses*, *unclassified phages*, *unclassified viruses* and *unassigned viruses* (**Figure 6.5).** At species level, *Geobacillus virus E2, Bathycoccus sp. RCC1105 virus BpV, Deep-sea thermophilic phage D6E*, *Cafeteria roenbergensis virus BV-PW1, Lactococcus phage 949* and *Cotesia congregata bracovirus* were the predominant neighbors of viral groups within the samples. The top ten neighbors of viral species functional classifications within the three sample types (**Table 6.3**).

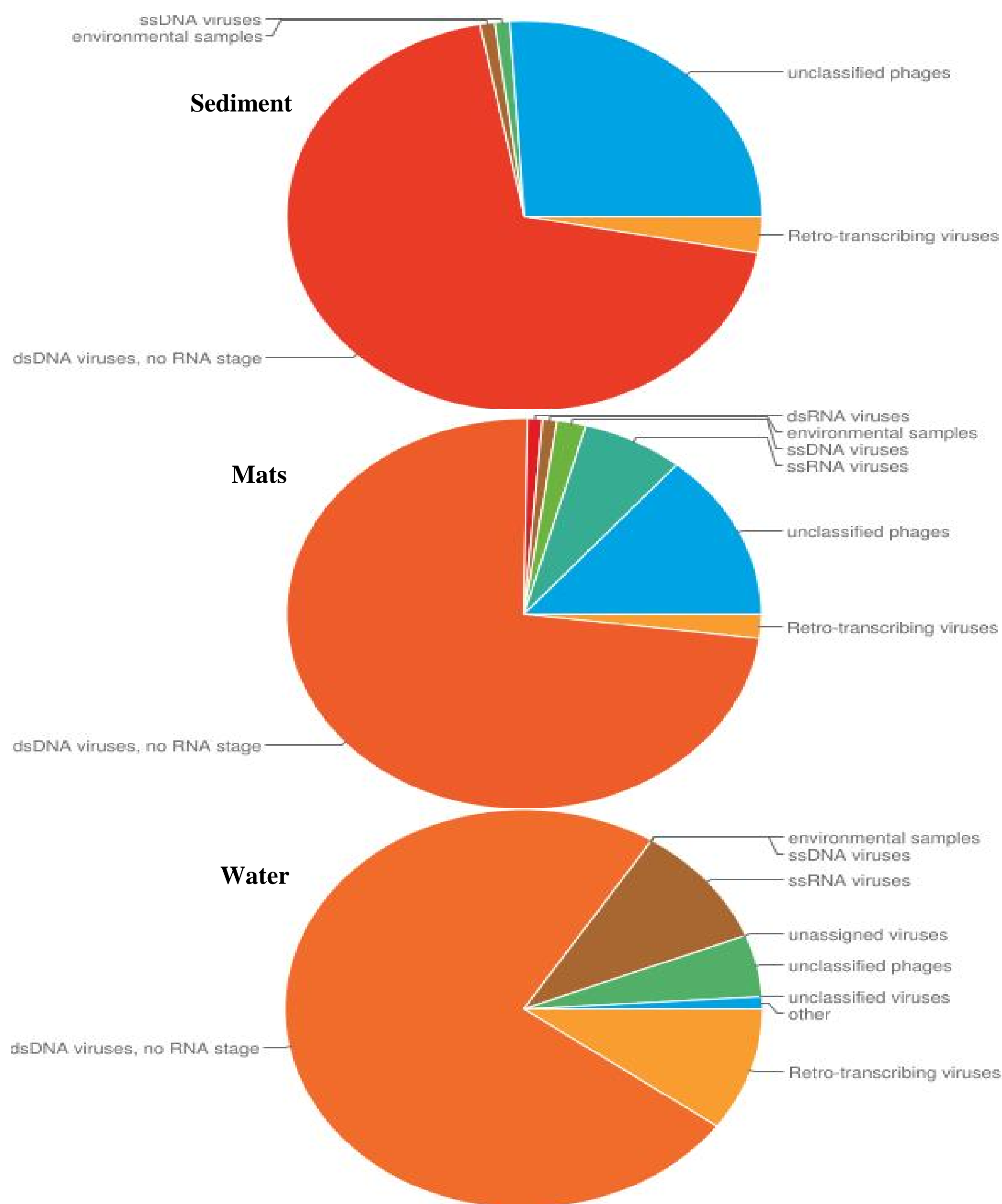


**Figure 6.6: Taxonomic classification of the most predominant viral groups in various samples collected from the hot springs of Little Magadi.** Unique viral groups include *unclassified phages, Retro-transcribing viruses, dsNDA viruses and unassigned viruses*

**Table 6.3:** Top ten functional classifications of relatives of viral species within the three sample types collected from the hot springs of Little Magadi

| **Sediment** | | **Mats** | | **Water** | |
|---|---|---|---|---|---|
| **Species (Blast)** | **Absolute Count** | **Species (Blast)** | **Absolute Count** | **Species (Blast)** | **Absolute Count** |
| *Geobacillus virus E2* | 45 | *Geobacillus virus E2* | 19 | *Bathycoccus sp. RCC1105 virus BpV* | 41 |
| *Deep-sea thermophilic phage D6E* | 24 | *Lactococcus phage 949* | 13 | *Cafeteria roenbergensis virus BV-PW1* | 36 |
| *Lactococcus phage 949* | 20 | *Deep-sea thermophilic phage D6E* | 11 | *Bovine viral diarrhea virus 1* | 33 |
| *Bacillus phage 0305phi8-36* | 11 | *Bacillus phage SIOphi* | 4 | *Geobacillus virus E2* | 24 |
| *Bacillus phage SIOphi* | 8 | *Bovine viral diarrhea virus 2* | 4 | *Bovine viral diarrhea virus 2* | 21 |
| *Clostridium phage c-st* | 5 | *Chaetoceros tenuissimus RNA virus 01* | 4 | *Cotesia congregata bracovirus* | 20 |
| *Bacillus phage G* | 4 | *Bacillus phage 0305phi8-36* | 3 | *Diachasmimorpha longicaudata entomopoxvirus* | 19 |
| *Bacillus phage phiNIT1* | 4 | *Bathycoccus sp. RCC1105 virus BpV* | 3 | *Moloney murine sarcoma virus* | 19 |
| *Bacillus phage vB_BceM-Bc431v3* | 4 | *Cannes 8 virus* | 3 | *Pandoravirus salinus* | 19 |
| *Clostridium phage D-1873* | 4 | *Enterobacteria phage phiX174 sensu lato* | 3 | *Deep-sea thermophilic phage D6E* | 17 |

### 6.3.5 Fungal taxonomic composition

The dataset revealed sixty one sequences that were affiliated to eukaryotic groups. Nucleotide blast against sequences on NCBI database showed twenty six of them to be related to the fungal phyla. Some of their neighbors included uncultured groups of *Hypocreales sp*, *Aspergillus oryzae*, *Harpochytrium sp*., *Acremonium sp*., and *Cladosporium sp*. among others, some of which had been observed in the DNA dataset. There were no common nucleotide sites found within the obtained sequences using Mega software, hence the groups could not be phylogenetically compared.

### 6.3.6 Categories of Clusters of Orthologous Groups (COG) in the three transcriptomes

COGs in the reads were sorted into categories. The sediment samples had COGs in the following categories: metabolism 39 %, cellular processes and signaling 24 %, information storage and processing 19 % and poorly characterized COGs scored 18 %. The mat samples COGs included: metabolism 37 %, cellular processes and signaling 25 %, information storage and processing 21 % and poorly characterized COGs scored 17 %. The water samples COGs included: metabolism 29 %, cellular processes and signaling 31 %, information storage and processing 21 % and poorly characterized COGs scored 15 %. Metabolism category included energy production and conversion, inorganic ion transport and metabolism, coenzyme transport and metabolism, nucleotide transport and metabolism, lipid transport and metabolism, secondary metabolites biosynthesis, transport and catabolism, amino acid transport and metabolism; and carbohydrate transport and metabolism. The cellular processes and signaling category included cell wall/membrane/envelope biogenesis, signal transduction mechanisms, intracellular trafficking, secretion, and vesicular transport, post-translational modification, protein turnover, chaperones, cell motility and defense mechanisms. Information storage and processing category included; transcription, translation, ribosomal structure and biogenesis, replication, recombination and repair

while the poorly characterized category comprised general function prediction and unknown function.

### 6.3.7 Diversity of genes and pathways between samples

Sixty nine percent (69 %) of genes in the wet sediment samples were affiliated to metabolic pathways. These include pathways that participate in the metabolism of carbohydrates, carbon, amino acids, energy, cofactors and vitamins, lipids, nucleotides, oxocarboxylic acids, terpenoids and polyketides, biosynthesis of amino acids and other secondary metabolites, biodegradation and metabolism of xenobiotics, glycan biosynthesis and metabolism and degradation of aromatic compounds. Ten percent (10 %) of the genes were affiliated to environmental information processing which include signal transduction, membrane transport, signaling molecules and interaction. Eight percent (8 %) of the genes were affiliated to genetic information processing which include transcription, translation, replication and repair, Folding, sorting and degradation. The rest of the genes were affiliated to cellular processes (3 %), organismal systems (4 %) and human diseases (6 %). Seventy percent (70 %) of genes in the mat samples were affiliated to metabolic pathways. The others included genetic information processing (9 %), environmental information processing (8 %), cellular processes (3 %), organismal systems (4 %) and human diseases (6 %). Sixty one percent (61 %) of genes in the water samples were affiliated to metabolic pathways. The others included genetic information processing (18 %), environmental information processing (6 %), cellular processes (2 %), organismal systems (4 %) and human diseases (9 %). A heat map of gene profiles grouped the metabolic pathways into five gene clusters (I-V) as shown in **Figure 6.7** and **6.8**.

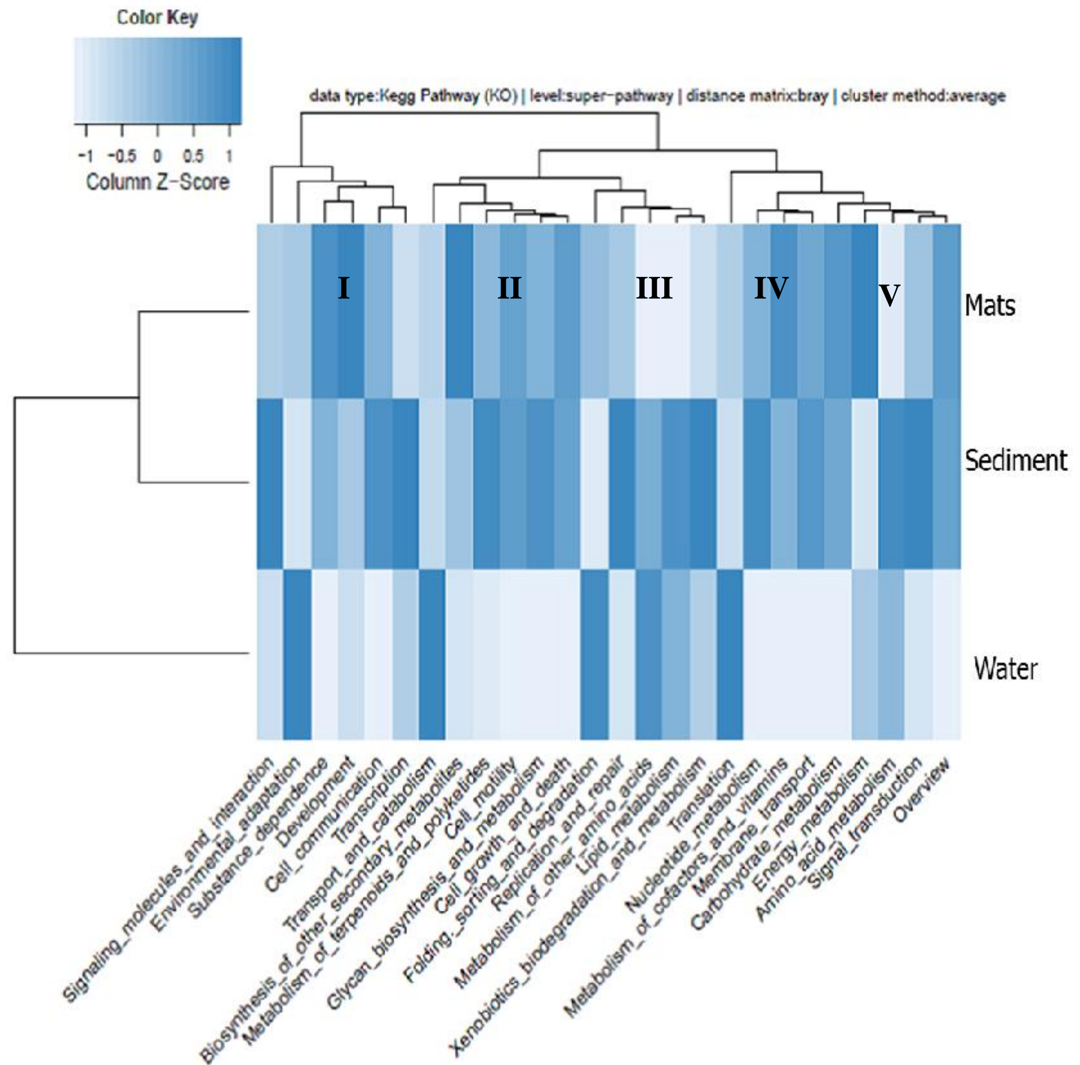


**Figure 6.7: Sorting of obtained gene clusters into pathways. Genes present in the five regions of the heat map (I-V) were assigned to different KEGG pathways.**

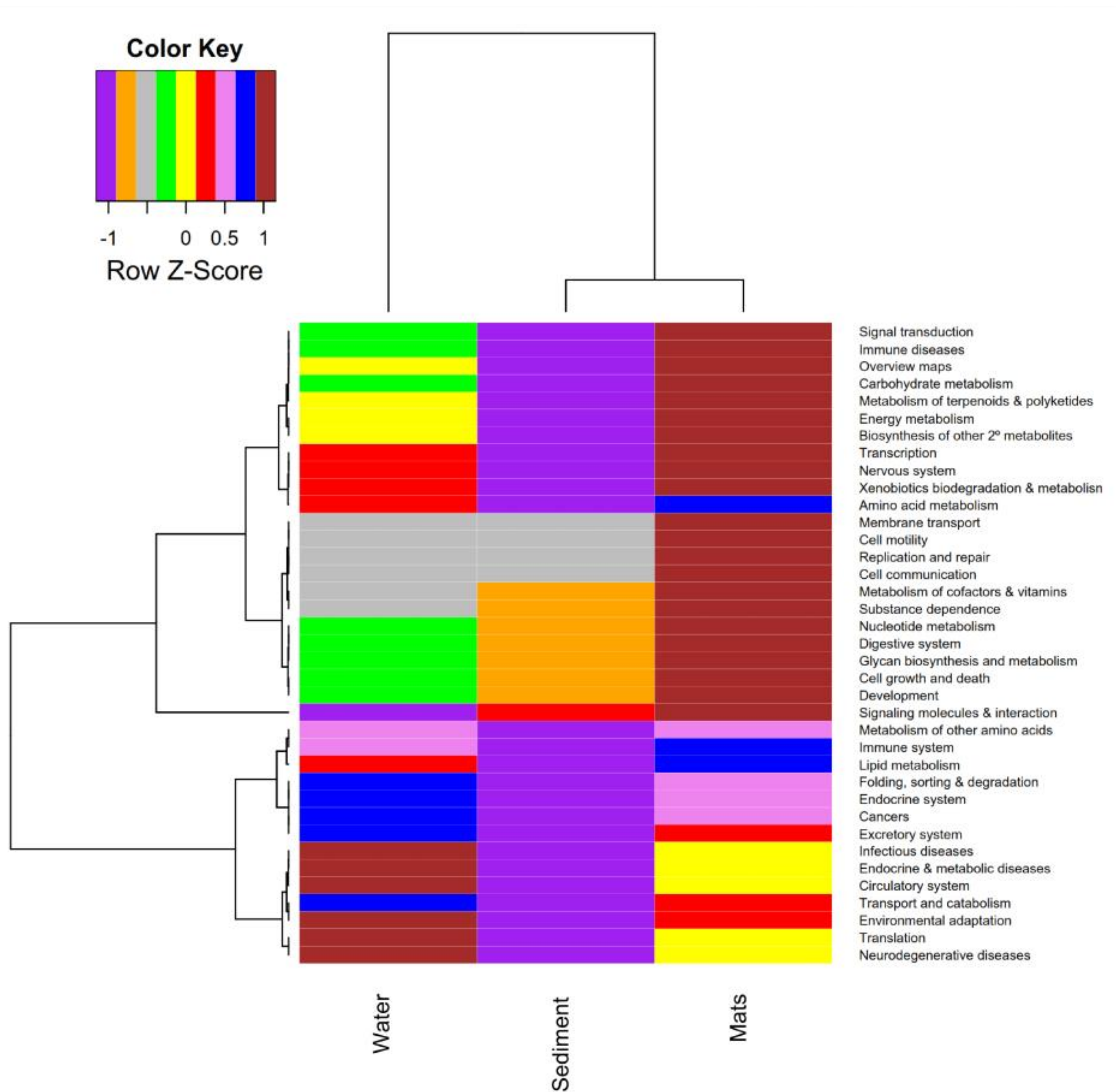


**Figure 6.8: A heatmap showing gene clustering of obtained metabolic pathways according to their distribution within the samples. Mat samples were found to harbor most diverse genes**

Analysis of KEGG pathways Enzyme Commission (EC) between the samples indicated mat samples to have the highest abundance of pathways with dominant pathways being xenobiotics biodegradation and metabolism, and biosynthesis of other secondary metabolites. This was followed by sediment samples while translation pathway scored the least abundance across all samples. Other major pathways include metabolism of carbohydrates, energy, terpenoids and polyketides, cofactors and vitamins as well as biosynthesis and metabolism of glycan. The EC super pathways obtained are shown in **Figure 6.9**. Sugar utilization in the samples was inferred from starch and sucrose metabolism pathway, in which carbohydrates and derivatives such as starch, sucrose, isomaltose, trehalose, and cellulose among others were potential sugar sources based on level three of Enzyme Commission (EC) numbers 5.1.3. There was representation of carotenoid biosynthesis and oxidative phosphorylation in mat samples showing an absolute abundance of 312 and 254 counts followed by water samples that scored an absolute abundance of 181 and 174 counts respectively. Although amino acid metabolism has relatively high absolute abundant, the translation pathway is shown to score very low abundance. This could be attributed to the fact that amino acids occur as isomers, several of which could code for biosynthesis of the same protein.

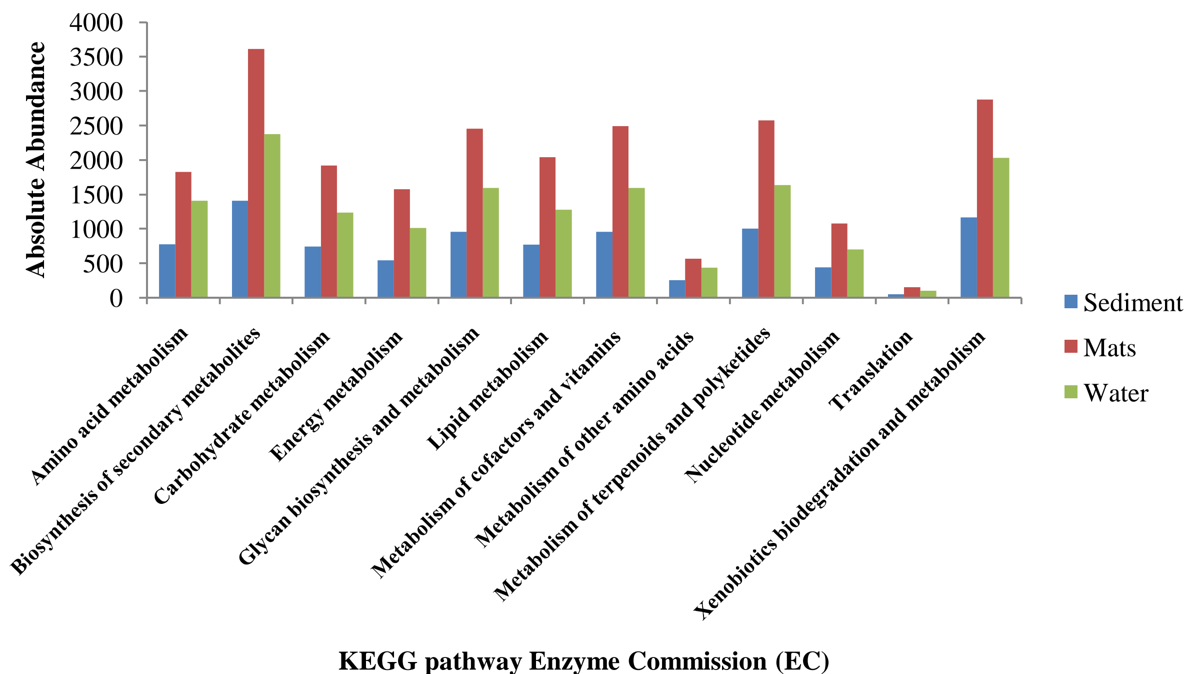


**Figure 6.9: Sorting of sample-specific KEGG pathways Enzyme Commission (EC) between the samples.** Microbial mat samples were shown to score the highest absolute abundance as compared to water and sediments.

### 6.3.8 Gene Ontology

The Gene Ontology functional classification revealed three gene categories namely biological process, cellular component and molecular function. Some of the biological processes included metabolism, reproduction, growth, response to stimulus, localization, biological regulation, multicellular and single-organism process while the major cellular components included macromolecular complex, cell, membrane, organelle and virion (**Figure 6.10).**

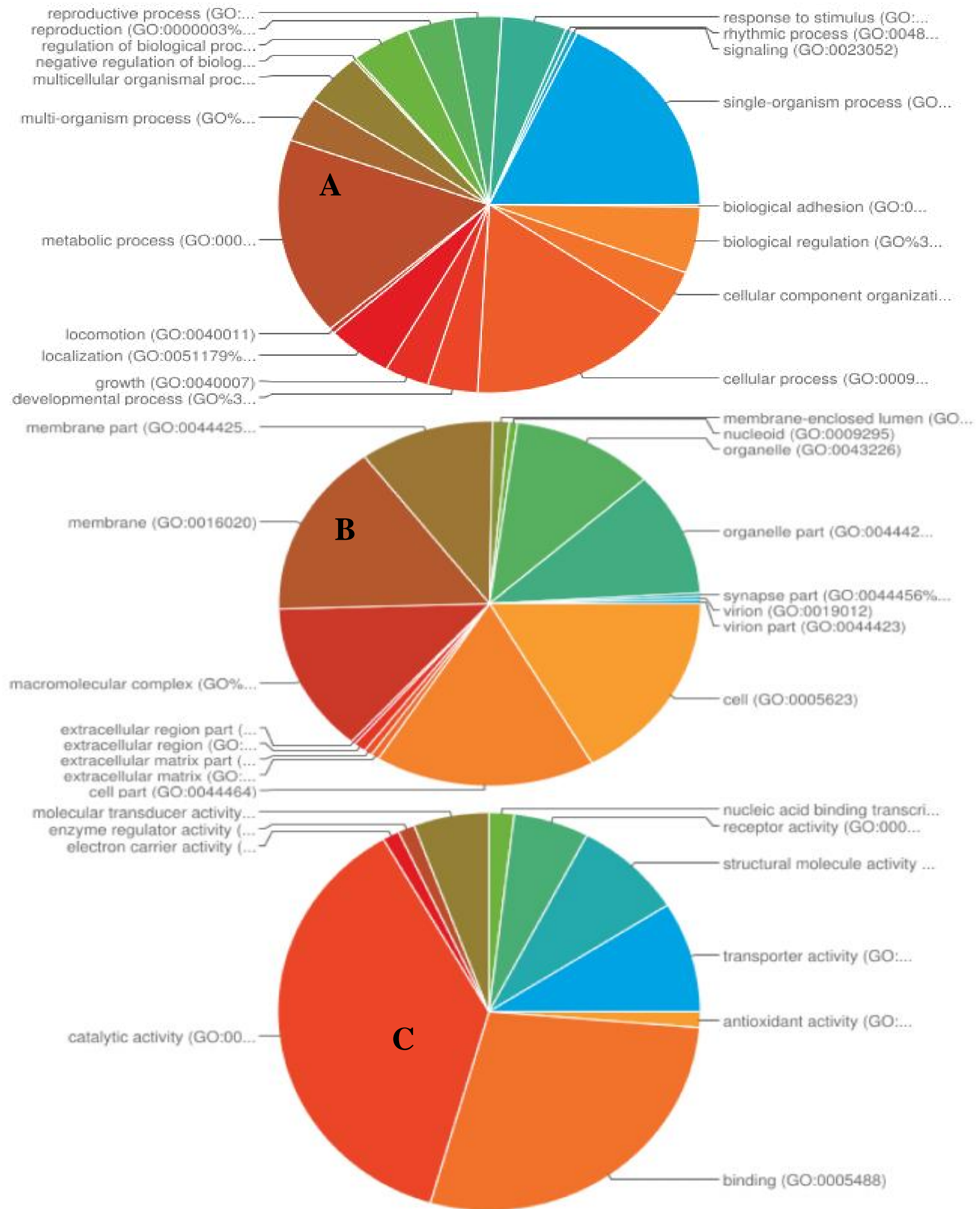


**Figure 6.10: Summary of Gene Ontology functional classification of the samples collected from the hot springs of Little Magadi.** A, B and C represent; biological processes, cellular components and molecular function, respectively.

Further classification of Gene ontology at molecular function level 2 revealed four clusters, most of which were harbored in mats and sediment samples. These are shown on **Figure 6.11**.

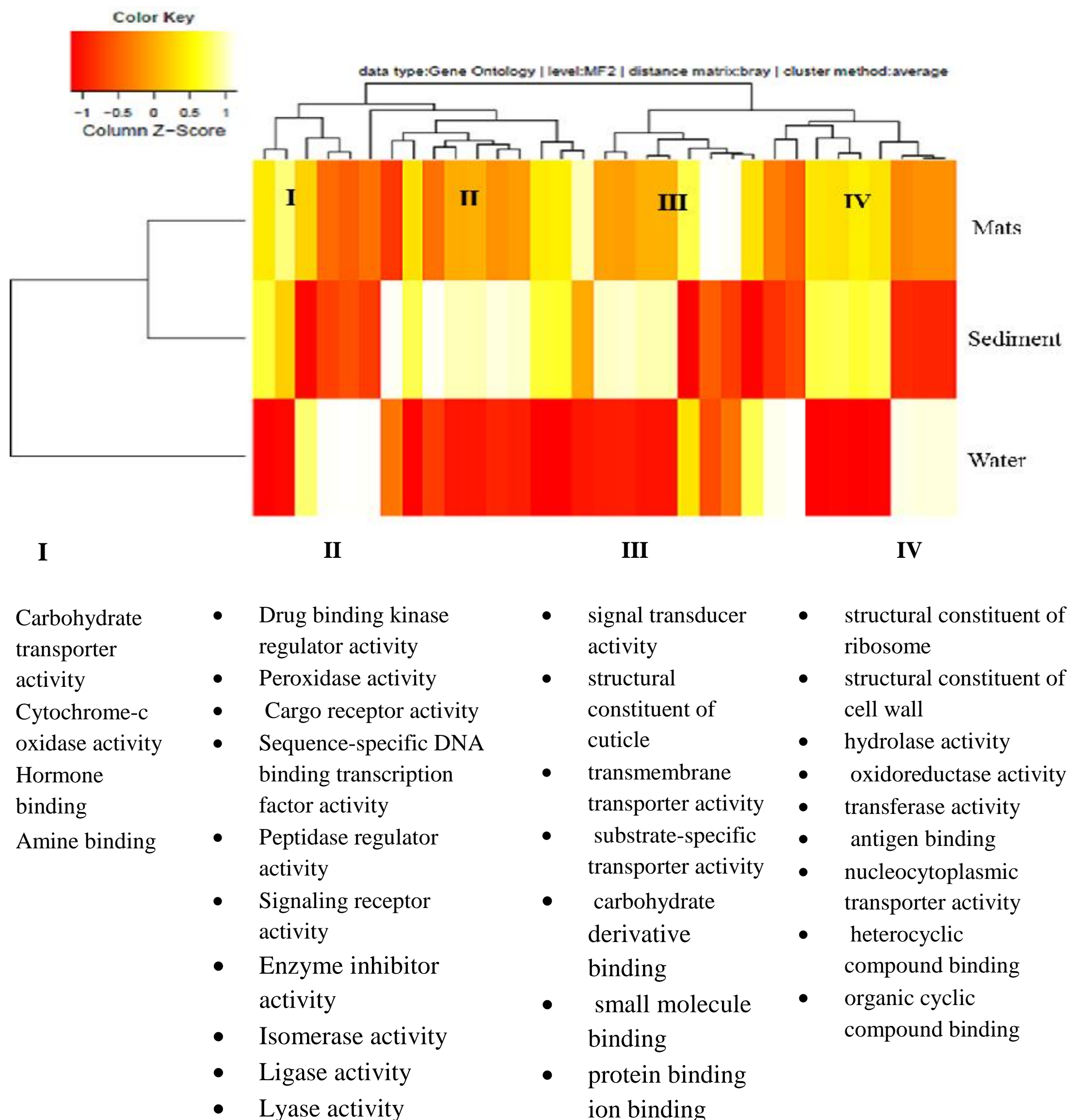


**Figure 6.11: Comparison of gene ontology (molecular function level 2) between the samples collected from the hot springs of Little Magadi**

### 6.3.9 Metacyc pathways

The major metacyc pathways obtained included biosynthesis of energy, lipids, fatty acids, secondary metabolites and co-factors, degradation / utilization / assimilation, detoxification, generation of precursor metabolites and energy, signal transduction pathways, superpathways, transport and tRNA processing pathway (**Figure 6.12**). Signal transduction and tRNA processing pathways scored zero hits across all samples.

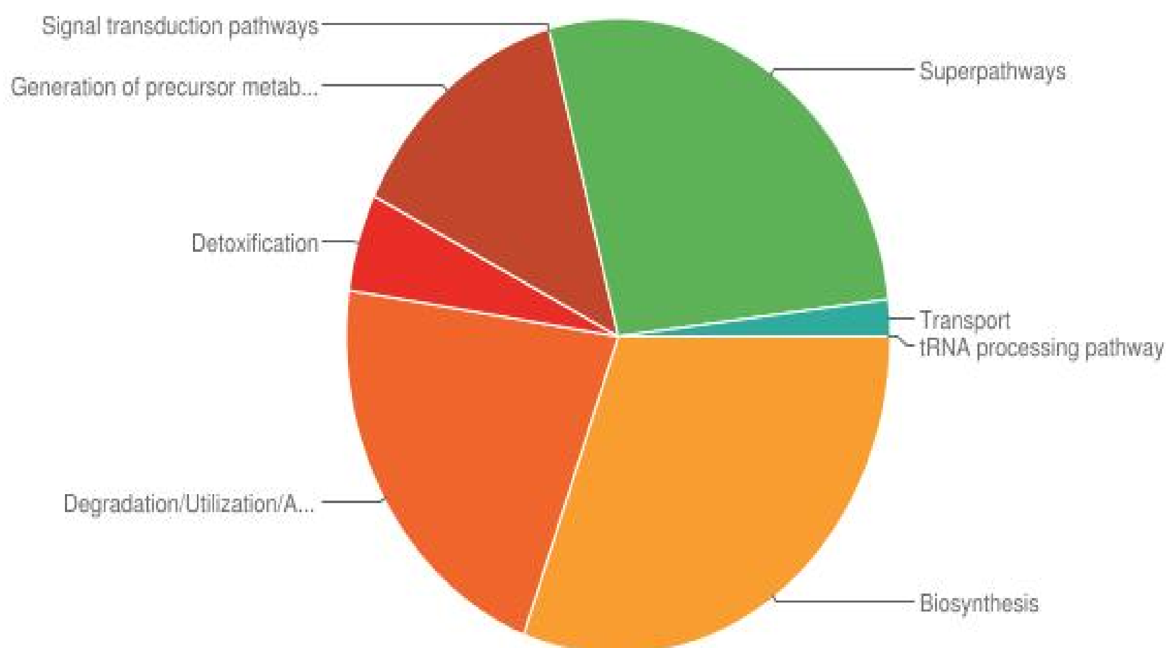


**Figure 6.12: Summary of metacyc pathways in the samples collected from the hot springs of Little Magadi**

The results indicated that mat samples harbored the highest abundance of metabolic pathways as compared to sediments and water samples (**Figure 6.13**). The least scoring pathways included acid resistance, aldehyde degradation, arsenate detoxification and protein degradation among others. Transport pathways identified within the hot spring samples include copper transport I, calcium transport I, calcium transport II, calmodulin regulated calcium transport, nucleobase ascorbate transport I and starch plastidic translocation I. Anaerobic aromatic compound degradation (*Thauera aromatica*) and aerobic respiration-electron donor III pathways also scored high abundance across all samples (**Figure 6.13**).

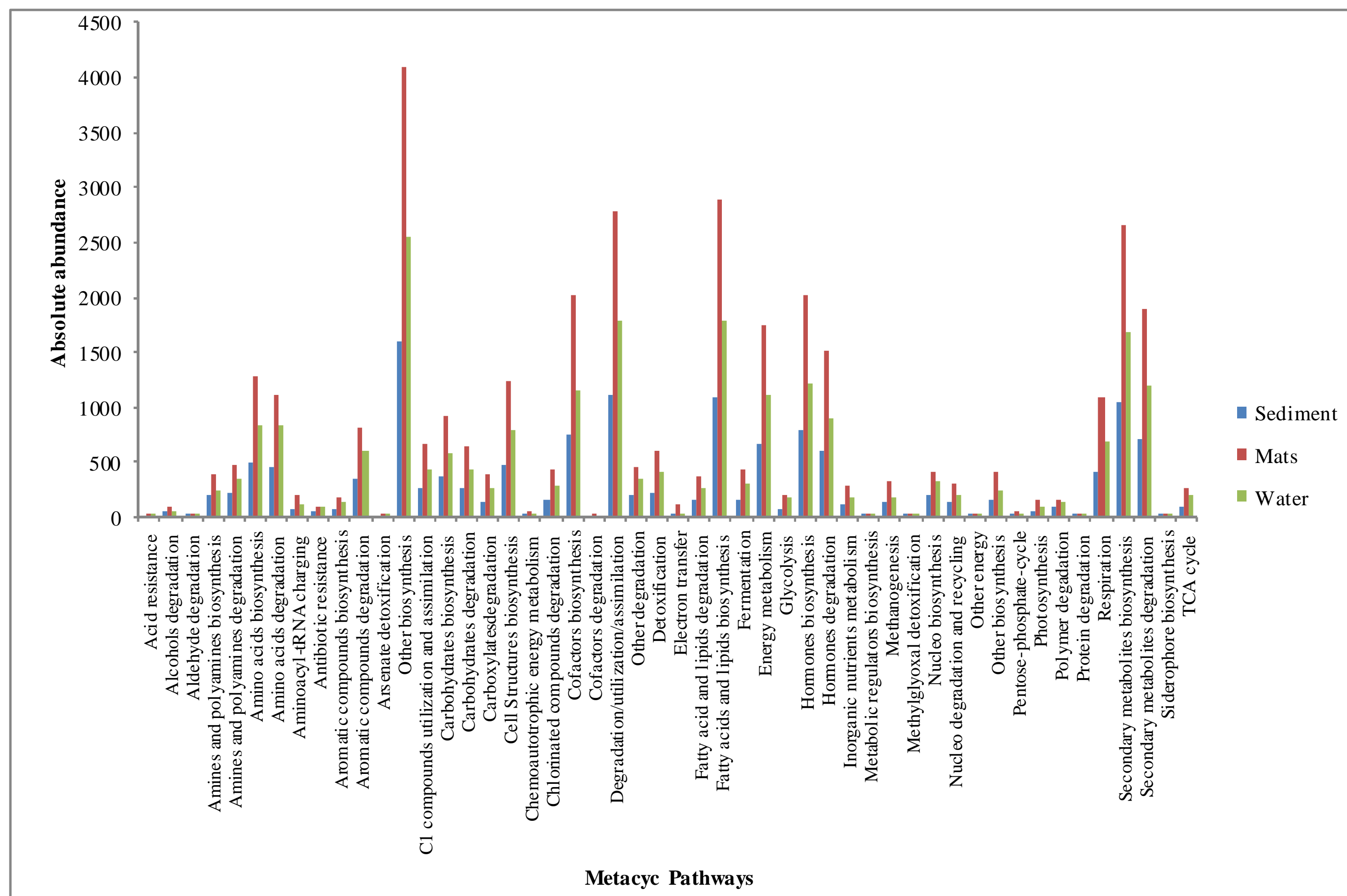


**Figure 6.13: Comparison of metacyc pathways between the samples collected from the hot springs of Little Magadi.** Microbial mat samples were shown to score the highest absolute abundance across all samples.

### 6.3.10 Functional classification of enzymes and their application in biotechnology

The analysis of enzyme groups obtained indicated the presence of the six groups of enzymes. These include: EC 1: oxidoreductases, EC 2: transferases, EC 3: hydrolases, EC 4: lyases, EC 5: isomerases and EC 6: ligases Enzyme comparison indicated the mat samples to contain high levels of Oxidoreductases, Ligases and Transferases while sediment samples contained high levels of Lyases and water samples had high levels of Hydrolases as shown in **Figure 6.14**. The average enzyme clustering based on bray curtis dissimilarity at level one (1) grouped them into two clusters. The dendogram indicated the mat samples to contain high levels of Oxidoreductases, Ligases and Transferases while sediment samples contained high levels of Lyases and water samples had high levels of Hydrolases (**Figure 6.14**).

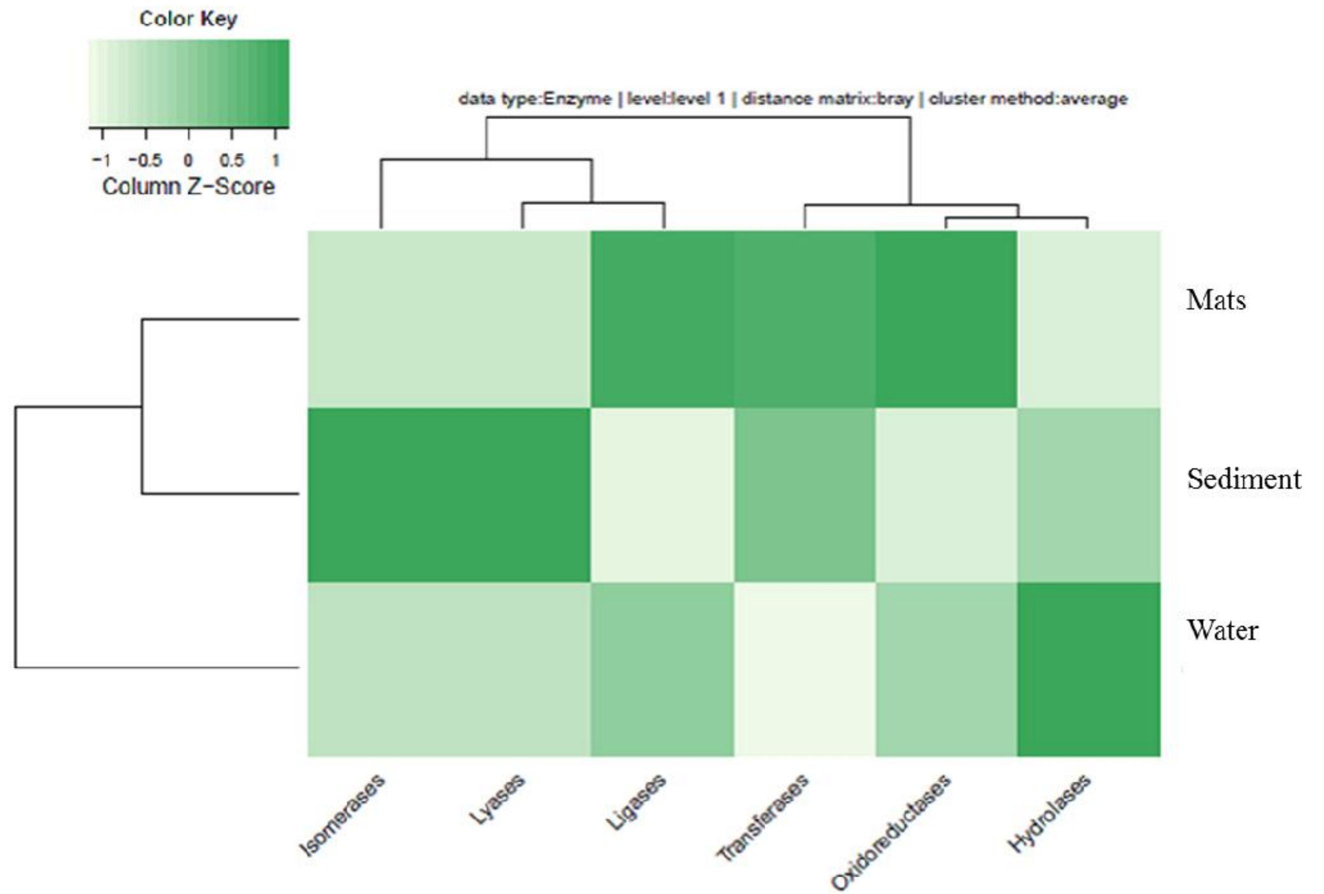


**Figure 6.14: Clustering of the six enzyme groups between the samples collected from the hot springs of Little Magadi**

**Table 6.4: Absolute abundance of enzymes revealed within samples**

| Enzyme Category (Class) | Industrial applications | Absolute abundance | | |
|---|---|---|---|---|
| | | Wet Sediment | Mats | Water |
| **Hydrolases (3.-.-.-)** | Pharmaceutical | 418 | 997 | 790 |
| **Isomerases (5.-.-.-)** | Catalyze biosynthetic reactions by: | 80 | 183 | 118 |
| **Ligases (6.-.-.-)** | - Addition of groups<br>- Rearrangement of atoms | 165 | 459 | 285 |
| **Lyases (4.-.-.-)** | - Join molecules together with covalent bonds | 136 | 342 | 220 |
| **Oxidoreductases (1.-.-.-)** | Catalyze biological oxidation/reduction reactions | 389 | 1030 | 650 |
| **Transferases (2.-.-.-)** | - label DNA<br>- produce plasmid vectors<br>- create biosensors | 518 | 1393 | 764 |

## 6.4 Discussion

### 6.4.1 Active microbial diversity

The comparison of prokaryotic diversity within 16S rDNA, 16S rRNA cDNA and mRNA cDNA datasets indicated *Proteobacteria* as the most dominant phyla across all samples. The dominant taxa related to those obtained from a previous research on deep sea and marine sediments community composition (DeLong *et al*., 2006; Brown *et al*., 2009; Liao *et al*., 2011). The groups are also similar to the groups obtained from the deep-sea brine sediment samples collected from the Red Sea, which indicated that

proteobacteria are well adapted to extreme conditions and could be responsible for various functional processes within the ecosystem (Yong *et al*., 2013). Hyperthermophillic communities have been found to be complex systems of primary producers and decomposers of organic matter. All hyperthermophillic primary producers are chemolithoautotrophs (i.e., sulfur oxidizers, sulfur reducers, and methanogens) (Vieille & Zeikus, 2001).

Most of the observed taxa are relatives to organisms that are unique to the high temperature and saline environments such as the *Rhodothermus marinus* which is an example of obligately aerobic, halophilic, thermophilic, Gram-negative bacterial inhabiting the hot springs (Bjornsdottir *et al*., 2006). *Thioalkalivibrio sulfidophilus* denotes obligately chemolithoautotrophic, haloalkaliphilic sulfur-oxidizing bacterial groups that play a key role in primary production within the ecosystem (Muyzer *et al*., 2011). *Bacillus species* are remarkably diverse bacterial groups, capable of growth within diverse environments. Comparative genomic analyses have also revealed that members of this genus exhibit considerable genome diversity (Earl *et al*., 2008); hence their ability to survive in the highly alkaline saline hot springs of Little Magadi. *Thermotogales* and *Aquificales* have been described in deep-sea hydrothermal environments (Stetter, 1996). Previous studies have shown archaea from hyperthermophillic environments to be composed of three branches: *Crenarchaeota, Euryarchaeota* and *Korarchaeota* (Barns *et al*., 1996). *Crenarchaeota* include halophiles while among the *Euryarchaeota*, methanogens have mesophilic relatives.

Microbial populations have various associations within their habitats. Some viruses are common in the environment especially bacteriophages that reside within bacterial hosts. There was presence of some mammalian viruses like *Bovine viral diarrhea virus 1* and *Moloney murine sarcoma virus*. These could be contaminants from livestock and other wildlife or could have been dispersed through erosion from previous rainfall and later found suitable hosts within the hot spring. The viral lineages obtained in this study could be responsible for a significant proportion of microbial mortality and thus have a

profound influence on carbon and other nutrient cycles as previously reported by Suttle**,** (2007). The viruses could be important vehicles for lateral gene transfer through lysogeny and transduction; and probably promote microbial diversity by preferentially lysing the most abundant species (Weinbauer & Rassoulzadegan, 2004). These could also have a role in lateral gene transfer and non-orthologous gene replacement in cellular genomes. They may have played critical roles in the evolution of DNA and DNA replication mechanisms and separation of the three domains of life within the hot spring ecosystem (Filee, *et al***.,** 2003; Forterre, 2006).

### 6.4.2 Diversity of genes and pathways

Some of the gene clusters revealed were affiliated to carbon fixation and methane metabolism pathways, demonstrating a high level of totality of different genes contained within the organisms present in the hot spring ecosystem. This was similar to the results of a study on metagenomes and metabolic pathways to differentiate adjacent Red Sea Brine Pools (Yong *et al*., 2013). A high level of pathway completeness suggested their adaptive importance within the alkaline hot spring ecosystem. For instance, region V on **figure 6.7** comprised genes affiliated with pathways that participate in the carbohydrate, energy and amino acid metabolism as well as global and overview pathways which include: carbon metabolism, biosynthesis of amino acids, oxocarboxylic acid metabolism, xenobiotics biodegradation and metabolism and degradation of aromatic compounds. Other major pathways include butanoate metabolism, nitrogen metabolism, unsaturated fatty acids synthesis and lipid metabolism. The extensiveness of these pathways and carbohydrate transport and metabolism obtained from the COG category analysis suggest that some of the microbes present were largely heterotrophic prokaryotes that depended heavily on extracellular aromatic compounds as well as other sources of organic carbon.

### 6.4.3 Application of various enzymes obtained from the hot spring in biotechnology

Oxidoreductases such as *oxidases* and *laccases* catalyze biological oxidation/reduction reactions (May and Padgette 1983). These enzymes account for the highest proportion (32.4%) of the six categories of enzymes (Straathof, 2002). An example of an oxidoreductase enzyme isolated from the marine hyperthermophillic strain of the archaeon *Pyrococcus furiosus* is a soluble hydrogenase I (H2: NADP+ oxidoreductase, EC 1.18.99.1), which has been partially purified by anion-exchange chromatography (Sogin *et al*., 2006).

Transferases enzymes such as *acyltransferases* and *glycosyltransferases* catalyze the transfer of groups of atoms from one molecule to another or from one position in a molecule to other positions in the same molecule. Terminal transferases can be used to label DNA or to produce plasmid vectors (Bowen, 2013), while glutathione transferases can be used to create biosensors. Glutathione transferases are also used in transgenic plants to increase resistance to biotic and abiotic stress (Chronopoulou and Labrou 2009). In this study, transferases were dominant in mat and sediment samples (**Figure 6.14**). Hydrolytic enzymes catalyze reactions with the substrate through the hydrolysis of chemical bonds. They include amylases, proteases, lipases and esterases.

The study of hydrolytic enzymes is of interest due their potential applications in the medical, chemical and pharmaceutical industries. Hydrolases from extremophiles have advantages when compared to chemical biocatalysts as their catalyses processes are clean, ecologically friendly, and highly specific, which take place in mild reaction conditions (Dalma *et al*., 2015). They can also be active in the presence of organic solvents, an important feature for the preparation of single-isomer chiral drugs in pharmaceutical industry (Dalma *et al*., 2015). Hydrolases such as peptidases are useful catalysts for inorganic synthesis and have many industrial applications in the pharmaceutical field, such as anti-inflammatory and digestive agents (Trincone, 2010).

Peptidases can be isolated from marine extremophiles such as from *Thermotoga maritime*, a hyperthermophillic isolate from a marine geothermal area near Vulcano, Italy (Hicks *et al***.**, 1998). In this study, members of *Thermotogae* phylum obtained from the hot spring mRNA transcripts could be candidates for use in bioprospecting for more pharmaceutical products. Peptidases from halophiles such as *Halobacterium halobium* have been used in peptide synthesis; which has been exploited for efficient peptide synthesis in Water-*N*'-*N*'-dimethylformamide (Kim & Dordick, 1997). Members of the class *Halobacteria* were also found within the hot spring samples under investigation in this study and could be exploited in search for peptidases. Lyases, Isomerases and Ligases catalyze the addition of groups, rearrangement of atoms and join molecules together with covalent bonds in biosynthetic reactions.

## 6.5 Conclusion

This study compared microbial communities, genes and their potential functions, metabolic pathways as well as enzymes within the sediments, mats and water samples collected from the hot springs of Little Magadi. The results revealed differences in microbial composition and metabolic pathways between the hot spring transcriptomes; as well as prokaryotic taxa and metabolic pathways previously un-observed. The results provide evidence that microbes existing within the hot spring depended heavily on organic carbon, aromatic compounds, heterocyclic compounds, amino acids, storage and structural sugars within the surrounding environment. Some of the major sugar metabolism pathways across the samples were represented by glycolysis, TCA cycle and Pentose phosphate pathway among others. The major enzymes involved include: EC 1.1.1.267: 1-deoxy-D-xylulose-5-phosphate reductoisomerase (2.66 %), EC 3.2.1.52: beta-N-acetylhexosaminidase (2.66 %), EC 1.2.1.2: formate dehydrogenase. (1.93%) and EC 1.2.1.3: aldehyde dehydrogenase (NAD+) (1.93%). Several like pathways were revealed in a recent study on metabolic traits of an uncultured archaeal lineage -MSBL1- from brine pools of the Red Sea (Mwirichia *et al*., 2016). This confirms that different groups of microorganisms have the capacity to adapt and thrive even in the most hostile

environments. The major pathways in energy metabolism included carbon fixation in prokaryotes and photosynthetic organisms, oxidative phosphorylation, methane metabolism, and nitrogen metabolism. These are important in supporting various metabolic pathways and enhancing microbial survival within the hot spring ecosystem. The presence of DNA binding proteins within the samples possibly maintained microbial DNA in a double-stranded form at high temperatures within the hot spring environment (Pereira and Reeve 1998). Amino acid biosynthetic pathways enabled the cytoplasmic proteins of the microbes present to adapt to the environment by accumulating anionic amino acids on the cell surfaces, improving their stability and activity in saline alkaline waters (Hendry 2006). The study gives an insight of possible metabolic processes within the hot spring ecosystem and the genes possibly responsible for specific pathways. More than 50% of microbial species, KEGG Orthologs, Gene Ontology and Enzymes were unassigned within in the various functional classifications, being denoted as ‘other’. A comprehensive survey of microbial communities should therefore be undertaken at more sampling sites to establish an actual picture of various pathways within the hot spring and how they occur.

# CHAPTER SEVEN

# GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

## 7.1 General discussion

Studies in microbial ecology have helped researchers to recognize and appreciate the extent of diversity owing to the recent awareness created by metagenomic studies of components within the interactive assemblage of microorganisms, most of which are not achievable by standard culturing techniques (Woese, 1990; Head *et al.,* 1998; Bull, 2004; Handelsman, 2007).

Hypersaline-alkaline ecosystems such as soda lakes are one of the unique ecological niches considered to harbor a physiologically highly diverse group of microorganisms, adapted to survive under at least dual extreme environmental conditions (Nissenbaum, 1980; Singh *et al.,* 2009). Their salt pans are sites for the evolution of halophilic microbial groups which dominate while suppressing the less halophilic and halotolerant representatives (Ahmad *et al.,* 2008). The hot springs of Lake Magadi and Little Magadi are environments that experience multiple extremophilic features. Enzymes, compatible solutes and exopolysaccharides amongst all other metabolites derived from microorganisms native to these environments may be bioaugumented to drive processes with such development requirements as their natural habitat and are likely to offer more versatile bioactive compounds than the existing pool obtained from mesophiles (Ibrahim, 2013)

In this study, the total and active microbial diversity of Bacteria, Archaea and Fungal communities within the hot springs of Lake Magadi and Little Magadi were assessed using amplicon sequencing of 16S rDNA and 16S rRNA cDNA libraries generated from rRNA on Illumina platform. The diversity of metabolic genes expressed within the hot

spring at 83.6 ° C was explored using Short Gun Sequencing of cDNA libraries generated from mRNA.

Results from this study indicated that bacteria were the most dominant taxa in the three datasets explored. The most predominant bacterial phyla included *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, *Chloroflexi*, and *Deinococcus-thermus* among others; while Archaea were distributed among eight (8) phyla; *Euryarchaeota, Crenarchaeota, Thaumarchaeota, Korarchaeota*, *Nanoarchaeota*, *Nanohaloarchaeota*, *Parvarchaeota* and *unclassified archaea*.

Most prokaryotic and eukaryotic taxa diversity revealed in this study is presumed to be involved in nutrient cycling processes in the ecosystem i.e., carbon, sulfur and nitrogen cycling. Key processes like carbon fixation, methane and ammonia oxidation were also confirmed by mRNA transcript analysis. Considering the high pH, salinity and temperature of the hot springs under investigation, there were inexplicably diverse microbial communities; almost similar to that of fresh water environments. The relative abundance of microbial diversity within samples from the three hot springs appeared to be a function of sample type and hot spring temperature, with wet sediments harboring majority taxa across all samples.

The 16S rDNA and 16S rRNA cDNA amplicon datasets obtained were a representative for majority of taxa in the underlying prokaryotic community. This was confirmed using complementary shotgun sequencing of samples collected from the hot spring at 83.6 $^{o}$C. The resulting dataset conformed to sample type specific clustering patterns and shared almost similar relative abundances for most taxa when compared to corresponding amplicon datasets.

The study revealed presence of various groups that appeared to contribute to primary production within the hot springs. Some of these include a diversity of non-sulfur purple bacteria from the family *Rhodobacteraceae*, purple sulfur bacteria from the family *Ectothiorhodospiraceae* while an internal sulfur cycle was suggested by presence of

sulfate reducers from the family *Desulfonatronaceae*. Other taxa encountered at high abundance include aerobic heterotrophs (*Bacteroidetes*) and fermentative anaerobes (*Thermoplasmatales*). This conforms to the widespread and diverse microbial groups previously observed in soda lake environments (Zavarzin & Zhilina, 2000; Sorokin *et al*., 2011).

The findings of this study also revealed presence of moderate and weak alkalitolerant fungi such as *Alternaria alternata, Penicillium* sp., *Cladosporium* sp. and *Fusarium* sp. previously reported to grow optimally at neutral or below neutral pH values. Majority of the obtained fungal groups at species level included *Aspergillus*, *Ascomycota sp.*, *Penicillium*, *Neurospora*, *Termitomyces*, *Malassezia*, *Trichocomaceae*, *Stagonospora*, *Ramularia* and *Hypocreales*.

When linking microbial communities, genes and their potential functions, metabolic pathways and enzymes within the wet sediments, mats and water samples collected from the hot spring at 83.6 $^{o}$C, there were marked differences in microbial composition and metabolic pathways between the hot spring transcriptomes; prokaryotic taxa and metabolic pathways previously un-detected. There was evidence that the survival of microorganisms existing within the hot spring depend heavily on organic carbon, aromatic compounds, heterocyclic compounds, amino acids, storage and structural sugars within the surrounding environment.

A diversity of viruses and eukaryotes other than fungi was observed in this study. This could be attributed to the fact that some viruses exist in association with prokaryotic groups either as hosts or in symbiotic relationship. Several viruses are also common in the environment and due to their fast growth rates; they are candidate for utilization during cloning. However, the extent to which they regulate microbial community diversity and structure within these extreme environments remains unknown. A few presumed bacteriophage transcripts were found among the mRNA reads from samples collected from the hot spring at 83.6 $^{o}$C. Previously, similar results were observed in a

study conducted on five Ethiopian soda lakes (Lanze´n *et al*., 2013), thus pointing to possible areas of further investigation.

The heavy presence of transferases, Oxidoreductases and hydrolases within mat samples could be attributed to the mats containing a higher biomass that possibly acted as organic carbon sources for diverse microorganisms.

The outcome of this study supports previous observations which propose that soda lakes are home for unique microbial communities which may be regarded as centres of evolution of microbial diversity (Zavarzin *et al.,* 1999; Mesbah *et al.,* 2007). It confirms earlier reports that indicated soda lakes are an important unique niche that harbors numerous bacterial, archaeal and fungal taxa, some of which could possibly have originated from the terrestrial environments. However, the ecological roles of these microorganisms and their adaptive mechanisms remain poorly investigated.

### 7.2 Conclusions

- The assessment of total and active diversity of bacterial, archaea and fungal communities in the hot springs of Lake Magadi and Little Magadi revealed 3502 and 1913 Operational Taxonomic Units (OTUs) that were recovered from 16S rDNA and 16S rRNA cDNA, respectively.
- The predominant Bacterial phyla obtained from the OTUs were *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, *Chloroflexi*, and *Deinococcus-thermus*.
- The most dominant Archaeal phyla included *Euryarchaeota, Crenarchaeota, Thaumarchaeota* and *Korarchaeota*.
- There was presence of moderate and weak alkalitolerant fungi such as *Alternaria alternata, Penicillium* sp., *Cladosporium* sp. and *Fusarium* sp. previously reported to grow optimally at neutral or below neutral pH values.
- There was a broad microbial distribution with water samples from the hot spring at 83.6°C found to be the richest sample, constituting 680 observed species.

- The study of diversity of metabolic genes expressed in the hot springs provide evidence that microorganisms existing within the hot spring depended heavily on organic carbon, aromatic compounds, heterocyclic compounds, amino acids, storage and structural sugars within the surrounding environment.
- Some of the microorganisms obtained are relatives of viral lineages that could be controlling microbial mortality rates within the ecosystem.
- This study has provided new insights into the prokaryotic and eukaryotic composition of the environment as well as their possible ecological function. This data could serve as starting points for the development of new culture dependent techniques for as yet uncultivated microorganisms and unrecognized species.

## 7.3 Recommendations

1. Integration of data from metagenomics, metatranscriptomics, proteomics, and classical microbiology can help culture more uncultured organisms from these unique environments and manipulate them for practical applications in biotechnology.
2. More studies should be done on fungi from soda lake hot springs in order to reveal their diversity and functional role within the ecosystem.
3. Studies on viruses and other eukaryotes exposed in this study should also be done in order to reveal their role within the hot spring environment.
4. More applied research should be done in order to retrieve some of the enzymes shown in this study for future application in biotechnology

## REFERENCES

**Abd-el-Malek, Y. and Rizk, S. G. (1963).** Bacterial sulphate reduction and the development of alkalinity. III. Experiments under natural conditions in the Wadi Natrun. *Journal of Applied Bacteriology, **26**,* 20-26.

**Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., and Boonham, N. (2009).** Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology*. *Journal, **10**,* 537-545.

**Ahmad, N., Sharma, S., Khan, F. G., Rajinder, K., Johri, S., Abdin, M. Z. and Qazi, G. N. (2008).** Phylogenetic Analyses of Archaeal Ribosomal DNA Sequences from Salt Pan Sediment of Mumbai, India. *Current Microbiology, **57***, 145-152.

**Akanga, J. O., Boga, H. I. and Klenk, H. P. (2015).** Studies of sediments, dry mud and water samples from Lake Elmenteita reveal the presence of *Thaumarchaeota* and *Euryarchaeota*. *Journal of Biodiversity and Environmental Sciences, **6***(6), 341-355.

**Alam, I., Antunes, A., Kamau, A. A., Ba alawi, W., Kalkatawi, M. and Vladimir, B. B (2013).** INDIGO-INtegrated data warehouse of microbial genomes with examples from the red sea extremophiles. *PLoS One,* **8**,e82210.

**Alexey, A. G., Marina, L. G., Sofiya, A. B., Alfons, J. M. D. and Elena, N. B. (2015).** On the diversity of fungi from soda soils. *Fungal Diversity Journal* DOI10.1007/s13225-015-0320-2.

**Amann, R. I., Ludwig, W. and Schleifer, K. H. (1995).** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews, **59**,* 143-69.

**Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. …. and F. Rohwer. (2006)**. The marine viromes of four oceanic regions. *PLOS Biology,* **4,** e368.

**Barea, J. M., Pozo, M. J., Azcón, R., and Azcón-Aguilar, C. (2005).** Microbial co-operation in the rhizosphere. *Journal of Experimental Botany, **56***(417), 1761-778.

**Baker, H. (1958).** *Geology of the Magadi Area,* Reports of the Geological Survey of Kenya, no. 42. Nairobi: Government Printer.

**Barns, S. M., Delwiche, C. F., Palmer, J. D. and Pace, N. R. (1996).** Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences, **93,*** 9188 - 9193.

**Baumgarte, S. (2003).** *Microbial diversity of soda lake habitats.* Unpublished PhD thesis, Braunschweig: Carolo-Wilhelmina University.

**Behr, H. J., and Röhricht (2000).** Record of seismotectonic events in siliceous cyanobacterial sediments (Magadi cherts), Lake Magadi, Kenya. *International Journal of Earth Science* **89**: 268 - 283.

**Bell, E. M. and Callaghan, T. V. (2012),** 'What are Extreme Environments and What Lives in Them', *in* Bell, E. M. (Ed). *Life at Extremes: Environments, Organisms and Strategies for Survival,* London: CAB International.

**Bengtsson, M. M., Sjøtun, K. and Øvreas, L. (2010).** Seasonal dynamics of bacterial biofilms on the kelp Laminaria hyperborea. *Aquatic Microbial Ecology, 60,*71– 83.

**Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., and House, C. H. (2008).** Metagenomic signatures of the Peru Margin subsea floor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences,* USA. vol. **105** (30): 10583–10588.

**Bjornsdottir, S. H., Blondal, T., Hreggvidsson, G. O., Eggertsson, G., Petursdottir, S., Hjorleifsdottir, S., Thorbjarnardottir, S. H. and Kristjansson, J. K.**

**(2006).** Rhodothermus marinus: physiology and molecular biology. *Extremophiles, 10*(1), 1-16.

**Boltianskaya, V., Antipoz, A. N., Kolganova, T. V., Lysenko, A. M., Kostrikina, N. A., and Zhilina, T. N. (2004).** "Halomonas campisalis, an obligatorily alkaliphilic, nitrous oxide-reducing denitrifier with a Mo-cofactor-lacking nitrate reductase.". *Mikrobiologiia Journal, **73***(2004), 326-34.

**Borsodi, A. K., Ma´ Rialigeti, K., Szabo´, G., Palatinszky, M., Polla´ K, B., Ke´ Ki, Z., …** and **To´ Th, E. M. (2008).** Bacillus aurantiacus sp. nov., an alkaliphilic and moderately halophilic bacterium isolated from Hungarian soda lakes. *International Journal of Systematic and Evolutionary Microbiology, **58***, 845-851.

**Bowen, R. (2013).** Terminal Transferase. *Biotechnology and Genetic Engineering.* Colorado: Colorado State University.

**Brazelton, W. J. and Baross, J. A. (2009).** Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *International Society for Microbial Ecology Journal, **3***,1420-1424.

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

**Brem, F. M. and Lips, K. R. (2008).** *Batrachochytrium dendrobatidis* infection patterns among Panamanian amphibian species, habitats and elevations during epizootic and enzootic stages. *Diseases of Aquatic Organisms, **81***(3), 189-202.

**Brock, T. D. (1986).** *Thermophiles: General, Molecular and Applied Microbiology*. New York: John Wiley and Sons.

**Brown, L. (1973).** *The Mystery of the Flamingos*. (2nd ed.). East African Publishing House, Nairobi.

**Brown, M. V., Philip, G. K., Bunge, J. A., Smith, M. C., Bissett, A. Lauro, F. M., Fuhrman, J. A. and Donachie, S. P. (2009).** Microbial community

structure in the North Pacific Ocean. *International Society for Microbial Ecology Journal, 3,* 1374–1386.

**Buee, M., Reich, M., Murat, C., Morin, E., Nilsson, R. H., Uroz, S. and Martin, F. (2009).** Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist Journal* **184**:449-456.

**Bull, A. T. (2004).** *Biotechnology, The Art of Exploit ing Biology, in Bull, A. T., (Ed), Microbial Diversity and Bioprospecting*, , Washington, D. C.: ASM Press.

**Burgaud, G., Woehlke, S., Rédou, V., Orsi, W., Beaudoin, D., Barbier, G., … and Molina, M. (2006).** Characterization of *fungi* from hypersaline environments of solar salterns using morphological and molecular techniques. *Mycological Research, **110**,* 962-970.

**Burgaud, G., Woehlke, S., Rédou, V., Orsi, W., Beaudoin, D., Barbier, G., Biddle, J. and Edgcomb, V. P. (2013)**. Deciphering the presence and activity of fungal communities in marine sediments using a model estuarine system. *Aquatic Microbial Ecology, **70**,* 45-62.

**Bryantseva, I. A., Gorlenko, V. M., Kompantseva, E. I., Achenbach, L. A. and Madigan, M. T. (1999a).** *Heliorestis daurensis*, gen. nov. sp. nov., An alkaliphilic rod-to-coiled-shaped phototrophic heliobacterium from a Siberian soda lake. *Archives of Microbiology, **172***, 167-174.

**Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … and Knight, R. (2012).** Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *International Society for Microbial Ecology Journal, **6**,* 1621-1624.

**Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushma, F. D., Costello, E. K., … and Knight, R. (2010a).** QIIME allows analysis of high-throughput community sequencing data. *Nature Methods,* **7**, 335-336.

**Cantrell, S. A., Marti´nez, C. L. and Molina, M. (2006).** Characterization of *fungi* from hypersaline environments of solar salterns using morphological and molecular techniques. *Mycological Research, **110**,* 962-970.

**Cardenas, E. and Tiedje, J. M. (2008).** New tools for discovering and characterizing microbial diversity. *Current Opinion* in *Biotechnology, **19***, 544-549.

**Carola, S. and Rolf, D. (2009).** Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology* and *Biotechnology*, ***85,*** 265-276.

**Chistoserdova, L. (2010).** Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology Letters, **32***, 1351- 1359.

**Chomczynski, P. and Sacchi, N. (1987).** Single-step method of RNA isolation by acid guanidinium thiocyanate – phenol - chloroform extraction. *Analytical Biochemistry, **162,*** 156.

**Chronopoulou, E. G. and Labrou, N. E. (2009).** Glutathione transfer ases: emerging multidisciplinary tools in red and green biotechnology. Recent Patents on *Biotechnology, **3***(3), 211–23.

**Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., … and Farrow, J. A. (1994).** The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *International Journal of Systematic Bacteriology, **44***, 812-826.

**Coolen, M. J. L., Orsi, W. D., Balkema, C., Quince, C., Harris, K., Sylva, S. P., …. and Giosan, L. (2013).** Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. *Proceedings of the National Academy of Sciences,* USA.

**Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-Davis, R. M., Sung, W., Power, D. M.,… and Thomas, W. K. (2010)**. Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology, **19,*** 4-20.

**Dadachova, E., Bryan, R. A., Huang, X., Moadel, T., Schweitzer, A. D., Aisen, P., Nosanchuk, J. D. and Casadevall, A. (2007)**. Ionizing radiation changes the electronic properties of melanin and enhances the growth of melanized fungi. *PLoS One, **2***(5), e457.

**Dalma, G. Z. L., Ferreira, D. and Vermelho, A. B. (2015).** Marine Extremophiles: A Source of Hydrolases for Biotechnological Applications. *Marine Drugs, 13,* 1925-1965.

**Davis, L. E. (2012).** *An astrobiological study of an alkaline-saline hydrothermal environment relevant to understanding the habitability of Mars*. Unpublished PhD thesis, London: University College London.

**DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., … and Karl D, M. (2006).** Community Genomics among Stratified Microbial Assemblages in the Ocean's Interior. *Science, **311***(5760), 496-503.

**Dubinin, A. V., Gerasimenko, L. M. and Zavarzin, G. A. (1995).** Eco-physiology and species diversity of cyanobacteria from Lake Magadi. *Microbiology Journal, **64***, 717 – 721.

**Duckworth, A. W., Grant, W. D., Jones, B. E., Meijer, D., Marquez, M. C. and Ventosa, A. (2000).** *Halomonas magadii* sp. nov., a new member of the genus *Halomonas*, isolated from a soda lake of the East African Rift Valley. *Extremophiles, **4,*** 53-60.

**Duckworth, A. W., Grant, W. D., Jones, B. E., and van Steenbergen, R. (1996).** Phylogenetic diversity of soda Lake Alkaliphiles. *FEMS Microbiology and Ecology, **19**,* 181–191.

**Earl, A. M., Losick, R and Kolter, R. (2008).**Ecology and genomics of *Bacillus subtilis***.** *Trends in Microbiology, 16*(6), 269.

**Edgcomb, V. P. (2013).** Deciphering the presence and activity of fungal communities in marine sediments using a model estuarine system. *Aquatic Microbial Ecology, **70***, 45-62.

**Edgcomb, V. P., Beaudoin, D., Gast, R., Biddle, J. F.** and **Teske, A. (2011a)**. Marine subsurface eukaryotes: the fungal majority. *Applied and Environmental Microbiology, **13,*** 172-183.

**Elahi, E., and Ronaghi, M. (2004).** Pyrosequencing: a tool for DNA sequencing analysis. *Methods* in *Molecular Biology, **255**,* 211-219.

**Embley, T. M., and Stackebrandt, E. (1997).** Species in practice: exploring uncultured prokaryote diversity in natural samples. *In* Species: the Units of Biodiversity (M. F. Claridge, H. A. Dawah and M. R. Wilson, eds.), pp. 61-81. London: Chapman and Hall.

**Eugster, H. P and Jones, B. F., 1968).** Gels composed of Sodium-Aluminium Silicate, Lake Magadi, Kenya. *Science, **161***, 160-163.

**Eugster, H. P. (1970).** Chemistry and origins of the brines of Lake Magadi, Kenya. *Mineral Society of America Special Publication, **3***, 213-235.

**Filee, J., P. Forterre, and J. Laurent.** 2003. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Research in* Microbiology. ***154*,** 237–243.

**Forterre, P. (2006**). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research, **117**,* 5–16.

**Florenzano, G., Sili, C., Pelosi, E., and Vicenzini, M. (1985).** Cyanospira ripkae and Cyanospira capsulatus (gen. nov. sp. Nov.) a new filamentous heterocystous cyanobacterium from Lake Magadi (Kenya). *Archives of Microbiology, **140***, 301-307.

**Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., and Delong, E. F. (2008).** Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences,* USA.

**Frontier, S. (1985).** Diversity and structure in aquatic ecosystems. *Journal of Oceanography* and *Marine Science, **23***, 253-312.

**Frostegard, A., Tunlid, A., and Baath, E. (1993).** Phospholipid fatty acid composition, biomass, and activity of microbial communities from two soil types experimentally exposed to different heavy metals. *Applied* and *Environmental Microbiology, **59***, 3605-17.

**Gadd, G. M. (2007).** Geomycology: Biogeochemical transformations of rocks, minerals, metals and radionuclides by fungi, bio-weathering and bioremediation. *Mycological Research,* ***111*** (Pt 1), 3 - 49.

**Gayantha, R. L. K. (2009).** *Hyperspectral mapping of surface mineralogy in the Lake Magadi area in Kenya.* Unpublished MSc Thesis. International Institute of Geo-information Science and Earth Observation. Netherlands: Enschede.

**Gerasimenko, L. M., Dubinin, A. V., Mityushina, L. L. and Zarvazin, G. A. (1999).** 'A Microscopic Green Alga from Soda Lakes', *Microbiology, translated from Mikrobiologiya,* ***68***, 696-700.

**Ghosh, A., Bhardwaj, M., Satyanarayana, T., Khurana, M., Mayilraj, S., and Jain, R. K. (2007).** Bacillus lehensis sp. nov., an alkalitolerant bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology,* ***57***, 238-242.

**Gifford, S. M., Sharma, S., Rinta-Kanto, J. M.,** and **Moran, M. A. (2010).** Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *International Society for Microbial Ecology Journal,* ***5****,* 461-472.

**Gontcharova, V., Youn, E., Sun, Y., Wolcott, R. D. and Dowd S. E. A. (2010).** Comparison of bacterial composition in diabetic ulcers and contralateral intact skin. *Open Microbiology Journal,* ***4****,* 8-19.

**Grant, S., Grant, D., Jones, B. E., Kato, C., and Li, L. (1999).** Novel Archaeal phenotypes from an East African alkaline environment. *Extremophiles,* ***3****,* 139–145.

**Grant, W. D. (1992).** Alkaline Environments. *In* Encyclopedia of microbiology (J. Lederberg, ed.), Vol. 1, pp. 73-80. London; Academic Press.

**Grant, W. D., Mwatha, W. E., and Jones, B. E. (1990).** Alkaliphiles: ecology, diversity and applications. *FEMS Microbiology Review,* ***75****,* 255-270.

**Grant, W. D. and Sorokin, D. Y. (2011).** *Distribution and diversity of soda lake alkaliphiles. . In* Extremophiles Handbook (K. Horikoshi, G. Antranikian,

A. T. Bull, F. T. Robb and K. O. Stetter, eds.), Vol. 1, pp. 27–54. London: Springer.

**Grant, W. E. (2006).** Alkaline Environments and Biodiversity, in Extremophiles. [Eds. Charles Gerday, and Nicolas Glansdorff], in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Oxford: Eolss Publishers.

**Grant, W. D. and Tindall, W. J. (1986)**. 'The Alkaline Saline Environment', *in* Herbert, R. A. and Codd, G. A., (Eds) *Microbes in Extreme Environments,* 1st edition, Academic Press, London. 25 -54.

**Gunde-Cimerman, N., Butinar, L., Sonjak, S., Turk, M., Uršic, V., Zalar P. and Plemenitaš A (2005).** Halotolerant and halophilic fungi from coastal environments in the Arctics. In: Gunde-Cimerman N, Oren A, Plemenitaš A (eds) Adaptation to life at high salt concentrations in Archaea, Bacteria and Eukarya. Springer, Netherlands, pp 397-423.

**Gunde-Cimerman, N., Ramos, J. and Plemenitaš, A. (2009).** Halotolerant and halophilic fungi. *Mycological Research, 113,*1231-1241.

**Handelsman, J. (2004).** Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology* and *Molecular Biology Reviews,* ***68***, 669.

**Hansel, C. M., Fendorf, S., Jardine, P. M. and Francis, C. A. (2008).** Changes in bacterial and Archaeal community structure and functional diversity along a geochemically variable soil profile. *Applied* and *Environmental Microbiology,* ***74,*** 1620-1633.

**Head, I. M., Saunders, J. R., Pickup, R. W. (1998),** Microbial Evolution, Diversit y, and Ecology: A Decade of Ribosomal RNA Analysis of Uncult ivated Microorganisms. *Microbial Ecology,* ***35,*** 1-21.

**Healy, P. C., Hocking, A., Tran-Dinh, N., Pitt, J. I., Shivas, R. G., Mitchell, J. K., Kotiw, M. and Davis, R. A. (2004).** Xanthones from a microfungus of the genus *Xylaria*. *Phytochemistry,* ***65***(16), 2373-2378.

**Hecky, R. E. and Kilham, P. (1973).** Diatoms in alkaline saline lakes: ecology and geochemical implications. *Journal of Limnology and Oceanography, **18**,* 53-71.

**Hendry, P. (2006).** There is more to life. *Environmental Chemistry Journal, **3**,* 75-76

**Hicks, P.M., Rinker, K.D., Baker, J.R. and Kelly, R.M. (1998).** Homomultimeric protease in the hyperthermophillic bacterium *Thermotoga maritima* has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria. *FEBS Letters, **440**,* 393–398.

**Holben, W. E. and Harris, D. (1995).** DNA-based monitoring of total bacterial community structure in environmental samples. *Molecular Ecology, **4**,* 627-631.

**Horikoshi, K. (1999).** Alkaliphiles: some applications of their products for biotechnology. *Microbiology and Molecular Biology Review, **63***, 735.

**Horikoshi, K. (1991).** *General view of alkaliphiles and thermophiles. In* Microorganisms in Extreme Environments (K. Horikoshi and W. Grant, eds.), pp. 3-14. Tokyo: Japan Scientific Societies Press.

**Hugenholtz, P. (2002).** Exploring prokaryotic diversity in the genomic era. *Genome Bioology Reviews, **3***(2), 0003.1–0003.8.

**Hyatt, D., Chen, G. L. and Lacascio, P. F. (2010).** Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics, 11,* doi: 10.1186/1471-2105-11-119.

**Ibrahim, O. H. (2013)**. *Actinobacterial and Archaeal Diversity in Lake Magadi, Kenya.* Department of Biotechnology, Pretoria: University of the Western Cape.

**Jones, S. E. and Lennon, J. T. (2010).** Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences,* USA, ***107*** (13): 5881–5886.

**Jones, B. E. and Grant, W. D. (2000).** *Microbial diversity and ecology of alkaline environments. In* "Journeys to diverse microbial worlds" (J. Seckbach, ed.), pp. 177-190. Netherlands: Kluwer Academic.

**Jones, B. E. and Grant, W. D. (1999).** 'Microbial Diversity and Ecology of the Soda Lakes of East Africa', *in* Bell, C. R., Brylinsky, M and Johnson-Green, P. (Eds), *Ecology and Diversity of Extremophiles,* Proceedings of the 8th International Symposium on Microbial Ecology, Atlantic Canada Society for Microbial Ecology, Halifax, Canada, 1-6.

**Jones, B. E., Grant, W. D., Duckworth, A. W., and Owenson, G. G. (1998).** Microbial diversity of soda lakes. *Extremophiles, **2**,* 191–200.

**Jones, B. F., Eugster, H. P., and Rettig, S. L. (1977).** Hydrochemistry of the Lake Magadi basin, Kenya. *Geochimica et Cosmochimica Acta*' *Journal, **41**,* 53-72.

**Jones, B. E., Grant, W. D., Collins, N. C., and Mwatha, W. C. (1994).** *Alkaliphiles: diversity and identification. In Bacterial Diversity and Systematics* (F. G. Priest, A. Ramos-Cormenzana and B. J. Tindall, eds.), pp. 195-230. New York.: Plenum Press.

**Kambura, A. K., Boga, H. I., Mwirichia, R. K., and Ngaira, J., N. ( 2012).** Isolation and characterization of bacterial isolates from Lake Magadi. *Journal of Tropical Microbiology and Biotechnology, **8**,* 20 - 28.

**Kambura, A. K., Mwirichia, R., Kasili, R., Karanja, E N., Makonde H. and Boga, H. I. (2016).** Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiology*, *16,* Pg 1. DOI: 10.1186/s12866-016-0748-x

**Keller, M., and Zengler, K. (2003).** Tapping into microbial diversity. *Nature Reviews*, ***2***,141.

**Kempe, S. and Ka mierczak, J. (2002).** Biogenesis and Early Life on Earth and Europa: Favored by an Alkaline Ocean? *Astrobiology, **2**,* 123-130.

**Kevbrin, V. V., Lysenko, A. M. and Zhilina, T. N. (1997).** Physiology of the alkaliphilic methanogen Z-7936, a new strain of *Methanosalsus zhilinaeae* isolated from Lake Magadi. *Microbiology, **66**,* 261-266.

**Kevbrin, V. V., Zhilina, T. N., Rainey, F. A. and Zavarzin, G. A. (1998).** *Tindallia magadii* gen. nov., sp. nov.: an alkaliphilic anaerobic ammonifier from soda lake deposits. *Current Microbiology, **37**,* 94-100.

**Kim, J. and Dordick, J.S. (1997).** Unusual salt and solvent dependence of a protease from an extreme halophile. *Biotechnology and Bioengineering, **55**,* 471–479**.**

**Kladwang, W., Bhumirattana, A. and Hywel-Jones, N. (2003)** Alkaline-tolerant fungi from Thailand. *Fungal Diversity Journal, **13**,* 69-83.

**Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M.** and **Turner, D. J. (2009).** Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods, **6**,* 291-295.

**Kristjansson, J. K. and Hreggvidsson, G. O. (1995),** 'Ecology and Habitats of Extremophiles'. *World Journal of Microbiology and Biotechnology, **11**,* 17-25.

**Lanze´n A., Simachew A., Gessesse A., Chmolowska D., Jonassen I., and Øvreas, L. (2013).** Surprising Prokaryotic and Eukaryotic Diversity, Community Structure and Biogeography of Ethiopian Soda Lakes. *PLoS One, **8***(8), e72577.

**Le Calvez, T., Burgaud, G., Mahé, S., Barbier, G. and Vandenkoornhuyse, P. (2009).** Fungal diversity in deep sea hydrothermal ecosystems. *Applied and Environmental Microbiology, **75**(20),* 6415-6421.

**Lee, J. C., Lee, G. S., Park, D. J. and Kim, C. J. (2008).** Bacillus alkalitelluris sp. nov., an alkaliphilic bacterium isolated from sandy soil. *International Journal of Systematic and Evolutionary Microbiology, **58**,* 2629-2634

**Lenchi, N., Inceo lu ö., Kebbouche-Gana, S., Gana, M. L., Llir s M., Servais, P and Garcia-Armisen, T. (2013).** Diversity of Microbial Communities in Production and Injection Waters of Algerian Oilfields Revealed by 16S rRNA Gene Amplicon 454 Pyrosequencing. *PLoS ONE, **8** (6),* e66588. doi:10.1371/journal.pone.0066588.

**Liao, L., Xu, X. W., Jiang, X. W., Wang, C. S., Zhang, D. S., Ni, J. Y. and Wu, M (2011).** Microbial diversity in deep-sea sediment from the cobalt rich crust deposit region in the Pacific Ocean. *FEMS Microbiology Ecology, **78**,* 565–585

**Linhoss, C. A., Kiker, G., Shirley, M, and Frank, K. (2015).** Sea-Level Rise, Inundation, and Marsh Migration: Simulating Impacts on Developed Lands and Environmental Systems. *Journal of Coastal Research, **1**,* 36-46. doi: 10.2112/JCOASTRES-D-13-00215.1

**Lindahl, B. D., Ihrmark, K., Boberg, J., Trumbore, S. E., Högberg, P. Stenlid, J. and Finlay, R. D. (2007).** Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytologist, **173**(3),* 611-620.

**Lopez-Garcia, P. and Moreira, D. (2008).** Tracking microbial biodiversity through molecular and genomic ecology. *Research* in *Microbiology Journal, **159**,* 67-73.

**Lorrayne, G. M., Guilherme, C. F., Guilherme, L. M., Joseane, B. C., Franceli, R. K., and Rogerio, M. (2012).** Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. *Genetics and Molecular Biology, **35**,* 292-303.

**Ma, Y., Zhang, W., Xue, Y., Zhou, P., Ventosa, A. and Grant, W. D. (2004).** Bacterial diversity of the Inner Mongolian Baer Soda Lake as revealed by 16S rRNA gene sequence analyses. *Extremophiles, **8**,* 45-51.

**Magan, N. and Aldred, D. (2007).** Environmental fluxes and fungal interactions: maintaining a competitive edge. In: Van West P, Avery S (eds).

**Magan, N. (2006).** *Fungi in extreme environments: Environmental and microbial relationships,* (2nd ed.), The Mycota IV. In: Christian P. Kubicek, Irina S. Druzhinina (eds). London: Springer.

**Mardanov, A. V., Gumerov, V.M., Beletsky, A. V., Perevalova, A. A., Karpov, G. A., Bonch-Osmolovskaya, E. A. and Ravin, N. V. (2011).** Uncultured archaea dominate in the thermal groundwater of Uzon Caldera, Kamchatka. *Extremophiles, **15**,* 365-372.

**May W. S. and Padgette R. S. (1983).** Oxidoreductase Enzymes in Biotechnology: Current Status and Future Potential. *Nature Biotechnology, **1**,* 677-686.

**Melack, J. M. and Kilham, P. (1974).** Photosynthetic rates of phytoplankton in East African Lakes. *Limnology and Oceanography, **19**,* 743–755.

**Mesbah, N. M., Abou-El-Ela, S. H. and Wiegel, J. (2007).** Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microbiology and Ecology, **54**,* 598-617.

**Moran, M. A. (2009)**. Metatranscriptomics: eavesdropping on complex microbial communities. *Microbes and Environments Journal, **4**,* 329-335.

**Mwatha, W. E and Grant, W. D. (1993).** *Natronobacterium vacuolata* sp. nov., a haloalkaliphilic archaeon isolated from Lake Magadi, Kenya. *International Journal of Systematic Bacteriology, **43**,* 401-406.

**Mwirichia, R., Alam, I, Rashid, M., Vinu, M., Ba- Alawi, W., Kamau, A. A., … and Stingl, U. (2016).** Metabolic traits of an uncultured archaeal lineage - MSBL1- from brine pools of the Red Sea. *Scientific Reports, **6**,* 19181.

**Mwirichia, R. (2009).** Microbial diversity of Lake Elementaita, Kenya . Unpublished PhD Thesis. Juja: JKUAT.

**Mueller, G. M. and Schmit, J. P. (2006).** Fungal biodiversity: what do we know? What can we predict? *Biodiversity and Conservation,* ***16,*** 1–5.

**Muyzer, G., Sorokin, D. Y., Mavromatis, K., Lapidus, A., Clum, A Ivanova, N.,… and Kyrpides, N. C. (2011).** Complete genome sequence of "*Thioalkalivibrio sulfidophilus*" HL-EbGr7**.** *Stand Genomic Sci,* ***4****(1),* 23–35.

**Nagahama, T., Takahashi, E., Nagano, Y., Abdel-Wahab, M. A. and Miyazaki, M. (2011).** Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. Applied and *Environmental Microbiology,* ***13,*** 2359-2370.

**Ndwigah, F. I., Boga, I. H., Wanyoike, W. and Kachiuru, R. (2015).** Characterization, Enzymatic Activity, and Secondary Metabolites of Fungal Isolates from Lake Sonachi in Kenya. IOSR *Journal* of *Pharmacy* and *Biological Sciences,* ***10,*** 65-76.

**Nakai, R., Abe, T., Takeyama, H. and Naganuma, T. (2011**). Metagenomic analysis of 0.2-mm-passable microorganisms in deep-sea hydrothermal fluid. *Marine Biotechnology,* ***13,*** 900-908.

**Neefs, J. M., van de Peer, Y., Hendriks, L. and de Wachter, R. (1990)**. Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Research,* ***18,*** pp. 2237–2317

**Neira, C., Sellanes, J., Levin, L. A. and Arntz, W. E. (2001).** Meiofaunal distributions on the Peru margin: relationship to oxygen and organic matter availability. *Deep-Sea Research Part I: Oceanographic Research Papers,* ***48,*** 2453-2472.

**Nelson, W. C., Wollerman, L., Bhaya, D. and Heidelberg, J. F. (2011).** Analysis of insertion sequences in thermophilic cyanobacteria: exploring the mechanisms of establishing, maintaining, and withstanding high insertion

sequence abundance. *Applied and Environmental Microbiology, **77**,* 5458-5466.

**Nielsen, P., Fritze, D. and Priest, F. G. (1995).** Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine new species. *Microbiology, **141**,* 1745-1761.

**Nielsen, P., Rainey, F. A., Outtrup, H., Priest, F. G. and Fritze, D. (1994).** Comparative 16S rDNA sequence analysis of some alkaliphilic *bacilli* and the establishment of a sixth rRNA group within the genus *Bacillus*. *FEMS Microbiology Letters, **117**,* 61-66.

**Nissenbaum, A. (1980).** Hypersaline Brines and Evaporitic Environments: Proceedings of the BatSheva Seminar on Saline Lakes and Natural Brines, *in* (Ed) *Developments in Sedimentology*, Elsevier, **3** and **4**, 23-60.

**Nogi, Y., Takami, H., and Horikoshi, K. (2005).** Characterization of alkaliphilic Bacillus strains used in industry: proposal of five novel species. *International Journal of Systematic and Evolutionary Microbiology, **55**,* 2309-2315.

**Noguchi, H., Taniguchi, T. and Itoh, T. (2008).** MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Research Journal, **15**,* 387–396, doi: 10.1093/dnares/dsn027 (2008).

**Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee G. J., Hattori, M., Kanai, A., Atomi, H., Takai, K. and Takami, H. (2011).** Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Research, **39**,* 3204-3223.

**Nura A. A. and Amir, D. (2015).** Halobacterium – A Model of Life in the Dead Sea. *http://www.ecomena.org/halobacterium/.*

**Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R B., … and Wagner, H. (2012). v**egan: Community Ecology Package. R package version 2.0-3**.**

**Olivera, N., Sineriz, F. and Breccia, J. D. (2005).** Bacillus patagoniensis sp. nov., a novel alkalitolerant bacterium from the rhizosphere of Atriplex lampa in Patagonia, Argentina. *International Journal of Systematic and Evolutionary Microbiology, **55**,* 443–447.

**Oren, A. (1999).** Bioenergetic Aspects of Halophilism. *Microbiology and Molecular Biology Reviews, **63***(2), 334-348.

**Patel, J. B., Leonard, D. G. B., Pan, X., Musser, J. M., Berman, R. E., and Nachamkin, I. (2000).** Sequence based Identification of Mycobacterium species using the MicroSeq 500 16S rDNA bacterial identification system. *Journal* of *Clinical Microbiology, **38**,* 246-251.

**Pawlowski, J., Lejzerowicz, F. and Esling, P. (2014).** Next-Generation Environmental Diversity Surveys of Foraminifera: Preparing the Future. *Biology Bulletin, **227***(2), 93–106.

**Pereira, S. L. and Reeve, J. N. (1998).** Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. *Extremophiles, **2**,* 141-148.

**Petti, C. A. (2007).** Detection and identification of microorganisms by gene amplification and sequencing. *Clinical Microbiology and Infectious Diseases, **44**,* 1108-1114.

**Pickup, R. W. (1991).** Development of molecular methods for the detection of specific bacteria in the environment. *Journal of General Microbiology, **137**,* 1009-1019.

**Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A., and Hollibaugh, J. T. (2005).** Analysis of Microbial Gene Transcripts in Environmental Samples. *Applied and Environmental Microbiology, **71**,* 4121-4126.

**Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P., and Moran, M. A. (2009).** Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology Reports, **11**,* 1358-1375.

**Posada, F., Aime, M. C., Peterson, S. W., Rehner, A. S. and Vega, E. F. (2007).** Inoculation of coffee plants with the fungal entomopathogen *Beauveria bassiana* (Ascomycota: Hypocreales). *Mycological Research, **111**,* 748-757.

**Prowe, S. G., and Antranikian, G. (2001).** Anaerobranca gottschalkii sp. nov., a novel thermoalkaliphilic bacterium that grows anaerobically at high pH and temperature. *International Journal of Systematic and Evolutionary Microbiology, **51**,* 457- 465.

**Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2013).** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, **41**(DI)*, D590-D596.

**Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H. T., Urich, T. and Schleper, C. (2012).** Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community. *Environmental Microbiology, **14**,* 1308–1324.

**R Development Core Team (2012).** R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

**Raghukumar, C. and Raghukumar, S. (1998)** Barotolerance of fungi isolated from deep-sea sediments of the Indian Ocean. *Aquatic Microbial Ecology, **15**(2),* 153-163.

**Raymon, J. C., and Sistrom, W. R. (1967).** The isolation and preliminary characterization of a halophilic photosynthetic bacterium. *Archives of Microbiology, **59**,* 255-268.

**Reeder, J., and Knight, R. (2010).** Rapidly denoising pyrosequencing amplicon reads exploiting rank-abundance distributions. *Nature Methods, 7,* 668-669.

**Rees, H. C., Grant, W. D., Jones, B. E. and Heaphy, S. (2004).** Diversity of Kenyan soda Lake Alkaliphiles assessed by molecular methods. *Extremophiles, **8**,* 63-71.

**Riesenfeld, C. S., Schloss, P. D. and Handelsman, J. (2004**). Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics, **38**,* 525-552.

**Roche, C. M., Loros, J. J., McCluskey, K. and Glass, N. L. (2014).** *Neurospora crassa*: Looking back and looking forward at a model microbe. *American Journal of Botany, **101***(12), 2022-2035.

**Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., Daroub, S. H., Camargo, F. A., Farmerie, W. G. and Triplett, E. W. (2007).** Pyrosequencing enumerates and contrasts soil microbial diversity. *International Society for Microbial Ecology Journal, **4**,* 283-90.

**Rothschild, L. J. and Mancinelli, R. L. (2001).** Life in extreme environments. *Nature, **409**,* 1092–1101.

**Ruiz-Suárez, J. (2004).** *Arenicolus filamentous fungi in Mayagüez Bay shoreline, Western Puerto Rico.* Unpublished MSc Thesis, Puerto Rico: University of Puerto Rico.

**Russell, J. R., Huang, J., Anand, P., Kucera, K., Sandova, A. G., Dantzler, K. W.,… and Strobel, S. A. (2011).** Biodegradation of polyester polyurethane by endophytic fungi. *Applied* and *Environmental Microbiology*, ***77**,* 6076-6084.

**Salano, O. (2011).** Isolation and characterization of fungal communities from L. Magadi. Unpublished MSc Thesis, Juja: Jomo Kenyatta University of Agriculture and Technology.

**Sambrook, K. J., Fritsch, E. F. and Maniatis, T. (1989).** *Molecular Cloning: a Laboratory Manual.*(2nd Eds.). New York, USA: Cold Spring Harbor

Laboratory,

**Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., and Huttenhower, C. (2013).** Computational meta'omics for microbial community Studies. *Molecular Systems Biology, **9**,* 666.

**Shi, Y., Tyson, G. W., and DeLong, E. F. (2009).** Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature, **459**,* 266-269.

**Singh, S. P., Purohit, M. K., Raval, V. H., Pandey, S., Akbari, V. G. and Rawal, C. M. (2010).** Capturing the Potential of Haloalkaliphilic Bacteria from the Saline Habitats Through Culture Dependent and Metagenomic Approaches, *in* Mendz-Vilas, A. (Ed) Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology. 81 - 87.

**Sjöling, S., and Cowan, D. A. (2008).** *Metagenomics: microbial community genomes revealed. In* Psychrophiles: from biodiversity to biotechnology (R. Margesin, F. Schinner, J. C. Marx and C. Gerday, eds.), pp. 313-332. Berlin Heidelberg: Springer.

**Sleator, R. D., C., S., and Hill, C. (2008).** Under the Microscope Metagenomics. *Letters in Applied Microbiology*, 361-366.

**Sogin, M. L., Morrison, G. H., Huber, J A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J. (2006).** Microbial diversity in the deep sea and the underexplored rare biosphere. *Proceedings of the National Academy of Sciences* of the USA, ***103***(32)*,* 12115 –12120.

**Sorokin, D. Y., deJong, G. A. H., Robertson, L. A., and Kuenen, G. J. (1998).** Purification and characterization of sulfide dehydrogenase from alkaliphilic chemo-litho-autotrophic sulfur-oxidizing bacteria. *FEBS Letters, **427**,* 11-14.

**Sorokin, D. Y., Trotsenko, Y. A., Doronina, N. V., Tourova, T. P., Galinski, D. A., Kolganova, T. V., and Muyzer, G. (2007).** Methylohalomonas lacus gen. nov., sp. nov. and Methylonatrum kenyense gen. nov., sp. nov., methylotrophic gamma-proteobacteria from hyper saline lakes.

*International Journal of Systematic and Evolutionary Microbiology,* ***57***(Pt 12), 2762-9.

**Sorokin, D. Y., Kuenen, J. G. and Muyzer, G. (2011).** The microbial sulfur cycle at extremely haloalkaline conditions of soda lakes. *Frontiers in Microbiology,* ***2****,* 44.

**Stackebrandt, E., and Goebel, B. M. (1994).** Taxonomic note: a place for DNA-DNA reassociation and 16SrRNA sequence analysis in the present species definition in bacteriology. *International journal of systematic bacteriology,* ***44****,* 846-849.

**Stackebrandt, E., Liesack, W., and Goebel, B. M. (1993).** Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. *Federation of American Societies for Experimental Biology Journal,* ***7****,* 232-236.

**Staley, J. T. and Konopka, A. (1985)**. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology,* ***39****,* 321-46.

**Stetter, K. O. (1996).** Hyperthermophiles in the history of life. *In* "Ciba Found Symposium", Vol. **202**, pp. 1-18.

**Stetter, K. O. (1996)**. Hyperthermophillic prokaryotes. *FEMS Microbiology* Reviews, ***18****,* 149-158.

**Stetter, K. O., and Zillig, W. (1985).** *Thermoplasma and the thermophilic sulfur-dependent archaebacteria. In* "The Bacteria" (R. S. Wolfe and C. R. Woese, eds.), Vol. 8, pp. 85-170. New York: Academic Press.

**Straathof, A. J., Panke, S. and Schmid, A. (2002).** The production of fine chemicals by biotransformations. *Current Opinion in Biotechnology,* ***13***(6), 548-56.

**Suttle, C. A.** 2007. Marine viruses—major players in the global ecosystem. *Nature Reviews Microbiology,* ***5****,* 801–812.

**Svetlitshnyi, V., Rainey, F., and Wiegel, J. (1996).** Thermosyntropha lipolytica gen. nov., sp. nov., a lipolytic, anaerobic, organo-heterotrophic, alkalitolerant

thermophile utilizing short- and long chain fatty acids in syntrophic co-culture with a methanogen. *International Journal of Systematic Bacteriology, **46,*** 1131-1137.

**Takai, K. Nakamur, K., Toki, T., Tsunogai, U. Miyazaki, M., Miyazaki, J.,… and Horikoshi K. (2008).** Cell proliferation at 122ºC and isotopically heavy CH4 production by a hyperthermophillic methanogen under high-pressure cultivation. *Proceedings of the National Academy of Sciences, USA,**105,***10949–10954.

**Tanner, M. A., Goebel, B. M., Dojka, M. A., and Pace, N. R. ( 1998).** Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Applied and Environmental Microbiology, **64,*** 3110-3113.

**Tekere, M., Lötter, A., Olivier, J., Jonker, N. and Venter, S. (2011).** Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa**.** *African Journal of Biotechnology, **10,*** 18005-18012.

**Tindall, B. J., Ross, H. N. M., and Grant, W. D. (1980).** An alkalophilic red halophilic bacterium with a low magnesium requirement from Kenyan Soda Lake. *Journal of General Microbiology, **11,*** 257-260.

**Tindall, B. J., Ross, H. N. M. and Grant, W. D. (1984).** *Natronobacterium* gen.nov. and *Natronococcus* gen. nov., two genera of haloalkaliphilic archaebacteria. *Journal of Systematic Applied Microbiology, **5,*** 41-57.

**Tindall, B. J., Ross, H. N. M., and Grant, W. D. (1984).** *Natronobacterium* gen.nov. and *Natronococcus* gen. nov., two genera of haloalkaliphilic archaebacteria. *Journal of Systematic Applied Microbiology, **5,*** 41-57.

**Tindall, B. J. (1988).** Prokaryotic life in the alkaline, saline, athalassic environment. *In* "Halophilic Bacteria" (F. Rodriguez-Valera, ed.), pp. 31-67. CRC Press, Boca Raton, FI.

**Thiel, V. (2011)**. *Extreme Environments.* In J. Reitner and V. Thiel (eds), Encyclopedia of Geobiology, Encyclopedia of Earth Science Series, Heidelberg: *Springer*.

**Torsvik, V., and Øvreås, L. (2002).** Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion* in *Microbiology, **5**,* 240-245.

**Torsvik, V., Sørheim, R., and Goksøyr, J. (1996).** Total bacterial diversity in soil and sediment communities - a review. *Journal* of *Industrial Microbiology and Biotechnology, **17**,*170 - 178.

**Tourova, T. P., Garnova, E. S. and Zhilina, T. N. (1999).** Phylogenetic diversity of alkaliphilic anaerobic saccharolytic bacteria isolated from soda lakes. Microbiology. Translated from *Mikrobiologiya*. ***68***, 701- 709.

**Trincone, A.** (**2010**), Potential biocatalysts originating from sea environments. *Journal* of *Molecular Catalysis B-Enzymes, **66***, 241–256.

**Tsukagoshi, N., Kobayashi, T. and Kato, M. (2001).** Regulation of the amylolytic and hemi-cellulolytic genes in aspergilli. *Journal* of *General and Applied Microbiology, **47*** 1-19.

**Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., and Schuster S. C. (2008).** Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. *PLoS ONE,* ***3*****(6),** e2527.

**Vaupotic, T., Veranic, P., Jenoe, P. and Plemenitas, A. (2008).** Mitochondrial mediation of environmental osmolytes discrimination during osmoadaptation in the extremely halotolerant black yeast *Hortaea werneckii*. *Fungal Genetics and Biology, **45***(6), 994-1007.

**Vieille, C. and Zeikus, G. J. (2001).** Hyperthermophillic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews, 65*(1), 1092-2172.

**Vladimir, G., Alexandre, T., Zorigto, N., Tracy, T., Tatyana, T., E., D., Randy, M., and Kenneth, N. (2004)**. Anaerobranca californiensis sp. nov., an

anaerobic, alkalithermophilic, fermentative bacterium isolated from a hot spring on Mono Lake. *International Journal of Systematic Evolution in Microbiology, **54,*** 739-743.

**Von Wintzingerode, F., Gobel, U. B., and Stackebrandt, E. (1997).** Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews, **21**,*213-229.

**Yong, W., Cao, H., Zhang, G., Bougouffa, S., Lee, O. O., Suwailem, A. and Qian, P. (2013).** Autotrophic Microbe Metagenomes and Metabolic Pathways Differentiate Adjacent Red Sea Brine Pools. *Scientific Reports, **3**,* 1748.

**Warnecke, F., and Hess, M. (2009).** A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology, **142***, 91-95.

**Warren, J. K. (2006)**. Subaqueous salts: salinas and perennial lakes. Evaporites sediments, resources and hydrocarbons. *Springer,* 221-285.

**Weinbauer, M. G., and Rassoulzadegan F. (2004)**. Are viruses driving microbial diversification and diversity? *Environmental Microbiology, **6**,* 1–11.

**White, T. J, Bruns, T., Lee, S. and Taylor, J. (1990).** *Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.* In Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR Protocols: a guide to methods and applications. New York: Academic Press.

**Woese, C. R., Kandler, O. and Wheelis, M. L. (1990).** Towards A Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya, *Proceedings of the National Academy of Science of the United States of America, **87**,* 4576-4579.

**Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., … and Xu, A. (2011).** Comparative Metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *International Society for Microbial Ecology Journal, **5**,* 414-426.

**Xu, J. (2006).** Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology, **15**,* 1713-1731.

**Yu, K., and Zhang, T. (2012).** Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *PLoS ONE, 7*(**5**), e38183.

**Yumoto, I., Yamaga, S., Sogabe, Y., Nodasaka, Y., Matsuyama, H., Nakajima, K., and Suemori, A. (2003).** Bacillus krulwichiae sp. nov., a halotolerant obligate alkaliphile that utilizes benzoate and hydroxybenzoate. *International Journal of Systematic and Evolutionary Microbiology,* ***53***, 1531-1536.

**Yuriko, N. and Takahiko, N. (2012).** Fungal diversity in deep-sea extreme environments. *Fungal Ecology,* ***5***, 463-471.

**Zavarzin, G. A. and Zhilina, T. N. (2000).** Anaerobic chemotrophic alkaliphiles. In: Seckbach J, editor. Journey to Diverse Microbial Worlds: *Adaptation to Exotic Environments*: Kluwer Academic Publishers. pp. 191–108.

**Zavarzin, G. A., Zhilina, T. N., and Kevbrin, V. V. (1999).** The alkaliphilic microbial community and its functional diversity. *Journal of Basic Microbiology,* ***68***, 503-521.

**Zhilina, T. N., Zavarzin, G. A., Detkova, E. N. and Rainey, F. A. (1996a).** *Natroniella acetigena* gen. nov. sp. nov., an extremely haloalkaliphilic, homoacetic bacterium: A new member of *Haloanaerobiales*. *Current Microbiology,* ***32***, 320-326.

# APPENDICES

**Appendix 1: Total diversity environmental data sheet of the samples collected from the hot springs of L. Magadi and Little Magadi**

| Sample name | Bio project ID | Isolation source | Collection date | Geo_loc _name | Barcode Sequence | Barcode Name | Latitude/ Longitude | Linker Primer Sequence |
|---|---|---|---|---|---|---|---|---|
| 1 | PRJNA289092 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGAGTGT | 515Fbar1 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| 2 | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGATCAG | 515Fbar2 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| 3 | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGATCTC | 515Fbar3 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| A | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGTACTC | 515Fbar7 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| B | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGTAGAC | 515Fbar8 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| C | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGTAGTG | 515Fbar9 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| X | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGATGAC | 515Fbar4 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| Y | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGATGTG | 515Fbar5 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| Z | PRJNA289092 | Alkaline Hot Spring | 10/1/2013 | Kenya | GAGTACAG | 515Fbar6 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |

**Appendix 2: Active diversity environmental data sheet of the samples collected from the hot springs of L. Magadi and Little Magadi**

| Sample name | Bio project ID | Isolation source | Collection date | Geo_loc _name | Barcode Sequence | Barcode Name | Latitude/ Longitude | Linker Primer Sequence |
|---|---|---|---|---|---|---|---|---|
| 1 | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGAGTGT | 515Fbar1 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| 2 | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGATCAG | 515Fbar2 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| 3 | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGATCTC | 515Fbar3 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| A | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGTACTC | 515Fbar7 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| B | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGTAGAC | 515Fbar8 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| C | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGTAGTG | 515Fbar9 | 02 S 036 E | GTGCCAGCMGCCGCGGTAA |
| X | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGATGAC | 515Fbar4 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| Y | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGATGTG | 515Fbar5 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |
| Z | PRJNA290229 | Alkaline Hot Spring | 1-Oct-13 | Kenya | GAGTACAG | 515Fbar6 | 01 S 036 E | GTGCCAGCMGCCGCGGTAA |

## Appendix 3: Absolute abundance of KEGG Pathway (KO) within the samples

| KEGG Pathway (KO) | Sediment | Mats | Water |
|---|---|---|---|
| Signaling molecules and interaction | 11 | 13 | 8 |
| Immune diseases | 9 | 19 | 13 |
| Excretory system | 5 | 21 | 25 |
| Substance dependence | 10 | 32 | 12 |
| Circulatory system | 6 | 21 | 33 |
| Development | 11 | 46 | 18 |
| Digestive system | 16 | 41 | 23 |
| Cell communication | 21 | 46 | 21 |
| Nervous system | 14 | 58 | 41 |
| Transcription | 21 | 52 | 40 |
| Environmental adaptation | 10 | 46 | 68 |
| Neurodegenerative diseases | 22 | 47 | 80 |
| Endocrine and metabolic diseases | 12 | 51 | 87 |
| Immune system | 26 | 67 | 64 |
| Transport and catabolism | 34 | 94 | 127 |
| Biosynthesis of other secondary metabolites | 55 | 155 | 106 |
| Cell motility | 76 | 200 | 81 |
| Glycan biosynthesis and metabolism | 83 | 191 | 111 |
| Cell growth and death | 79 | 213 | 105 |
| Endocrine system | 75 | 167 | 185 |
| Metabolism of terpenoids and polyketides | 88 | 212 | 142 |
| Cancers | 78 | 178 | 202 |
| Replication and repair | 160 | 275 | 159 |
| Metabolism of other amino acids | 115 | 253 | 247 |
| Lipid metabolism | 146 | 300 | 269 |
| Xenobiotics biodegradation and metabolism | 161 | 316 | 257 |
| Folding, sorting and degradation | 64 | 321 | 375 |
| Infectious diseases | 100 | 284 | 460 |
| Nucleotide metabolism | 224 | 514 | 304 |
| Membrane transport | 269 | 635 | 275 |
| Metabolism of cofactors and vitamins | 225 | 714 | 289 |
| Signal transduction | 336 | 745 | 500 |
| Overview maps | 315 | 842 | 530 |
| Energy metabolism | 262 | 887 | 566 |
| Amino acid metabolism | 357 | 821 | 673 |
| Translation | 135 | 655 | 1421 |
| Carbohydrate metabolism | 414 | 1131 | 682 |

**Appendix 4: Absolute abundance of Metacyclic Pathway within the samples**

| Metacyclic pathways | Sediment | Mats | Water |
|---|---|---|---|
| Acid resistance | 0 | 5 | 3 |
| Alcohols degradation | 45 | 85 | 56 |
| Aldehyde degradation | 13 | 26 | 12 |
| Amines and polyamines biosynthesis | 195 | 396 | 237 |
| Amines and polyamines degradation | 218 | 472 | 354 |
| Amino acids biosynthesis | 500 | 1275 | 840 |
| Amino acids degradation | 452 | 1101 | 840 |
| Aminoacyl-tRNA charging | 69 | 207 | 125 |
| Antibiotic resistance | 43 | 100 | 85 |
| Aromatic compounds biosynthesis | 67 | 168 | 126 |
| Aromatic compounds degradation | 345 | 814 | 611 |
| Arsenate detoxification | 0 | 1 | 1 |
| Other biosynthesis | 1593 | 4107 | 2547 |
| C1 compounds utilization and assimilation | 256 | 671 | 432 |
| Carbohydrates biosynthesis | 379 | 927 | 577 |
| Carbohydrates degradation | 253 | 642 | 431 |
| Carboxylates degradation | 132 | 388 | 259 |
| Cell Structures biosynthesis | 474 | 1247 | 801 |
| Chemoautotrophic energy metabolism | 13 | 42 | 10 |
| Chlorinated compounds degradation | 163 | 435 | 288 |
| Cofactors biosynthesis | 750 | 2017 | 1159 |
| Cofactors degradation | 0 | 1 | 0 |
| Degradation/utilization/assimilation | 1120 | 2787 | 1797 |
| Other degradation | 195 | 463 | 357 |
| Detoxification | 222 | 593 | 409 |
| Electron transfer | 21 | 125 | 31 |
| Fatty acid and lipids degradation | 163 | 363 | 260 |
| Fatty acids and lipids biosynthesis | 1092 | 2900 | 1784 |
| Fermentation | 163 | 441 | 308 |
| Energy metabolism | 669 | 1736 | 1120 |
| Glycolysis | 80 | 201 | 185 |
| Hormones biosynthesis | 788 | 2030 | 1207 |
| Hormones degradation | 595 | 1522 | 891 |
| Inorganic nutrients metabolism | 122 | 277 | 170 |
| Metabolic regulators biosynthesis | 17 | 28 | 32 |

| Metacyclic pathways | Sediment | Mats | Water |
|---|---|---|---|
| Methanogenesis | 138 | 322 | 185 |
| Methylglyoxal detoxification | 13 | 26 | 12 |
| Nucleo-biosynthesis | 190 | 420 | 322 |
| Nucleo degradation and recycling | 136 | 309 | 208 |
| Other energy | 10 | 21 | 13 |
| Other biosynthesis | 158 | 404 | 244 |
| Pentose-phosphate-cycle | 9 | 42 | 22 |
| Photosynthesis | 50 | 159 | 102 |
| Polymer degradation | 84 | 166 | 133 |
| Protein degradation | 2 | 6 | 1 |
| Respiration | 415 | 1090 | 697 |
| Secondary metabolites biosynthesis | 1045 | 2653 | 1685 |
| Secondary metabolites degradation | 714 | 1884 | 1193 |
| Siderophore biosynthesis | 2 | 17 | 37 |
| TCA cycle | 91 | 253 | 205 |

## Appendix 5: DNA Extraction Reagents

- Solution 1
    - 50 mM Tris pH 8.5
    - 50 mM EDTA pH 8.0
    - 25 % Sucrose solution
- Solution 2
    - 10 mM Tris pH 8.5
    - 5 mM EDTA pH 8.0
    - 1 % SDS
- Lysozyme 20 mg/ml
- RNase A 20 mg/ml
- Proteinase K 20 mg/ml
- Phenol
- Chloroform
- Absolute ethanol.
- 3 M NaCl
- Isopropanol

## Appendix 6: RNA Extraction Reagents

- TRIZOL LS reagent
- Chloroform
- Isopropanol
- Ethanol
- Glycogen or GlycoBlue
- RNAse free water