ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Metagenomic Exploration of Ethiopian Soda Lake Sediments: Microbial Diversity and Community Structure

By

Mesfin Tafesse Gemeda

A Thesis Submitted to

The Department of Microbial, Cellular and Molecular Biology and

Presented to the School of Graduate Studies of the Addis Ababa University in Partial
Fulfillment of the Requirements for the **Degree of**
DOCTOR OF PHILOSOPHY in Biology (Applied Microbiology)

July, 2014

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Metagenomic Exploration of Ethiopian Soda Lake Sediments: Microbial Diversity and Community Structure

By

Mesfin Tafesse Gemeda

A Thesis Presented to the School of Graduate Studies of the Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology (Applied Microbiology)

Approved by Examining Board:

Name	Signature
Dr. Kassahun Tesfaye	(Advisor)
Prof. Francis Mulaa	(Examiner)
Dr. Fassil Assefa	(Examiner)
Dr. Fassil Assefa	(Chairman)

Acknowledgements

The work presented in this PhD thesis has been carried out at the Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Ethiopia. Funding for laboratory bench has been supported at various times from University of Western Cape, South Africa and University of Copenhagen, Denmark. Addis Ababa science and Technology University, my current host University, has also supported my PhD study to persist to the end while paying my fulltime salary.

I am thankful to Dr. Amare Gessesse to support the primary phase of my PhD work on 16S rRNA clone library construction for Chitu and Shalla Lakes through the trilateral collaborative NUFU funded project between Addis Ababa University, Institute of Microbial Biotechnology and Metagenomics (IMBM), University of the Western Cape, South Africa, and University of Bergen, Norway.

I would like to express my deepest gratitude to my supervisor Dr. Kassahun Tesfaye: Your excellent supervision, enthusiasm and positive attitude all through my work. Thanks for your willingness to supervise my PhD and bring it back to a continuation during the spat. Thank you for all your constructive advices and for sharing your scientific knowledge, and also for all fruitful discussions.

I am indebted to all members of the Institute of Microbial Biotechnology and Metagenomics (IMBM), led by Prof Don Cowan University of the Western Cape, South Africa, who have contributed to the work presented in this thesis. I am especially thankful to Prof Don Cowan and Dr. Anna Casanueva for their helpful supervision and fruitful discussions during my laboratory work on construction of 16S rRNA clone libraries at IMBM. I would like to also thank Heide Goodman for her all the technical support during my long stay and laboratory work at IMBM.

I am thankful to Molecular Microbial Ecology group, Department of Microbiology, University of Copenhagen, Denmark: Prof. Soren Sorenson and Prof Lars Hansen for supervising and supporting my laboratory work on the soda lakes 454 pyrosequencing, and Fosmid library construction and Illumina sequencing of fosmid clones; Michael Roggerbuck for his valuable contributions to the laboratory and data analysis technical support of this study; and the rest of the group in Dr. Waleed Abu Al-Soud, Asker Brejnrod, Stefan Milani, Karin Vestberg, and Anette Lot for their enormous technical support during my intensive laboratory work there.

I am very grateful to Dr. Michael Richter from Max Planck Institute for Marine Microbiology, Microbial Genomics and Bioinformatics Research Group, Germany for his assistance in the analysis of Illumina sequences of Fosmid library.

I would also like to acknowledge Dr. Fasil Assefa for plying a huge role in making this PhD work to endure to the last. I thank him for his helpful scientific advices and for proofreading the thesis despite his busy times for office duties.

I am very thankful to my bosses and colleagues at Addis Ababa Science and Technology University: Dr. Tarekegn Tadesse, Dr. Nurelegne Tefera, Dr. Kurkura Kebeto, Dr. Dhugassa Mulugeta, Dr. Habtamu Itefa, Dr. Leakemariam Berhe, Ato Shambel Aregay, Ato Israel Tesema, W/o Meseret Bekele, W/t Zinash Asefa, W/t Kidest Wondimu, for your heartening words and reliable care during the PhD write up. You have always been source of my enjoyable, encouraging, supportive and good friends.

My warmest appreciation goes to my friends Ato Geremew Tasew and Pr. Temesgen Bulti: for your regular encouragement to the end of this work. Thank you for standing at my sides and for the joyful times we have had together. My heartfelt gratirtude also goes to Dr. Dereje Beyene for his all the advices and encouragement regarding the dissertation, and for always having the latest knowledge to share and communicate.

To my father, Tafesse Gemeda and my mother, Shumete Gudina, living in Gimbie: You are the root of everything to me. Your love, support, endurance, and undrestanding through all the years of my study are costly, and have promoted me to this heigh. This PhD study is dedicated to you.

To my family: thank you for your encouragement and love, and for all your tolerance during my busy times. Thank you, my dear wife, Haki, for your encouraging support at all times of frustration, and for your everlasting love and care. This PhD is the fruit of your strength and endurance during turbulences, and your dictated care for me and our children all the way through. And to my two daughters, Bethany and Harerta, thank you for opening the most thrilling new gateways into my life. This PhD study is also dedicated to you.

Abstract

It is estimated that less than 1% of all known environmental bacteria are culturable. It is currently believed that the uses of genetic approaches including metagenomics can bypass culturing of microorganisms and allow the identification of novel gene sequences and complete metabolic pathways of unculturable microorganisms directly from the environment.

In this study, 16S- rRNA-based metagenomics was used to investigate the microbial community composition and phylogenetic relationships of Chitu, Shalla and Abijata soda lakes in the Ethiopian Rift Valley. Four 16S- rRNA-clone libraries were constructed for Chitu and Shalla lakes by using environmental DNA from sediment samples and 16S rRNA primers specific to bacterial and actinobacterial domains. In addition, 454-pyrosequencing was used to identify microbial phylogeny and compositions of Chitu, Shalla and Abijata soda lakes using the V3-V4 hypervariable specific regions of the bacterial 16S-rRNA genes with bar-coding strategy. Moreover, metagenomic approaches were used to construct a 40Kb fosmid (pCC1FOS vector) metagenomic DNA library of Chitu Soda Lake. Potential enzymes, metabolic genes and pathways were detected using functional and sequence- based metagenomic.

16S rRNA clone sequences of Chitu and Shalla Lakes were related to low and high G+C Gram-positive bacteria (*Firmictes* and *Actinobacteria*), CFB, and *Protobacteria* of the **γ**-, α-, *ß*- and Δ-***Proteobacteria*** subdivisions. Most of the clones were phylogenetically allied with members of the **γ**-subdivision of *Protobacteria* (accounted for 34–43% of sequences), followed by *Firmicutes* (affiliated sequences at 16–26%). Both lakes were found to harbor biotechnologically and ecologically significant representatives of microbes that belong to the members of the group *Rhodobaca*, *Alkalimonas*, *Nitrinicola*, *Thioalkalivibrio*, *Methylophaga*, *Desulfuromusa*, *Bacillus*, *Clostridium*, *Dethiobacter*, *Cytophagales*, *Actinobacteria*, and *Bacteroidetes*. Barcode- tagged 454 pyrosequencing of prokaryotic community of Chitu, Shalla, and Abijata soda lakes revealed much wider diversity and novel representatives of microbial species that mostly included the methanogenic archaea (*Methanocalculus*), methane-oxidizing (*Methylomicrobium sp.*), Cyanobacteria, photosynthetic purple bacteria (*Halorhodospira*), non-sulfur purple bacterial genera (*Rhodobaca)*, SOB (*Thioalkalivibrio*, *Thioalkalispira, Sulfurimonas*), SRB (*Desulfobacterium*, *Desulfosarcina*, *Desulfurivibrio*, *Desulfonatronovibrio*), *Actonobacteria*, *Anoxybacillus*, *Bacillus*, *Clostridia*, *Nitrospira*, *Paracoccus*, *Rhodobacteraceae*, *Pseudomonas*, *Bacteroidetes*, *Deinococcus-Thermus*, *Planctomycetes*, and *Spirochetes*.

A total of 288 Mb metagenomic DNA Fosmid library size was constructed from Lake Chitu sediments in Ethiopian Rift Valley area. The plate screening of 7, 200 of total Fosmid clones of the lake sediments resulted significant proteolytic, amylolytic, cellulosic enzyme activities. Besides, sequence-based functional analysis of Lake Chitu Fosmid DNA using Hiseq 2000 (Illumina) provided novel Lake Chitu prokaryotes that have adapted to life styles in the harsh soda lake environments. Pathway analysis using KEGG metabolic maps of the Fosmid DNA sequences evidenced that the microorganisms *Burkholderiales, Delftia acidovorans SPH-1*, *Achromobacter*, *Comamonadaceae, Ralstonia,* and *Bordetella*, within the *Betaprotobacteria* in Lake Chitu majorly play roles in many ways of Methane, Nitrogen, Sulfur, and Carbon cycles.

Page

Table of Contents

List of Abbreviations

ARDRA	Amplified rDNA (Ribosomal DNA) Restriction Analysis
ATP	Adenosine Triphosphate
BAC	Bacterial Artificial Chromosom
BLAST	Basic Local Alignment Search Tool
CFB	Cytophaga-Flavobacterium-Bacteroidetes
COG	Clusters of Orthologous Groups of proteins
FACS	Fluorescence-activated cell sorting
Gb	Gigabase
GO	Genome Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
Mb	Megabase
MEGAN	MEtaGenome ANalyser
METREX	Metabolite-Regulated Expression Screening
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PIGEX	Product Induced Gene Expression
SEED	A peer-to-peer environment for genome annotation
SIGEX	Substrate Induced Gene Expression

List of Figures

List of Tables

CHAPTER ONE: INTRODUCTION

1. Introduction

1.1 General Introduction

It is estimated that more than 70 % of the Earth‘s surface is covered by various aquatic ecosystems that are characterized by high biomass and diverse groups of microorganisms. (Zinger *et al*., 2012). Sea water harbors nearly 10^8 prokarotic cells/ml containing a large number and diverse pool of species (Auguet *et al*., 2010),where the global Ocean contains more than 2 x 10^6 bacterial species per ml with vast array of metabolic functions (Curtis *et al*., 2002; Venter *et al*., 2004). On the contrary, relatively fewer number of prokaryotes (2.3 x 10^{26} total estimates number of cells$_)$ in freshwaters and saline lakes given an average density of 10^6 cells/ml (Fry, 1998; Whitman *et al*., 1998).

The soda lakes of the East African Rift valley have been shown to support a dense and diverse population of aerobic, organotrophic, halophilic, alkaliphilic and alkalitolerant representatives of major bacterial and archaeal phyla (Grant *et al*., 1990). Some of these lakes reflect extremely high primary productivity due to the relative high ambient temperatures, high light intensities, high salts, availability of phosphates and continuous access to CO_2 via the $CO2/HCO_3^-/CO_3^{2-}$ equilibrium in these carbonate-rich waters (Grant *et al*., 1990; Jones *et al*., 1998; Zavarzin *et al*., 1999).

Knowledge on microbial diversity of soda lakes is essential to expand the understanding of the existence of life at extreme pH and salinity and the prospects of finding novel useful biomolecules for biotechnological applications. Consequently, there is a considerable interest among researchers and industries to exploit these lakes as sources of enzymes and metabolites (Grant *et al*., 1992; Grant and Heaphy, 2010). Thus, understanding the composition of microbial communities is important in relation to reveal their roles in ecological systems in nature and harnessing their potential for different application.

Microbial taxonomic and diversity studies were primarily based upon morphotype/phenotypic description of the colony they would form on specific culture media. These include cultural, morphological, microscopic, and chemical characteristics of the cell and wall and outer membrane structures (Smibert and Krieg, 1994). Later, the difficulty of culturing microbial species under laboratory conditions was recognized. The fact that only 1% of the bacteria counted under the microscope could be cultivated on solid or in liquid media prompted the concept of *―great plate count anomaly"* (Staley and Konopka, 1985). Thereafter, the advent of molecular biology partly solved this difficulty by the use of genotypic criteria that include DNA–DNA hybridization (DDH), DNA G+C content and 16S rRNA gene sequence analysis for bacterial taxonomy and diversity studies (Tindall *et al*., 2010).

The use of ribosomal RNA (Olsen *et al*., 1986) is becoming popular for the description of microbial taxonomic diversity by means of several methods that comprise, automated rRNA intergenic spacer analysis (ARISA), terminal restriction fragment length polymorphism (T-RFLP), temperature or denaturing gradient gel electrophoresis (TGGE or DGGE) and single-strand conformation polymorphism (SSCP)], FISH (fluorescence in situ hybridization), and by cloning 16S rRNA gene fragments and subsequently sequencing the clones following the Sanger sequencing method (reviewed in Dorigo *et al*., 2005). However, the 16S rRNA gene also exhibits several limitations as a taxonomic marker including: the presence of nucleotide variations among multiple rRNA operons in a single genome (Rainey *et al*., 1996; Acinas *et al*., 2004) and the possibility of 16S rRNA genes being acquired by horizontal gene transfer that may distort relationships between taxa in phylogenetic trees (Jain *et al*., 1999).

Quite recently, it was revealed that only 1% of the bacterial counts under the microscope are culturable and the majority couldn't be identified from environmental samples. Later on, it was recognized that characterization of microorganisms to their taxonomic groups couldn't be achieved with a few phenotypic characteristics that necessitated the characterization with as many characteristics as possible using numerical taxonomy. Later, the need to characterize microbial communities has leaped in the area of metagenomics and the advent of new-generation sequencing techniques (NGS), leading towards a high-throughput description of microbial community at a higher level of details from environmental samples than cloning or previous sequencing methods on pure cultures (MacLean *et al*., 2009). However, it was also realized that the genetic methods per se could not exhaustively reveals microbial diversity and function of a given environment. Consequently, polyphasic methods of phenotypic, chemotaxonomic and genotypic characteristics are currently used (Tindall *et al*., 2010), in that DNA from environmental samples could be characterized and compared with DNA of already known microorganisms in order to get better insight into microbial diversity and to identify novel compounds for different applications.

The Ethiopian Rift Valley contains a number of fresh and soda lakes that include Koka, Beseka, Gemari, Abe, Ziway, Langano, Abijata, Chitu and Shalla, Awassa, Abaya, Chamo and Chew-Bahir. Many of them have only been characterized with much focus on geochemical, limnological, fish and algal compositions. (Kebede *et al*., 1994; Tilahun and Ahlgren, 2010). A few studies have been carried out on the microbial functions of Ethiopian Rift Valley lakes using culture-based/phenotypic techniques. These include isolation and characterization of isolates (*Mycobacterium* and *Bacillus sp*.) for xylanase and protease enzymes Lake Arenguadie (Gessesse and Gashe, 1997a; 1997b); *Bacillus spp*. for xylanases (Gessesse, 1998; Mamo and Gessesse, 2000; Mamo *et al*., 2006); alkaline proteases (Gessesse *et al*., 2003; Haile and Gessesse, 2012); amylase (Martins *et al*., 2001), and glucanase (Minig *et al*., 2009) from Arenguadie, Chitu, Shalla and Abijata Lakes. In another study, Delgado *et al*., (2006) reported a new isolate, *Nesterenkonia aethiopica* sp. Nov Strain DSM 17733, an alkaliphilic, moderately halophilic, Gram-positive, strictly aerobic, non-motile, non-endospore-forming bacterium was identified from a soiled feather sample collected on the shore of Lake Abjata. Although partial, the prokaryotic and eukarytic diversity, community structure and biogeography of these lakes were also recently described using amplicon- and shotgun sequencing of mRNA transcripts (Lanzen *et al*., 2013).

However, given the enormous divergent potential of the microbes from Ethiopian soda lakes, studies were still limited in scope with respect to modern approach-based microbial diversity studies. These require detailed genotypic approaches through the use of metagenomics and high throuput sequeuencing techniques to realize true diversity and functions from the huge reservoir of uncultivated bacteria. Thus, further culture-independent techniques and other polyphasic approaches are required to provide new insights into the roles and functions of these microbes in the ecosystem and the potential for bioactive genes of some important Ethiopian Rift Valley soda lakes.

1.2 Objectives of the Study

The aim of this PhD study was to assess the microbial diversity and functions of some of the Ethiopian Rift Valley soda lakes (Chitu, Shalla, and Abijata lakes) using genotypic approaches that include 16S rRNA gene–based library, high throughput sequencing, and functional metagenomics.

1.2.1 General Objectives

- ✓ To study the diversity of microbial populations of the soda lakes from their sediments using 16S rRNA genes.
- ✓ To clone the genomic pool (metagenome) of the prokaryotic community and by using expression and sequence strategies for mining novel bacterial bioactive genes and gene products.

1.2.2. Specific objectives

- ✓ Isolation of pure, intact high molecular weight DNA from Lake sediments
- ✓ Construction of 16S rRNA gene–based library using metagenomic DNA
- ✓ Construction of a Fosmid large insert metagenomic DNA library from Lake sediments
- ✓ Functional screening and Sequence characterization of bioactive genes using the Fosmid DNA library

CHAPTER TWO: Literature Review

2. Literature Review

2.1 Alkaline and Saine Environments

Microbial life is said to have originated 3.5 billion years ago and at present it is estimated that about 5 x 10 ^{30} prokaryotic cells inhabit our planet (Schopf *et al*., 2002; Whitman *et al*., 1998). That constitutes the huge and richest reservoir of genetic diversity on earth (Wooley *et al*., 2010). The diversity of microbial communities is expressed by their great variation in modes of energy generation and metabolic functions. These features allow them to thrive profusely in almost all habitats across a great range of natural, polluted and extreme environments.

From the anthropocentric view, living organisms don‗t always exist in conditions or habitats that would seem very ‗normal‗ to us and their occupancy on earth are far from being imagined and more inclusive than previously thought (Rampelotto, 2009). Under normal circumstances, extreme environmental conditions which include physical extremes (e.g., temperature, radiation and pressure) and geochemical extremes (e.g., desiccation, salinity, pH, and redox potential), may destroy the cellular molecules that compose most living organisms, making their survival and development difficult (Rothschild and Mancinelli, 2001). However, many microorganisms survive and flourish in hostile physical and chemical environmental conditions of temperature, pressure, drought, salinity, and pH. Microorganisms are found to thrive at 6.7 km depth inside the Earth‗s crust, and more than 10 km deep inside the ocean at pressures of up to 110 Mpa; from extreme acid (pH 0) to extreme basic conditions (pH 12.8); and from hydrothermal vents at 122 ^{0}C to frozen sea water at −20 ^{0}C (Rodrigues *et al*., 2009; Takai *et al*., 2008).

They are believed to possess unique habitats, compounds and metabolic processes that enable them to live in such extreme conditions. Such organisms are termed ‗Extremophiles‗, that encompass a wide variety of organisms from all three domains of life (eukarya, bacteria and archaea). These organisms belong to seven main categories based on their optimal growth in hostile environments such as of temperature (-2^{0}C to 15^{0}C; Psychrophiles; and 60^{0}C to 115^{0}C; Thermophiles), salinity (2-5M NaCl; Halophiles), pH (< 4 Acidophiles and > 9; Alkaliphiles), anaerobicity (Methanogens), and/or pressure (Barophiles) (Rainey and Oren, 2006). Some of them are not often mutually exclusive and some extremophilic microbes, called polyextremophiles are even able to thrive optimally in multiple types of such harshest conditions in their environment.

These polyextremophiles overcome acidic, superheated sulfur vent (thermoacidophiles), or highly alkaline hypersaline lakes (haloalkaliphiles), or an extremely freezing, high pressure deep ocean. For example, a soda lake is both salty and alkaline, so microbes that live in a soda lake must be both alkaliphiles and halophiles. It is thus apparent that there is an unlimited breadth of environment and amount of habitats on Earth that are suitable for microbial life. These extraordinary physiological capacities of microorganisms to colonize wide variety of extreme environments are coupled with their wide and versatile metabolic diversity (Rainey and Oren, 2006).

2.1.1 Alkaliphiles

Alkaliphiles are microorganisms that grow optimally or very well at pH values above 9, often between 10 and 12, but cannot grow or grow only slowly at the near-neutral pH value of 6.5 (Horikoshi, 1999). They were isolated from various alkaline environment; alkaline spring waters, soda lakes and soda deserts in West USA, East African Rift Valley, and Mongolian and Tibet plateaus (Grant, 2003). Various studies indicated that a diversity of organisms could flourish at pH 10.5 and 12.9 in the soda lakes (Martins *et al*., 2001; Pedersen *et al*., 2004). Alkaliphiles that grow in alkaline habitats that tend to have high concentrations of NaCl are termed as ―Haloalkalophiles" (Gareeb *et al*., 2009).

In order to survive in an alkaline environment, alkaliphiles must make their cytoplasm more acidic to buffer the alkalinity. In addition, enzymes—both excreted and surface located - must be resistant to the effects of extreme pH. Finally, the pH gradient must be reversed to carry out ATP synthesis. Internal pH maintenance is achieved by both active and passive regulation mechanisms in alkaliphilic bacteria (Horikoshi, 2006). Cytoplasmic pools of polyamines and low membrane permeability are two modes of passive regulation, whereas Na^+ channels drive the active regulation (Horikoshi, 2006). Alkaliphilic bacteria compensate for reversal of the pH gradient by having a high membrane potential or by coupling Na^+ expulsion to electron transport for pH homeostasis and energy transduction (Krulwich *et al*., 1998; Horikoshi, 2006).

2.1.2 Halophiles

Halophiles include salt- loving microorganisms that require high salinity conditions for growth. Studies on the two of the largest hypersaline environments; the Great Salt Lake in Utah and the Dead Sea in the Middle East showed that halopiles thrive aquatic habitats of varying salinity, salt marshes, surface salt lakes, subterranean salt lakes, and some other places (Litchfield and Gillevet, 2002). It is known that mild halophiles optimally grow at 0.2–0.85M (1–5%) sodium chloride (NaCl); moderate halophiles grow optimally at 0.85–3.4M (5–20%) NaCl; and extremehalophiles grow optimally at 3.4–5.1M (20–30%) NaCl (DasSarma and Arora, 2002). Examples may include archaeal halophiles such as *Haloarcula*, *Haloferax*, *Halococcus* and *Halobacterium*. Among halophilic bacterium are *Salinibacter rubber*, which are found in saltern crystallizer ponds and others such as *Chromohalobacter beijerinckii* (that were isolated from fermented salted beans) and *Tetragenococcus halophilus* (that were isolated from salted soy sauce) (Anton *et al*., 2002; Tanasupawat *et al*., 2002). A recent metagenomic study of a solar saltern showed the occurrence of a major new phylotype, called nanohaloarchaea, with small cells (< 0.8 um) (Narasingarao *et al*., 2011). Other non-culture based studies showed that novel species similar to haloarchaea may occur in the human gastrointestinal tract (Oxley *et al*., 2010).

Many archaea and bacteria evolved various adaptive mechanisms to cope up with high osmotic pressure they face in the high NaCl environment they inhabit (Galinsky, 1993 and 1994; Oren, 2004). They adjustment their internal osmotic pressure with the outside environment by accumulating high concentration of compatible solutes (K^+, Na^+, Cl^-) or sugars and other organic osmolytes in their cytoplasm, which function as osmoprotectants without interfering with their normal cell functions. Halophilic Archaea accumulate very high concentrations of KCl in their cytoplasm close to that of the surrounding medium (Nieto and Vargas, 2002). The other mechanism of adaptation of halophiles is that their proteins are optimally folded and function under saturated salt concentrations (Michael *et al*., 1999). Here, as a consequence of cytoplasmic salt concentration, proteins from halophiles are rich with negatively charged amino acid residues on their surface and their membrane phospholipids become more anionic (Beales, 2004; Russell *et al*., 1995).

2.1.3 Soda Lakes

Soda lakes and soda deserts are examples of naturally-occurring alkaline saline environments and they are widely distributed around the world (Grant and Sorokin, 2011). Table 2.1 describes the location, alkalinity and salinity of some of the known soda lakes largely distributed across the world. Soda lakes are the most stable of the alkaline aquatic ecosystems characterized by high pH values of 9.0-12.0 and salinity up to saturation concentrations with sodium chloride or other dissolved salts (Grant et *al*., 2006). They are caused by a combination of geological, geographical and climatic conditions. They usually appear more or less permanent closed bodies of water in shallow depressions or craters with a high marginal relief (having sufficient rainfall to sustain streams and ground water entering the basin) formed due to continental rifting and volcanic activities.

Table 2.1: List of Some Examples of Soda Lakes in the World by Country, pH and Salinity. NA indicates 'data not available'. Source: Wikipedia, the free encyclopedia (http://en.wikipedia.org/), 9 February 2014.

Name	Country	pH	Salinity (%)
Africa			
Wadi El Natrun lakes	Egypt	9.5	5
Malha Crater Lake	Sudan	9.5-10.3	NA
Lake Arenguadi (Green Lake)	Ethiopia	9.5-9.9	0.25
Lake Basaka	Ethiopia	9.6	0.3
Lake Shala	Ethiopia	9.8	1.8
Lake Chitu	Ethiopia	10.3	5.8
Lake Abijatta	Ethiopia	9.9	3.4
Lake Ratwe	Uganda	NA	NA
Lake Mahaga	Uganda	NA	NA
Lake Nyamunuka	Uganda	NA	NA
Lake Magadi	Kenya	10	>10
Lake Bogoria	Kenya	10.5	35
Lake Turkana	Kenya	8.5-9.2	0.25

Lake Nakuru	Kenya	10.5	NA
Lake Logipi	Kenya	9.5-10.5	2-5
Lake Sonachi (Crater Lake)	Kenya	NA	NA
Lake Manyara	Tanzania	9.5-10	NA
Lake Natron	Tanzania	9-10.5	>10
Lake Rukwa	Tanzania	8-9	NA
Lake Eyasi	Tanzania	9.3	0.5
Lake Ngami	Botswana		
Rombou Lake	Chad	10.2	2
Asia			
Kulunda Steppe Lakes	Russia	NA	NA
Lake Khatyn	Russia	10	NA
Lake Van	Turkey	9.7-9.8	NA
Lake Salda	Turkey	NA	NA
Lake Urmia	Iran	NA	30
Lonar Lake (Crater Lake)	India	9.5-10.5	1
Sambhar Salt Lake	India	9.5	7
Khyagar Lake	India	9.5	0.6
Tso Moriri Salt Lake	India	9.0	NA
Tso Kar Salt Lake	India	8.8	NA
Lake Surigh Yilganing Kol	Aksai Chin, India	NA	NA
Tso Tang Lake	Aksai Chin, India	NA	NA
Aksayqin Hu Lake	Aksai Chin, India	NA	NA
Lake Hongshan Hu	Aksai Chin, India	NA	NA
Tianshuihai lake	Aksai Chin, India	NA	NA
North Tianshuihai lake	Aksai Chin, India	NA	NA
Kushul lake	Aksai Chin, India	NA	NA
Pangong Salt Lake	India & China	9.4	0.9
Spanggur Tso (Pongur Tso)	India & China	NA	NA
Guozha lake	China	NA	NA
Qinghai Lake	China	9.3	2.2
Namucuo Lake	China	9.4	0.2
Lake Zabuye (Drangyer)	China	10	NA
Taboos-nor	Mongolia	NA	NA
Europe			
Lake Fehér (Szeged)	Hungary	NA	NA
Böddi-szék	Hungary	8.8-9.8	NA
Lake Neusiedl (Fertö)	Austria, Hungary	9-9.3	NA
Rusanda	Serbia	9.3	NA
Kelemen-szék	Hungary	9-9.7	NA
North America			
Mono Lake	USA	9.8	8
Big Soda Lake (Nevada)	USA	9.7	NA
Soap Lake	USA	9.7	0.7
Baldwin Lake	USA	NA	NA

Alkali Lake (OR)	USA	11	NA
Summer Lake	USA	NA	NA
Owens Lake	USA	NA	NA
Borax Lake	USA	NA	NA
Manitou Lake	Canada	NA	NA
Goodenough Lake	Canada	10.2	NA
Lake Texcoco	Mexico	8.8-11.5	8
Lake Alchichica	Mexico	8.9	NA
South America			
Antofagasta Lake	Chile	NA	NA
Australia			
Lake Werowrap	Australia	9.8	4

This type of topology prevents the outflow of water from the lake which raises the alkalinity and salinity of the lake as a result of high rate surface evaporation that precipitates large amounts of insoluble sodium carbonate/soda (or complexes of this salt) with considerable amount of sodium chloride accumulation (Jones *et al*., 1995). Other consequence of high alkalinity of the soda lake is due to the corresponding lack of Ca^{2+} and Mg^{2+} concentrations which would remove carbonate ions as insoluble salts producing an alkaline brine dominated by sodium carbonate or sodium bicarbonate with variable amounts of sodium chloride (Jones *et al*., 1995).

The best-studied soda lakes are located in the Kenyan–Tanzanian part the East Africa Rift Valley. The East Africa Rift Valley stretches in a roughly north-south direction from the Afar Depression, where the system of tectonic troughs joins the oceanic domains of the Red Sea and Gulf of Aden, towards the Zambezi River in Mozambique (Fig. 2.1). The salinity of the lakes rage from approximately 5% total salinity (w/v) to saturation ($\geq$ 30%) with roughly equal proportion of $NaCO_3$ and NaCl as the major salts and pH values ranging from 9 to > 11.5 (Grant, 1992). The Rift Valleys is also characterized by active volcanism having numerous more diluted and less alkaline hot springs in the shores that feed some of the lakes.

Regardless of the currently hostile nature of the Soda lakes, they harbor considerable diverse and dense microbial populations often resulting in distinct coloration (like green, pink, orange or red) of the lake water. This is reflected in some lakes by extremely high primary productivity presumably due to the relative high ambient temperatures, high light intensities, high salts, availability of phosphates and continuous access to CO_2 via the $CO2/HCO_3^-/CO_3^{2-}$ equilibrium in these carbonate-rich waters (Zavarzin *et al*., 1999). However, all soda lakes are not extremely productive and many factors such as nutrient limitations, toxicity, or trophic interactions mechanisms may control their primary productivity (Zavarzin and Zhilina, 2000).

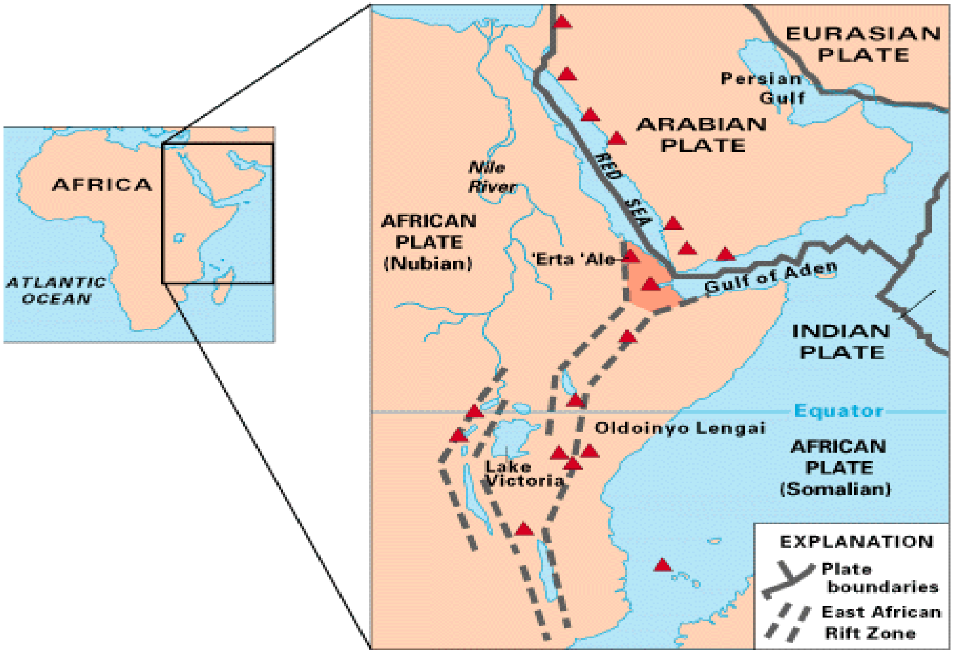


Figure 2.1: Map showing the location of the East African Rift Valley System (Source: http://en.wikipedia.org/wiki/East_Africa_Rift).

2.1.4 Ethiopian Rift Valley Soda Lakes

The Ethiopian Rift is part of the East African Rift system (Fig. 2.1), which extends from the Kenyan border in the south up to the Red Sea in the north. It divides the highlands of central Ethiopia and contains a number of fresh and soda lakes that are generally located in three basins: Awash basin (lakes Koka, Beseka, Gemari, Abe, etc.), the central Main Ethiopian Rift lake areas (lakes Ziway, Langano, Abiyata, Chitu and Shala) and southern basin (lakes Awassa, Abaya, Chamo and Chew-Bahir). The lakes rift ecosystem (Fig. 2.2) is well recognized by its geothermal potential.

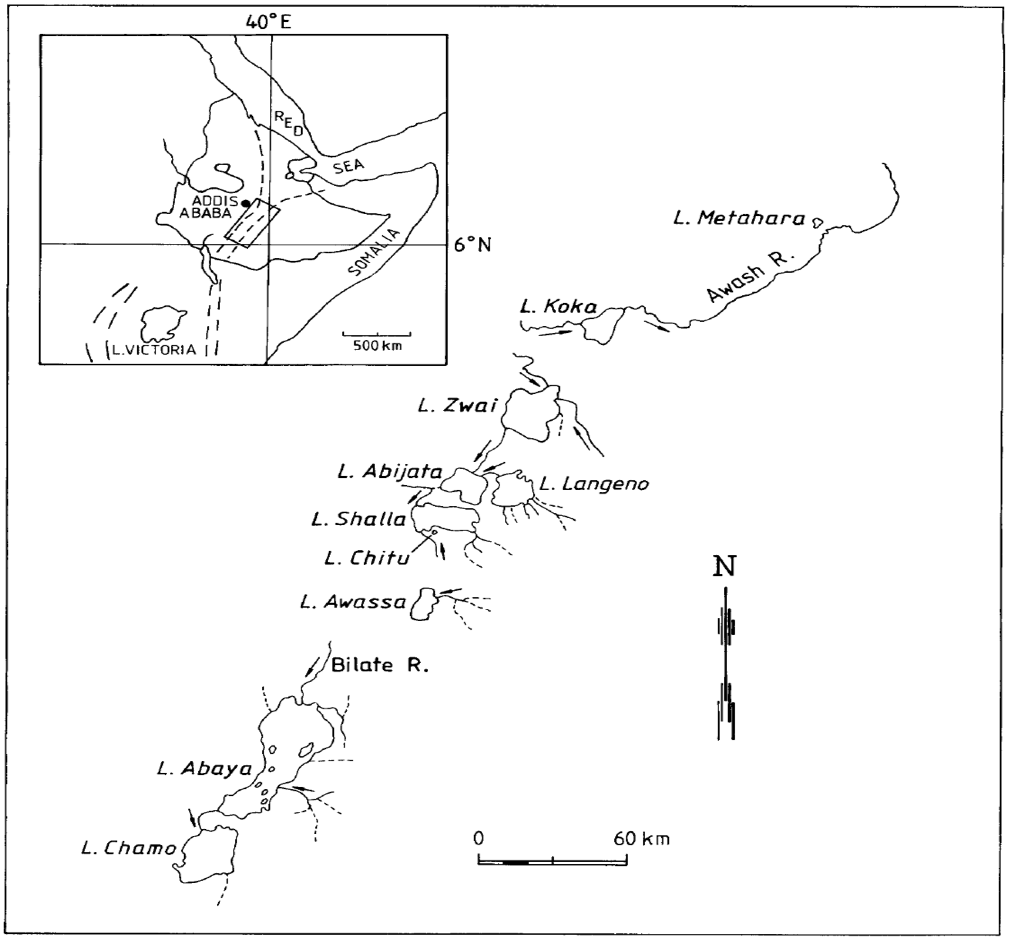


Figure 2.2: Location Map of Ethiopian Rift Valley Lake Areas (From Kebede *et al.*, 1994).

Hydrologically, the lakes form separate units by occupying the volcan-tectonic depressions and fed by local rivers that are originated from the eastern and western highlands. Some of the soda lakes including Lake Abijata, Lake Shalla, and Lake Chitu are highly productive, whose shores support one of the largest populations of flamingos (personal observation).

Recently, due to climatic and other anthropogenic factors, hydrochemical setting of some of the lakes have been changed and their levels have been reduced; indicating the need to apply serious conservation and ecosystem management to protect the perturbance of the rich biodiversity and biotechnological resources of these soda lakes (Kebede *et al.*, 1994; Ayenew, 2007).

2.2 Microbial Diversity in Soda Lakes

Microorganisms of the soda lake environments should undoubtedly be alkalipiles (alkali-loving) or at least alkali-tolerant with the additional competence for growing under high levels of NaCl (halophiles or salt-loving); as such called haloalkaliphiles. The occurrence of haloalkaliphiles was probably first mentioned by Tindall in 1984. He described two new genera; *Natronobacterium* gen. nov. and *Natronocococcus gen.* nov. Further, Bivin and Stoeckenius (1986) isolated 51 haloalkaliphilic strains from alkaline salt lakes in Kenya and the Wadi Natrun in Egypt. The isolates contained retinal pigments identical to halorhodopsin of *H. halobium.* There after, a lot of studies were initiated and many members of the bacterial and archaeal communities were reported of the soda lakes and elsewhere. Table 2.2 provides the summary of some of the major and common microrganisms of diverse metabolic groups identified from East Africa Soda lakes and other soda lake environments using culture-based and culture independent approaches (Antony *et al*., 2013).

Table 2.2: Major Diverse Bacterial and Archeal Groups Detected in World Soda Lakes across Africa, North America and Eurasia as Revealed by Culture- dependent and Culture- independent Studies (Modefied from Antony *et al*., 2013).

Heterotrophs	**Phototrophs**	**Methanogens & Methanotrophs**	**Acetogens**	**Sulfur oxidizers & Sulfate Reducers**	**Nitrifiers & Denitrifiers**
Halomonas *Spirochaeta* *Bacillus Alcaligenes* *Rhodobaca* *Methanosalsum* *Arthrobacter* *Nitrincola* *Alkalilimnicola* *Marinospirillum* *Alkalimonas* *Idiomarina* *Halorubrum* *Halobiforma* *Natronococcus* *Natronolimnobius* *Halobiforma* *Natronococcus* *Natronolimnobius*	*Ectothiorhodospira* *Synechococcus* *Synechocystis* *Rhodobaca* *Thiocapsa* *Roseinatronobacter*	*Methylomicrobium* *Methylobacter* *Methanolobus* *Methanosalsum*	*Tindallia* *Clostridium*	*Thioalkalivibrio* *Thioalkalimicrobium* *Desulfonatronovibrio* *Desulfonatronum*	*Nitrosomonas* *Nitrobacter* *Pseudomonas* *Paracoccus*

Amongst the microorganisms, the Cyanobacteria that include *Arthrospira* and *Cyanospira* species were found to be dominant in these lakes, play important roles as major photosynthetic primary producers (Grant *et al*., 1999; Jones *et al*., 1998). The anoxygenic phototrophic purple sulfur bacterial genera, *Ectothiorhodospira* and *Halorhodospira* in hypersaline lakes, plus Cyanobacteria also have unrestricted functions for primary productivity that may be the source of significant amount of organic compounds for the support and growth of other heterotophic biomass in the soda water bodies (Grant *et al*., 1999; Jones *et al*., 1998; Jones and Grant, 2000).

The phototrophic purple sulfur bacterial species, *Ectothiorhodospira* and *Halorhodospira* in the hypersaline lakes (where oxygen is very reduced) and *Thiorhodospira* and *Thioalkalicoccus* as strictly anaerobic in low saline alakaline lakes (Bryantseva *et al*., 1999a) oxidize H_2S and elemental sulfur under anoxic condition as photosynthetic electron donors. *Rhodobaca spp*. (Boldareva *et al*., 2008) and *Rubribacterium sp*. (Boldareva *et al*., 2009) have also been recorded for primary productivity in various soda lake habitats. A number of soda lake sulfate-reducing bacteria have also been brought into culture, including *Desulfononatronospira thiodismutans* (Sorokin *et al*., 2008a), *Desulfonatronum cooperatum* (Zhilina *et al*., 2005), *Desulfonatronovibrio hydrogenovorans* (Zhilina *et al*., 1996), *Desulfonatronum lacustre* (Pikuta *et al*., 1998), and recent isolates (*Dethiobacter* and *Desulfurivibrio spp*.) reported by Sorokin *et al* (2008b).

Soda lake environments also harbor diverse microbial communities associated with major element cycles including Sulphur and Nitrogen. Haloalkaliphilic sulfur oxidizers including *Thioalkalivibrio*, *Thioalkalimicrobium* and *Thioalkalispira* species have been described from Lake Hadyn in the Transbaikal region (Sorokin *et al*., 2000b) and also recently from sediments of northeastern Mongolian soda lakes (Sorokin *et al*., 2004). Among these, the genus *Thioalkalivibrio* is most diverse phylogenetically and its representatives are physiologically the best adapted for life in hypersaline brines that also potentially mediate the denitrification process (Sorokin *et al*., 2003).

Sulfate-reducing bacterial studies in Siberian soda lakes also showed the presence of phylotypes related to the orders *Desulfovibrionales* and *Desulfobacterales* (Foti *et al*., 2007). Several lithotrophic sulfate-reducing bacteria of the genera *Desulfonatronum*, *Desulfonatronovibrio*, *Desulfonatronospira* and several heterotrophic sulfate- reducing bacteria affiliated to *Desulfobotulus alkaliphilus*, *Desulfobacteraceae*, *‗Desulfobulbus alkaliphilus'* and *Synthrophobacteraceae* have been isolated from soda lakes in Kulunda Steppe (reviewed in Sorokin *et al*., 2011). The presence of both autotrophic sulfur oxidizers and heterotrophic sulfur reducers in soda lakes possibly indicates their functional connections through carbon and sulfur cycles (Drake *et al*., 2008).

Lin *et al* (2004; 2005) isolated many of the methanotrophs (Methane-utilizing microbes) affliated to *Methylomicrobium*, *Methylobacter*, *Methylomonas, Methylothermus* and *Methylocystis* species from soda lakes of Transbaika, Suduntuiskii, Torom, Mono lake, and Gorbunka. In another study, a pure culture of an obligate methanotroph isolate was obtained from a Kenyan soda lake after enrichment at 0.6 M total Na^+ and pH 10 which was identified as a member of the genus *Methylomicrobium* in the *Gammaproteobacteria* (Sorokin *et al*., 2000a) and later classified as *M. kenyense* (Kalyuzhnaya *et al*., 2008). The ecological role of the methanotrophs isolated from Transbaikal and East African soda lakes might not be limited only by methane oxidation, since the strains were found to possess an additional potential to oxidize ammonia (to nitrite) and carbon disulfide (CS_2) (Khmelenina *et al*., 2000; Sorokin *et al*., 2000a) it is likely that they constitute an important functional link between C, N and S cycles in soda lake ecosystems (Trotsenko and Khmelenina, 2002).

The nitrogen cycle in soda lakes involves the production of ammonia by fermentative anaerobes such as *Tindallia magadiensis* (Sorokin *et al*., 2002). The most important groups of organisms, involved in the nitrogen cycle are the lithoautotrophic ammonia-oxidizing bacteria and the lithoautotrophic nitrite-oxidizing bacteria. The *in situ* culturing procedures of ammonia oxidation and the presence of taxonomic 16S rRNA gene sequences revealed that the nitrifiers and denitrifiers such as *Nitrosomonas* and *Nitrobacter* play important roles in the nitrogen cycle in soda lakes (reviewed by Grant and Sorokin, 2011) (Table 2.2). Although nitrification is undertaken by the nitrifying bacteria, *Nitrosomonas* and *Nitrobacter/Nitrospira*, the occurrence of important heterotrophic denitrifiers, such as *Halomonas*, *Pseudomonas* and *Paracoccus spp*. (reviewed by Antony *et al*., 2013) (Table 2.2), in soda lakes illustrate that the N and C cycles of these habitats are also interconnected under anoxic conditions.

According to Duckworth *et al* (1996), culture dependent studies have also revealed several strains of aerobic, heterotrophic alkaliphilic, and haloalkaliphilic organotrophs from a range of soda lakes in the East African Rift Valley. These included many proteobacteria such as the species of *Halomonas*, *Cellulomonas*, *Alkalimonas*, and *Dietzia* (Duckworth *et al*., 2000; Jones *et al*., 2005; Ma *et al*., 2004a; Duckworth *et al*., 2004) and other proteobacteria related to pseudomonads and vibrios (Duckworth *et al*., 1996). Heterotrophic Gram-positive bacterial lineages that belong to both the high G + C Firmicutes dominated by diverse *Bacillus*, the low G + C Actinobacteria dominated by streptomycetes, and many other new genera were also readily isolated from hypersaline brines.

In summary, much of the hetherto microbial diversity studies of the soda lake environments were limited to using traditional culture-baesd and taxonomic procedures. More recently, genetic studies showed the immense diversity of the total prokaryote biotope in soda lake environments using metagenomic, PCR-based 16S rRNA techniques and high throughput sequencing facilities. The water and sediments of the lakes are found to be dominated by microbial groups affiliated with the low G+C Gram-type-positive, *Firmicutes*-related, *Proteobacteria*, *Actinomycetes* and *Bacteroidetes.*

Sequence analyses has revealed unexplored biodiversity novel bacteria and archaea from several East African sites (Grant *et al*., 1999; Rees *et al*., 2004), Inner Mongolian sites (Ma *et al*., 2004b; Pagaling *et al*., 2009), Lake Van in Turkey (Lopez-Garcia *et al*., 2005), Wadi al Natrun (Mesbah *et al*., 2007), and Ethiopian soda lakes (Lanzen *et al*., 2013). Therefor, microbial diversity of soda lakes is not only essential to expand the understanding of the existence of life at extreme pH and salinity but also vital for mining novel useful biomolecules for biotechnological applications.

2.3 Biotechnological Prospective of Soda Lake Microbes

Soda lakes are characterized by extreme physiochemical conditions and thus microbes living in these lakes may possess novel adaptation mechanisms to multiple stresses for survival in high salinity and alkaline pH (Grant *et al*., 1990). Because of these unique properties, microbes dwelling in unexplored and extreme environments such as soda lake have recently prompted considerable interest among scientific communities and industries as sources for valuable biomolecules that encompass industrial enzymes and metabolites. Isolates from soda lakes are found to be sources of extracellular hydrolytic enzymes such as proteinases, amylases, cellulases, and lipases capable of functioning at high pH and possibly high temperature and salt concentration (Grant and Heaphy, 2010). Bioenzymes form such microbial sources are supposed to be eco-friendly, efficient and substitute many less desirable and harmful chemicals for various industrial processes including detergent, textile, leather, pharmaceutical, food, paper, pulp, chemical, and waste treatment.

Several bacterial isolates from soda lakes of different geographical regions have been characterized for their potential for sources of alkalophilic-degrading protease, amylase, lipase, and chitinase enzymes (Jones *et al*., 1998; LeCleir *et al*., 2007; Joshi *et al*., 2008); *Bacillus* and other species are the key prodcers and make up about half of the total industrial enzyme market (Simonen and Palva, 1993). Currently, two different cellulases (IndiAge Neutra and Puradax) derived from Gram-positive bacteria from Kenyan soda lake are marketed for use in laundry and textile processes (Sheridan, 2004; Grant and Heaphy, 2010). Ectoine, an organic solute osmolyte produced by microbes from soda lakes is fund to have valueable biotechnological potential in cosmetics and pharmaceutical industries (reviwed in Pastor *et al*., 2010).

With the advancement in molecular tools and metagenomics approaches, there is a possibility of revealing more novel genes, enzymes and other biomolecules from soda lakes for various applications (Pawaraya *et al*., 2009; reviewed in Moussard *et al*., 2011). In the future the additional aid of improved high-throughput screening techniques and sequencing technology, discovery of many commercially important enzymes and biomolecules from saline and alkaline lake ecosystems could be realized.

2.4 Methods in Microbial Diversity Studies

Microbes with their diverse metabolic and enzymatic attributes control fundamental ecosystem processes such as the biogechemical cycles of elements, and decompose waste materials and pollutants, and hence influence the atmosphere, the climate, animal, and plant health (Whitman *et al*., 1998). The versatility of microbes in ecosystem recycling prompts the importance of understanding microbes at the biochemical and genomic level to exploit their genetic potential for the production of useful materials and products of enzymatic, nutritional, diagnostic, and therapeutic values (Demain, 2000; Ferrer *et al*., 2009). For these purposes, a number of methods have been employed to study and get access to the vast diversity, gene pool and functions of microorganisms present in various environments.

The widely accepted classical method of identifying and accessing microbial diversity requires laboratory cultivation of the organism using standard methods. However, it was realized that most habitats contain thousands of different species, yet only a small fraction of the microbial community can be cultivated under standard laboratory conditions (Handelsman, 2004). This phenomenon was coined the *great plate count anomaly*- the report of the discrepancy between total cell counts using a microscope and viable colony count on nutrient agar plate (Staley and Konopka, 1985).

It is generally accepted that most microbes resist laboratory cultivation that may be due to their otherwise special physiological, physical and chemical conditions in their native habitats and/or specific interdependence with other microbial consortia (Staley and Konopka, 1985). This mostly restricts the use of classical culture methods for the characterization of diversity in microbial communities and led to the advent of culture-independent strategies to obtain almost a complete coverage of the uncultured microbial diversity and access to the biochemical and functional pathways within these uncultivable microorganisms from environmental samples. Some of the currently used culture-independent methods enhanced the analysis of bacterial diversity, structure and functions in various environments during the last two decades. These culture-independent methods, together with environmental sampling revealed that traditional cultivation methods could only detect less than 1% of the total microbial species found in nature (Handelsman, 2004).

2.4.1 Sequencing of 16S- rRNA genes

Various detection methods of bacterial identification that use the 16S rRNA gene have been widely used in microbial ecology to differentiate microorganisms (Bouchet *et al*., 2008). These include fluorescence ***in situ*** hybridization (FISH) (Wagner *et al*., 1993), denaturing-or temperature gradient gel electrophoresis (DGGE, TGGE) (Muyzer *et al.,* 1993), terminal-restriction fragment length polymorphism (T-RFLP) (Liu *et al.,* 1997), quantitative PCR (qPCR) (Zhang and Fang, 2006), microarray (Xia *et al*., 2010), clone library construction (discussed below).

PCR amplification and sequencing of the small subunit ribosomal RNA gene (SSU rRNA or 16S rRNA) has been the principal approach to assess the abundance and taxonomic identity of microbes in the environment. In this regard, the initial attempt was based on direct extraction and sequencing of 5S rRNA molecules from Yellowstone hot spring environmental sample to determine its microbial diversity (Stahl *et al*., 1985).

However, very limited information could be obtained from these molecules because of their variable and short nucleotides (200 bp). This has reduced the importance of these molecules for continued ecological application. Then the 16S rRNA gene was identified as a marker molecule for assessing microbial diversity (Woese, 1987) and employed for phylogenetic profiling of microbial communities (Schmidt *et al*., 1991). Further more, the average full length of 16S rRNA gene (1500 nucleotides) also contains highly conserved sequence domains interspersed with nine variable regions and provides more information than the 5S rRNA gene (Van de Peer *et al*., 1996). The high-sequence conservation of 16S genes allowed a wider analysis of microbial phylogeny and diversity as well as the identification of new taxa.

In general, the significance of 16S rRNA is emanated from (1) the direct analysis of 16S rRNA gene sequences and the diversity of microbes is possible without culturing, (2) they occur in high copy numbers in the microbial cells and (3) the 16S rRNA gene sequences could be generated for large number of bacteria from a variety of environments (Amann *et al*., 1995).

Technically, the 16S rRNA genes can be amplified by polymerase chain reaction (PCR) using oligonucleotide universal primers designed to target the conserved regions within the rRNA genes. The PCR products are then coned into appropriate vectors so as to transform host bacterial strains and generate clone libraries. These 16S rRNA gene libraries sequenced and compared with already known 16S rRNA databases to infer the diversity and phylogeny of community microorganisms in that environment. The 16S rRNA can also be identified by hybridization to specific probes. The application of this method provides a means to investigate the taxonomic composition of microbial population in any environment without the need for cultivation (Amann *et al*., 1995, Olsen *et al*., 1986; Ward *et al*., 1990).

The importance of PCR-based approaches of 16S rRNA amplicon for microbial diversity, albeit, it may alter the inference of microbial richness and evenness in microbial communities due to experimental bias affecting PCR and PCR induced sequence artifacts or chimeric sequences (Acines *et al*., 2005; Hong *et al*., 2009; Engelbretson *et al*., 2010). However, the emergence of new High-Throughput Sequencing (HTS) technologies (e.g. 454 and Illumina) has allowed the analysis of 16S rRNA for the exploration of microbial diversity at a remarkable level (Logares *et al*., 2012).

Recently, both sequencing technologies have gained superiority in microbial ecology for their reduced cost and efficiency while their performances and biases are investigated in parallel (Claesson *et al*., 2010; Haas *et al*., 2011; Minoche *et al*., 2011; Logares *et al*., 2013). The use of 454-pyrosequencing (short length reads, ~ 450bp) is undertaken by sequencing ribosomal DNA gene (rDNA) amplicons from environmental samples (Sogin *et al.,* 2006). Illumia sequencers (shorter length reads, ~ 200bp) also use PCR amplicons to explore natural composition of microbial population (Caporaso *et al*., 2012; Werner *et al*., 2012; Bokulich *et al*., 2013).

2.4.2 Metagenomics

Metagenomics is a culture-independent approach (doesn‘t depend on isolating microbes in pure cultures) which involves genomic analysis of DNA extracted from environmental samples. It is thus a much broader tool to uncover microbial genetic pool and many new microbial species compared to classical and culture-dependent techniques (Handelsman *et al*., 1998). It involves the direct isolation of entire mixture of environmental DNA from a defined habitat, followed by cloning into a suitable vector, transforming the clones into a host bacterium (mostly *E. coli*) to facilitate the analysis of the functions and the sequences within these clone libraries (Fig. 2.3 and 2.4).

The DNA library clones can be screened for phylogenetic markers or ―anchors,” such as 16S rRNA and *recA*, or for other conserved genes by hybridization or multiplex PCR (Stein *et al*., 1996). They may be used for expression of particular phenotypes, such as enzyme activity or antibiotic production (Schloss and Handelsman, 2003) and for random sequencing (Tyson *et al*., 2004).

Metagenomic methodologies, therefore, allow for the direct retrieval of phylogenetic and functional information from the DNA isolated from environmental samples. Researchers could examine microorganisms in the context of their natural environments they inhabit and to study both the phylogeny and putative functional genes of entire communities at the same time (Bohannon, 2007).

2.4.2.1 Preparation of Metagenomic Libraries

The preparation of metagenomic library construction begins from gathering of natural environmental samples (such as water, soil, sediments, fecals, biofilms samples) and extraction of DNA (unselective) from the samples (Fig. 2.3). In some cases samples are pretreated before making any further analysis. These may involve enrichment of cultures (targeted) by adding nutrients *in situ* to get improved quantity and quality of DNA to stimulate metabolic activity of the microbes in natural habitat (such as samples from extreme environments and scant microbes).

Graham *et al* (2011) successfully expressed archaeal recombinant cellulase enzyme gene by enriching three archaeal species collected from a geothermal source on lignocellulose at 90 ^{0}C. Other pretreatment strategies focus on enriching metabolic active microbes by growing them on selective substrates that are labeled with isotopic ^{13}C, ^{15}N and ^{18}O, called stable isotope probing (SIP) (Dumont and Murrell., 2005). This technique separates the heavier DNA of the enriched cultured microbes from the total pool of metagenomic DNA upon density centrifugation. Pretreatment strategies, however, can at the same time reduce microbial diversity just like anthropogenic factors that have impact on the community composition (Ferrer *et al.*, 2012; Nimchua *et al.*, 2012).

Various protocols for metagenomic DNA extraction methods are employed and a variety of commercial products or extraction kits are available depending on the type of metagenomic samples**.** For example**,** membrane filters for environmental water samples, are used to filter large debris at start and finally the filtration of single cell microorganisms can be achieved on small pore- sized membrane filter. Solid samples are treated by direct *in situ* lysis or indirect lysis which employs previous separation of membrane matrix-adhering cells. The lysis procedure is performed by physical treatment that may include ultrasonication, pestle or bead beating homogenization techniques and centrifugation methods (reviewed in Daniel, 2005; Robe *et al.*, 2003).

Chemical procedures commonly performed by using ion-exchange resins for cation removal, chelating agents such as ethylenediaminetetraacetic acid (EDTA), cetyltrimethyl-ammonium bromide (CTAB), polyvinylpolypyrrolidone (PVPP), anionic surfactants such as sodium dodecyl sulfate (SDS) for protein denaturation, and addition of highly concentrated NaOH for alkaline lysis (Miller *et al.*, 1999; Zhou *et al.*, 1996).

Enzymatic methods usually depend on using Proteinase K for protein digestion or lysozyme for the hydrolysis of the peptidoglycan backbone of bacterial cell walls, especially from Gram-positive bacteria. Different laboratory methods have been used to retrieve DNA from soil and sediment samples. Protocols for the isolation of high-molecular weight metagenomic DNA from different environmental samples (e.g. soil, sediments, alkaline samples and hot springs) are available (Verma and Satyanarayana, 2011). However, there is no single method which is universally applicable for isolation and purification of soil metagenomic DNA. This is mainly due to the fact that DNA extraction methods need optimization for every type of soil sample and different objectives of metagenomic works (Zhou *et al*., 1996; Lakay *et al*., 2007).

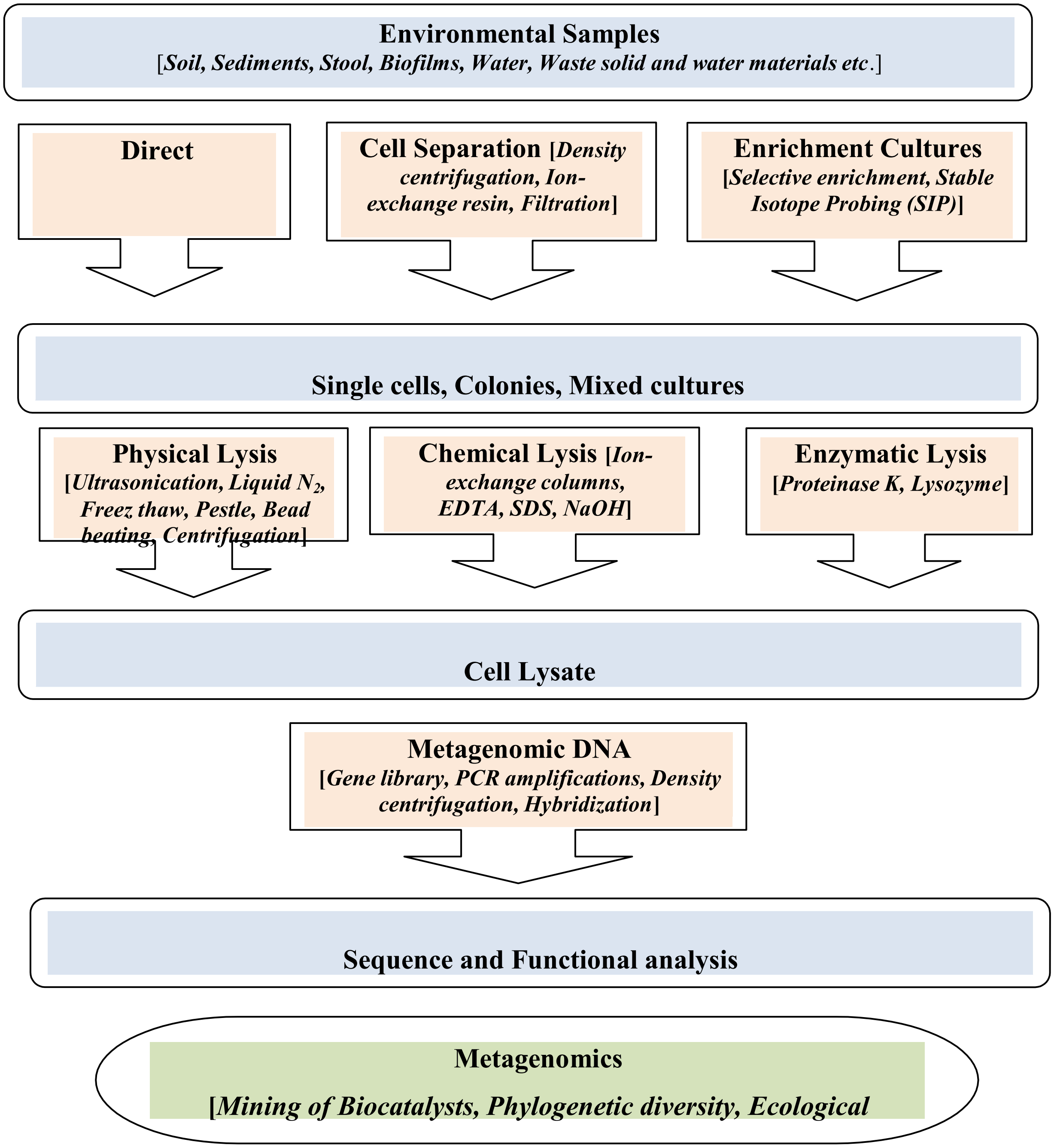


Figure 2.3: Major Metagenomic Preparation Steps from sample collection to DNA retrieval (Modified from Leis *et al*., 2013).

The other step in metagenomic preparation process is the construction of metagenomic libraries from extracted environmental DNA (Fig. 2.4). The outcome of a metagenomic library mainly relies on the quality of the DNA isolated and it should be suitable for subsequent downstream molecular biology processes such as enzymatic, cloning, ligation, hybridization and sequencing procedures. The determinant factors are purity from contaminants, fragment size and degree of mechanical shearing, quantity of the DNA, and other PCR-related, and sample bias. Humic compounds are major contaminants of DNA from soil and sediment samples (Wang *et al*., 2012). Humic substances include: i) humic acid, which is soluble in alkali and insoluble in acid, ii) fulvic acid, which is soluble in alkali and acid and iii) humin, which is insoluble in both alkali and acid (Senesi and Loffredo, 2001). Humic acids affect almost all molecular biological methods such as hybridization, restriction digestions of DNA, PCR and bacterial transformation (Tebbe and Vahjen, 1993). Therefore, in most of the soil and sediment metagenomic projects, the bottleneck remains the isolation of enough intact large molecular weight metagenomic DNA without humic substances.

Several strategies have been developed for the purification of soil metagenomic DNA. Caesium chloride density gradient centrifugation is a widely used and an efficient strategy for the purification of DNA from contaminants (Tien *et al*., 1999). However, due to longer processing time, this method is not suitable for purification of multiple samples. Recently, several simple and rapid purification methods have been reported for the successful removal of contaminants from metagenomic DNA. These methods include the

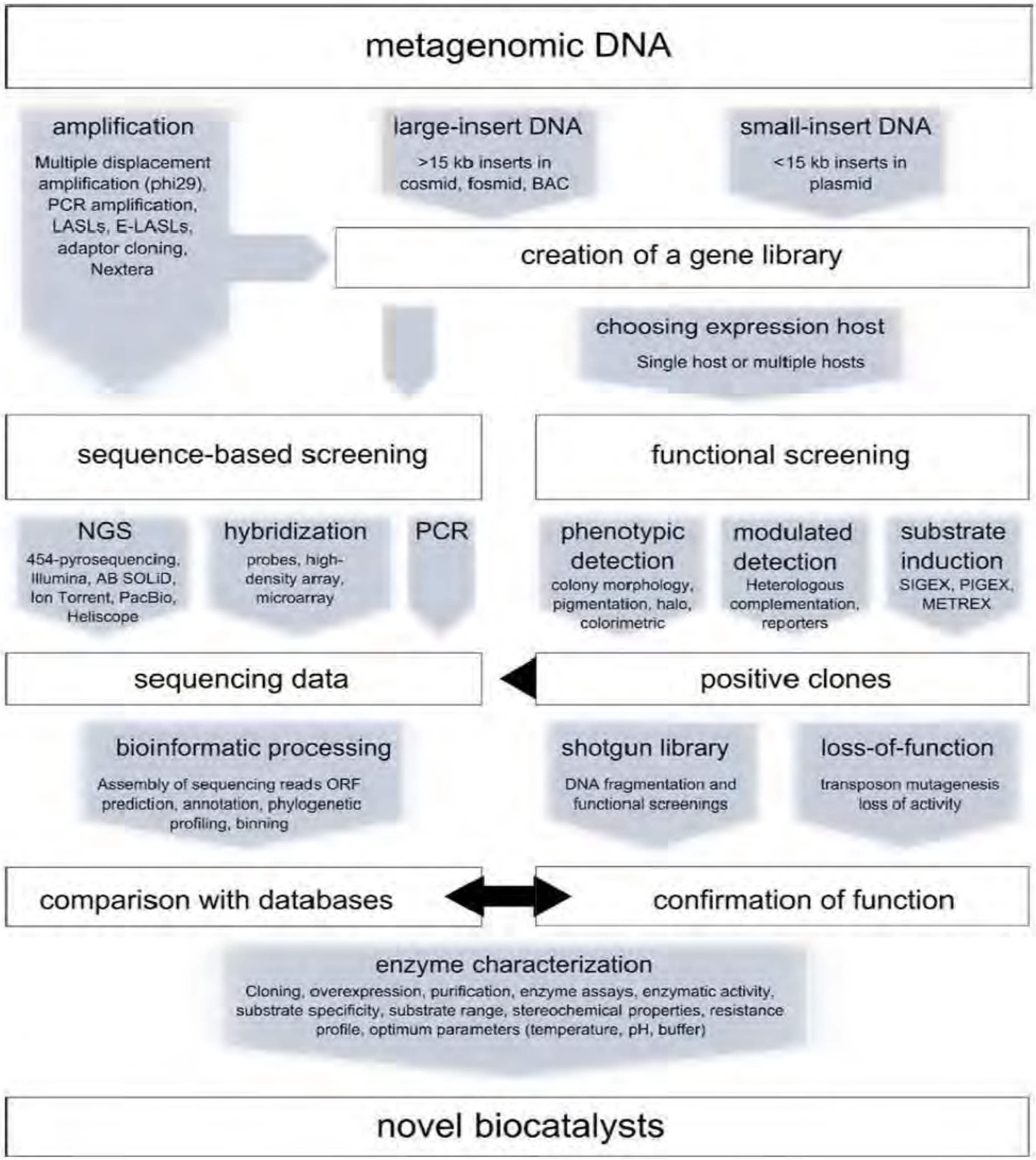


Figure 2.4: Work flow for Screening Metagenomic Libraries from environmental samples (from Leis *et al*., 2013).

preprocessing of soil and sediment samples, agarose gel purification (using Agarose, Polyacrylamide, Sephadex, Sephadex and Sephacryl) electroelution and various chromatographical separations. (Moreira, 1998; Young *et al*., 1993; Sharma *et al*., 2007; Sharma *et al*., 2014).

In case of limited amounts of metagenomic DNA due to low abundance of microbes from their natural environments, particularly for very extreme ecosystems (for example, highly alkaline and saline soda lakes), various DNA amplication strategies are employed (Ferre *et al*., 2009). One of these methods is Whole Metagenome Amplification (WGA) that is carried out by Multiple Displacement Amplification (MDA) using the *Bacillus subtilis* bacteriophage Φ29 DNA polymerase (Phi29 (reviewed in Binga *et al*., 2008). The other is the so-called linker-amplified shotgun libraries (LASLs) described by Breitbart *et al*., (2002) which is commercially as available Nextera (from Epicentre and Illumina) is based on limited-cycle amplification of small amounts of metagenomic DNA (about 50 ng) with specific 454-adapters and bar codes which then can readily be sequenced using 454-pyrosequencing (Caruccio, 2011; Marine *et al*., 2011).

Depending on insert size of interest (small-insert fragments or high-molecular weight fragments), the metagenomic DNA is cloned into plasmids vectors (<15 kb), fosmids (up to 40 kb) and BACs (up to 200 kb) (Kim *et al*., 1992). They are both low copy-number vectors based on the *Escherichia coli* F factor replicon, yielding only 1-2 copies per cell. Fosmid vectors can be packaged into lambda phage heads due to the presence of dual *cos*-sites, similarly as its multi-copy precursor cosmid vector. Large-insert metagenomic DNA is commonly cloned into commercial fosmid vectors carrying cos-sites for packaging of the ligated DNA into λ phage heads prior to infection of *E. coli* host cells (e.g. pCC1FOS, Epicentre) (Park *et al*., 2008).

The phage infection procedure ensures efficient transduction of the DNA while each clone carries one metagenomic insert of approx. 40 kb. If high-molecular weight DNA needs to be cloned into small-insert vectors, mechanical shearing into smaller random sized fragments can be performed using ultrasonic or Hamilton syringe treatment or enzymatic digestion and size select the fragment needed (Manufacturere‘s procedures, Epicentre). The expression of only a few genes and even weakly active enzyme genes is more likely when using small-insert libraries with plasimds having high-copy number and strong promoters. On the contrary, large-insert libraries rely on the expression of many genes and entire operons mainly on native promoters located on the insert and thus expected to hold more metagenomic information than smaller insert sequences.

The last stage of the metagenomic library preparation is screening for phylogenetic markers or ―achors" from clone libraries generated from metagenomic DNA. The Screening may include 16S rRNA for diversity studies, fnctional genes, or for other conserved genes, by hybridization or multiplex PCR (Stein *et al.*, 1996) or for heterologous gene expression of specific traits, such as enzyme activity/biocatalysts or antibiotic production other ecological functions (Schloss and Handelsman, 2003), or they can be sequenced randomly (Tyson *et al.*, 2004) for detailed genetic analysis. Generally, two strategies, namely function-based and sequence-based screening are used to screen and identify novel biocatalysts or genes involved in the control of bioproducts from metagenomic libraries (Schloss and Handelsman 2003).

2.4.2.2 Sequence-Based Metagenomics

Sequence-based metagenomics are applied to reveal the gene composition and phylogenetic diversity (using phylogenetic marker genes or phylogenetic anchor) of a complex microbial community through sequence homology (by PCR or hybridization) or direct sequencing. For example, for identification of biotechnologically important pathways or target or functional genes from metagenomic libraries, the target ORFs (open reading frames) of the sequences encoding conserved domains of known genes or deduced proteins of interest are identified by using PCR amplification (Vergin *et al.*, 1998) or hybridization techniques (Stein *et al.*, 1996).

PCR primer is used for the successful amplification and expression of full-length environmental genes specific to enzyme-encoding genes from different metagenome samples (Chandrasekharaiah *et al.*, 2012; Tekedar and Sanli-Mohamed, 2011). It si also reported that complete functional gene sequence of the target genes can be recovered by primer walking (Jiang *et al*., 2006), or using of PCR-denaturing gradient gel electrophoresis (DGGE) and metagenome walking (Morimoto and Fujii, 2009). PCR-based strategies can be directly applied on the metagenomic DNA sample or the constructed libraries. These approaches, however, are usually associated with the risk of PCR bias due to unequal amplification of mixed template DNA (Polz and Cavanaugh, 1998) and were already shown to hinder proper assessment of microbial diversity from environmental samples (Gonzalez *et al*., 2012). Recently, full-length of target functional genes could also be identified with hybridizing DNA probes using gene-targeted metagenomics (Iwai *et al*., 2011) and short oligonucleotide probes by microarray technology (Vilchez-Vargas *et al*., 2012).

For several decades, well–known genetic techniques such as fluorescent *in situ* hybridization (FISH) and its derivatives such as CARD-FISH (Pernthaler *et al*., 2002; Schippers *et al*., 2012), DGGE (Dias *et al*., 2012; Vaz-Moreira *et al*., 2013), terminal restriction fragment length polymorphism (T-RFLP) (Babcock *et al*, 2007) or real-time PCR (RT-PCR) methods (Dias *et al*., 2012; Schippers *et al*., 2012) have been used to study the structure and abundance of environmental microorganisms.

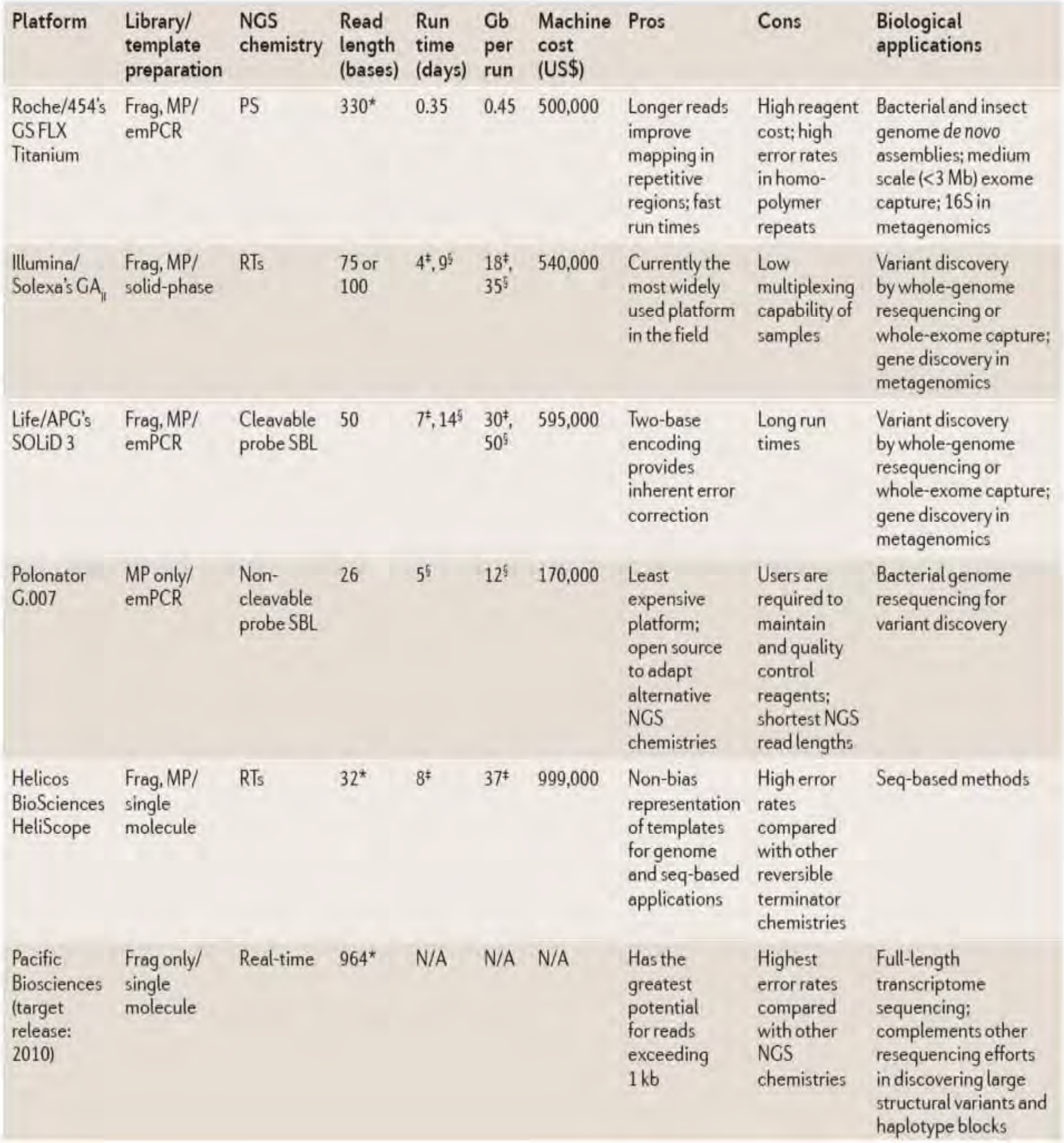
The conventional Sanger capillary-sequencing technique (Sanger *et al*., 1977) is also still in use for marker genes like ribosomal RNAs (Haas *et al*., 2011; Youssef *et al*., 2012). However, these methods have very low phylogenetic resolution and are time consuming and are not applicable for the characterization of complex microbial communities. These days, next-generation sequencing (NGS) techniques have dramatically changed the landscape of microbial ecology and large-scale and in-depth information of microbial genes are widely accessible. NGS platforms overcome some of the limitations by sequencing the vast diversity of thousands of different metagenomic DNA templates in parallel while reducing sequencing costs (Shokralla *et al*., 2012).

Seuqencing can be based on PCR or non-PCR plateforms. The NGS sequencing like the Roche 454 Genome Sequencer (Roche Diagnostics Corp.) (www.454.com), the AB SOLiD System and Ion Personal Genome Machine (both Life Technologies Corp.) (www.appliedbiosystems.com) and the HiSeq 2000 Genome Analyzer (/Solexa/Illumina Inc.) (www.illumina.com) are all PCR-based technologies. The Non-PCR sequencing methods like Heliscope (Helicos BioSciences Corp.) (www.helicosbio.com), and PacBio RS SMRT (single-molecule real-time) system (Pacific Biosciences) (www.pacificbiosciences.com) are besed on Single-molecule sequencing (SMS) technologies. All the sequencing technologies follow the methods of template preparation, sequencing and imaging, and data analysis. The unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform (Metzker, 2010). These differences in data output present challenges when comparing platforms based on data quality and cost.

The 454/Roche pyrosequencing is most widely used because it generates significantly longer reads than the other competing platforms. In the meantime, large-scale metagenome projects are making increased use of the Illumina and, to a lesser extent, SOLiD platforms. Although the latter two still provide shorter reads than pyrosequencing, they offer a much higher throughput and hence coverage for the same price. All the plateforms, however, are widely applied for mass sequencing of microbial communities and they are still under constant improvement (reviewed in Metzker, 2010; Shokralla *et al*., 2012; Thompson and Milos, 2011). The comparisons of the different next-generation sequencing plateforms (table 2.3) with their detailed technologies were well documented by Metzker (2010).

The new contemporary sequence-based metagenomic analysis offers the possibility to bypass cloning steps and the need for heterologous gene expression systems and instead rely on direct sequencing of community DNA. Tyson *et al* (2004) and Venter *et al* (2004) used large-scale sequencing for the characterization and discovery of new microbial species and genes using samples from an acid mine and Sargasso Sea. In a similar work massive sequence dataset was generated from the Global Ocean Sampling (GOS) expedition which has provided potential signal for the discovery of new protein families in Eath‘s oceans around the world and in nature (Rusch *et al.,* 2007; Williamson *et al.,* 2008; Yooseph *et al.,* 2007).

Table 2.3: Comparison of Next-generation Sequencing Plateforms (Taken from Metzker, 2010).



Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US$)	Pros	Cons	Biological applications
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo-polymer repeats	Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics
Illumina/ Solexa's GA$_{II}$	Frag, MP/ solid-phase	RTs	75 or 100	4[‡], 9[§]	18[‡], 35[§]	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7[‡], 14[§]	30[‡], 50[§]	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics
Polonator G.007	MP only/ emPCR	Non-cleavable probe SBL	26	5[§]	12[§]	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8[‡]	37[‡]	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks

*Average read-lengths. ‡Fragment run. §Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLiD, support oligonucleotide ligation detection.

For the last few years, a lot of work has been conducted on microbial phylogeny, structure and metabolic characteristics through analysis of metagenomic sequence data. They were generated by next generation sequencing platforms and/or Sanger sequence analysis from Soap Lake (Hawley and Hess, 2014), anoxic sediment of a Sub-saline shallow lake, Laguna de Carrizo, Spain (Ferrer *et al.*, 2011). Furthermore, different workers also got sequence data from hot springs of Kenyan soda lake (Lake Bongoria) habitats (Dadheech *et al.*, 2013), North Pacific Ocean (Wu *et al.,* 2013), antibiotic resistance genes from human gut microbiota (Hu *et al.*, 2013), agricultural soils of Argentina (Rascovan *et al.*, 2013), and from forest soil horizons in Norway (Uroz *et al.*, 2013). The vast Metagenomic DNA sequences datasets derived from such or similar studies have also initiated the use of *in-silico* screening and evaluation approaches for metagenomic data to identify novel gene sequences, which can further be used for *in vitro* testing (e.g. Thomas *et al*. 2012; Wooley *et al.*, 2010; and Wooley and Ye, 2009).

2.4.2.3 Function/Activity-based Metagenomics

Although the new ultra-fast sequencing technologies quickly generate a remarkable number of target gene candidates, information about the functional/activity assays are still needed for their applicability. Compared to sequence-based metagenomics, function/activity-based screening of metagenome libraries is dependent upon the expression of the gene products (Handelsman, 2004). This can be cloned into expression vectors (plasmids, cosmids, fosmids, or phages to detect of target gene products or biocatalytic/enzymatic activities through heterologous host/expression, in *E. coli*. (Handelsman, 2004; Uchiyama and Miyazaki, 2010). Three different function-driven screenings are being used in metagenomic functional screening strategies (Ekkers *et al.*, 2012). These systems are summarized below;

a. **Phenotypic detection:** involves the expression of metagenomic genes through phenotypic detection of active clones/host. These include changes in colony pigmentation, changed morphology or formation of clear halos upon enzymatic hydrolysis of the substrates, or color changes around the colony due to released chromophores. It is a classical, cheaper and simple approach in functional enzyme screening from metagenomes using plate assays on growth agar containing substrates and indicator substances. Clone mixtures from libraries are plated or spots from arrayed metagenomic libraries are grown on agar plates containing the substrate, or subsequent substrate overlays are applied after growth of the colonies.

b. **Modulated detection:** involves the use of separate reporter systems, or heterologous complementation that involves growth advantage of metagenomic library clones or mutant strains under selective conditions e.g antibiotic, *lacZ*- β-galactosidase-based system.

c. **Induced gene expression:** The approach that involves the use of an ‗operon trap' vector which contains a reporter gene (eg. green fluorescent protein -*gfp*) immediately downstream of a cloning site for the genomic insert(s) such that expression of the inserted gene(s) is coupled with substrate or poduct induction/expression of the reporter gene. This method is usually useful in high-throughput screening methods and when activities do not result in detectable phenotypes using FACS. The various methods that use this approach involve substrate induced gene expression (SIGEX) (Uchiyama *et al*., 2005), metabolite-regulated expression screening (METREX) (Williamson et *al*., 2005), and product induced gene expression/ product sensing reporter systems (PIGEX) systems (Uchiyama & Miyazaki, 2010).

Direct functional screens often used to access complete novel genes and pathways without knowledge of prior sequence data and hence may fundamentally reveal novel and previously unknown genes and gene products distinct from those currently known (Handelsman, 2004). However, function-based screening has several limitations. First, when functional enzymatic activity depends on more than one genetic subunit, the clone must contain the complete gene sequence, or even a gene cluster.

Second, although common *E. coli* host strain has been in use as screening heterologous host in several function-based metagenomic screening studies, it may not efficiently express many genes (low detection frequency of genes) from metagenomic samples. This is due to differences in codon usage, transcription and/or translation initiation signals, protein-folding elements, post-translational modifications, such as glycosylation, or toxicity of the active enzyme (Angelov *et al*., 2009; Liebl, 2011). In addition, efficient and economical screening methods for desired traits must be established to facilitate high-throughput-screening of vast libraries.

2.4.3 Metagenome-derived enzymes

Metagenomic libraries have been screened for a wide range of biocatalysts which are of industrial interest (Streit *et al*., 2004; Steele *et al*., 2009)). Some of the examples of foremost and well documented enzymes include, Agarase (used as an additive in food industries) from soil metagenome (Voget *et al*., 2003); Amylases (used in the detergent and food industries) from soil and cow gut metagenomes (Ferrer *et al*., 2005b; Voget *et al*., 2003), soil and compost (Yun *et al*., 2004; Lämmle *et al*., 2007).

Cellulases were also characterized from wide environments such as soils (Voget *et al*., 2006), soda lake sediments in Africa and Egypt (Grant *et al*., 2004; Rees *et al*., 2003), hot springs (Graham *et al*., 2011), rabbit cecum and termite guts (Feng *et al*., 2007; Nimchua *et al*., 2012). Xylanases were also identified from insect gut (Brennan *et al*., 2004), soils (Wang *et al*., 2012) and waste lagoon (Lee *et al*., 2006a).

Different studies also detected Lipases/Eastrases from deep sea hypersaline basin (Elend *et al*., 2006; Ferrer *et al*., 2005a), hot spring (Rhee *et al*., 2005), and arctic sediments (Fu *et al*., 2012). Proteases enzymes with important application in industry, particular in detergents and in the food industries) were also identified from microorganisms from soils (Rondon *et al*., 2000; Gupta *et al*., 2002) and alkaline soil (Niehaus *et al*., 2011).

2.4.4 Bioinformatics tools for mining metagenomic data

There are constantly increasing needs and challenges to analyse and obtain meaning full interpretations for the mounting datasets generated from metagenome sequences (reads- or contigs-) by next-generation sequencers. This was made possible since the first massive direct community DNA sequences generated by Tyson *et al* (2004) and Venter *et al* (2004). More innovative and sophisticated bioinformatics tools and program packages must be devised to assure microbial diversity studies from environmental samples. This helps to evaluate their gene functions, ecological functions, and metabolic pathways. Some of the important and widely used website/ online tools and softwares can be shown.

Well known *in silico* metagenomic sequence assebleres include, 454-pyrosequencing assembler Newbler (Roche), AMOS (Treangen *et al*., 2011), MIRA (Chevreux *et al*., 1999), SOAP (Li *et al*., 2008), Meta-IDBA (Peng *et al*., 2011) or MetaVelvet (Namiki *et al*., 2012). After assembly, Open reading frame (ORF) prediction would be done and then functional assignment (annotation) of the predicted proteins will be performed. Various ORF prediction programs used may involve MetaGene and MetaGeneAnnotator (Noguchi *et al*., 2006; Noguchi *et al*., 2008), MetaGeneMark (Zhu *et al*., 2010), Orphelia (Hoff *et al*., 2009), FragGeneScan (Rho *et al*., 2010) and Glimmer-MG (Kelley *et al*., 2012).

The next step in the process of metagenomic sequence data analysis is the assignment of functions to the predicted ORFs from assembled DNA sequences to study enzymatic functions and metabolic properties or even the ecological role of the organisms. For this purpose, different tools, like SWISS-PROT (Bairoch and Apweiler, 2000), Pfam (Punta *et al*., 2012), TIGRFAMs (Selengut *et al*., 2007) and integrated databases like nonredundant NCBI (National Center for Biotechnology Information) database (Sayers *et al*., 2012), UniProt (UniProt Consortium, 2010) or InterPro (Hunter *et al*., 2012) are used.

Other databases are based on the phylogenetic classification of proteins by Clusters of Orthologous Groups (COG) (Tatusov *et al*., 2001), by Gene Ontology (GO) (Ashburner *et al*., 2000). The properties of the product including molecular function, biological process and cellular component can be studied by mapping the enzymatic function to whole biochemical pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa *et al*., 2008) and the comparative genomics environment SEED (Overbeek *et al*., 2005). Databases for comprehensive enzyme information are available from the BRENDA enzyme information system (Scheer *et al*., 2011) that contains detailed information about catalyzed reactions, substrate specificities, structures and functional parameters.

For microbial diversity assessment, similarity searches of the metagenomic sequences are performed against known genes and sequences from public databases. For example, in the use of 16S rRNA genes as phylogenetic anchors (Woese, 1987), several ribosomal rRNA databases are available, such as SILVA (Pruesse *et al*., 2007), Greengenes (DeSantis *et al*., 2006) and the Ribosomal Database Project-II (RDP-II, Cole *et al*., 2003). Other binning programs (Phylotyping of contigs according to their likely taxonomic origins) asssit in inferring sequence homologies of metagenomic fragments include online tools like IMG/M (Markowitz *et al*., 2012), MG-RAST (provides a comparative functional and sequence-based analysis for uploaded samples (http://metagenomics.nmpdr.org/) (Meyer *et al*., 2008), MEGAN (Huson and Mitra, 2012), CARMA (Krause *et al*., 2008), SOrt- ITEMS (Monzoorul *et al*., 2009) and MetaPhyler (Liu *et al*., 2011). The Metagenome Analyzer MEGAN uses BLAST (Altschul *et al*., 1990) as source for taxonomic binning by comparison with NCBI reference databases NCBI-nr and NCBI-nt.

CHAPTER THREE: Materials and Methods

3. Materials and Methods

3.1 Descriptions of Study Areas

Three Ethiopian Rift Valley lakes, Chitu, Shalla and Abijata were selected as study areas of the current study. The morphometric and physical features about the the sampling sites are described in Table 3.1.

Lake Chitu (N 07^{o} 24.433, E 038^{o} 25.547 m above sea level (asl)) is a crater lake (crater diameter of 1.6 x 1.2 km) located at the altitude of 1560 m on an area of approximately 0.8 km^2 with a maximum depth of about 17m. It is situated in the southern part of Ethiopia, near Shashamane, approximately 287 Km away from the capital city, Addis Ababa. Lake Chitu is mainly characterized with high salinity (60 g/l) and it is highly saline and alkaline, where Na^+, Cl^-, HCO_3^- and CO_3^{2-} are the predominant ions, with an average pH value of 10.4. This high alkalinity is due to the high concentration of sodium carbonate. The lake is closed and does not receive any industrial seepage and no significant outflow. Lake Chitu is highly rich in microbial composition and usually dominated with the blooms of planktonic *Spirulina spp*. which are the principal food of the vast population of about 5,000-10,000 filamingos that inhabit the lake.

Lake Shalla (N 07 ^{o} 29.293, E 038^{o} 37.713 m above sea level (asl)) is another crater lake adjacent to and about 1.5 km away from Lake Chitu. It is the deepest of the Ethiopian Rift lakes, with an average depth of 87 m (maximum 226 m) and a surface area of 329 square kilometers. It is relatively diluted lake with total salinity of 10 g/l with an average pH value of 9.7. It receives water from the Jido River that drains the western escarpment. The area of the lake is characterized by a canopy of giant wild fig trees and acacias and a number of hot springs that drains into the lake. Lake Shala‘s islands are used as breeding sites by many birds.

Lake Abiyata (N 07 ^{o} 35.1, E 038^{o} 35.2) is a relatively shallow, small, alkaline closed lake, located in a saucer-shaped hollow within a deep faulted trough at an elevation of 1578 m (asl) with average depth of 7.6 m (maximum 14 m). Its total salinity is 35 g/l with an average pH value of 9.8. The main inflow is from direct rainfall and discharge from the Bulbula and Horakelo rivers, which are the outflows of lakes Ziway and Langano, respectively. As a closed lake, the only significant water loss from Lake Abiyata is evaporation, although recently, due to anthropogenic activities, the loss has been enhanced by development schemes in the catchment such as pumping of water from the lake for soda ash extraction, diversion of Feeder Rivers and direct use of the Lake Ziway water for irrigation. The lake has abundant cyanobacteria and high chlorophyll content reflecting its high primary productivity (Kebede *et al*., 1994). It also found to attract large inhabitants of flamingos and other several bird species.

3.1.1 Sample collection

Lake sediments were collected from Lake Shalla, Lake Chitu, and Lake Abijata (Fig. 3.1). The sediment samples from the three lakes were randomly collected from different depths of 0.5m, 2m, 5m, 9.5m, 12m, 16m using Eckman grab and from 10cm of the shores. The sediment samples from each depth layer for each sample sites were pooled and thoroughly mixed in sterilized containers and preserved in ice box till and taken to the laboratory and preserved at 4 ^{0}C until use. Some information about the sampling sites is described in table 3.1.

Table 3.1: Descriptions of Morphometric and Physiscal Features of Sampling sites (data adapted from Baye Sitotaw, 2014, unpublished data).

Parameter	**Lake Chitu**	**Lake Shalla**	**Lake Abijata**
Coordinate	N 07^{o} 24.433, E 038^{o} 25.547	N 07 ^{o} 29.293 E 038^{o} 37.713	N 07 ^{o} 35.1 E 038^{o} 35.2
Altitude (m)	1560	1558	1578
Surface area (km^{2})	0.8	329	176
Max. depth (m)	17	226	14
Mean depth (m)	-	87	7.6
Secchi depth (cm)	20-24	125	65
Catchment area(km ^{2})	-	2920	1630
pH	10.4	9.7	9.8
Conductivity (ms/cm)	56	24	44
Total Salinity (g/l)	60	10	35
Total Alkalinity($CaCO_3$) mg /l	583	167	340
Surface temp (^{0}C)	25	25	26

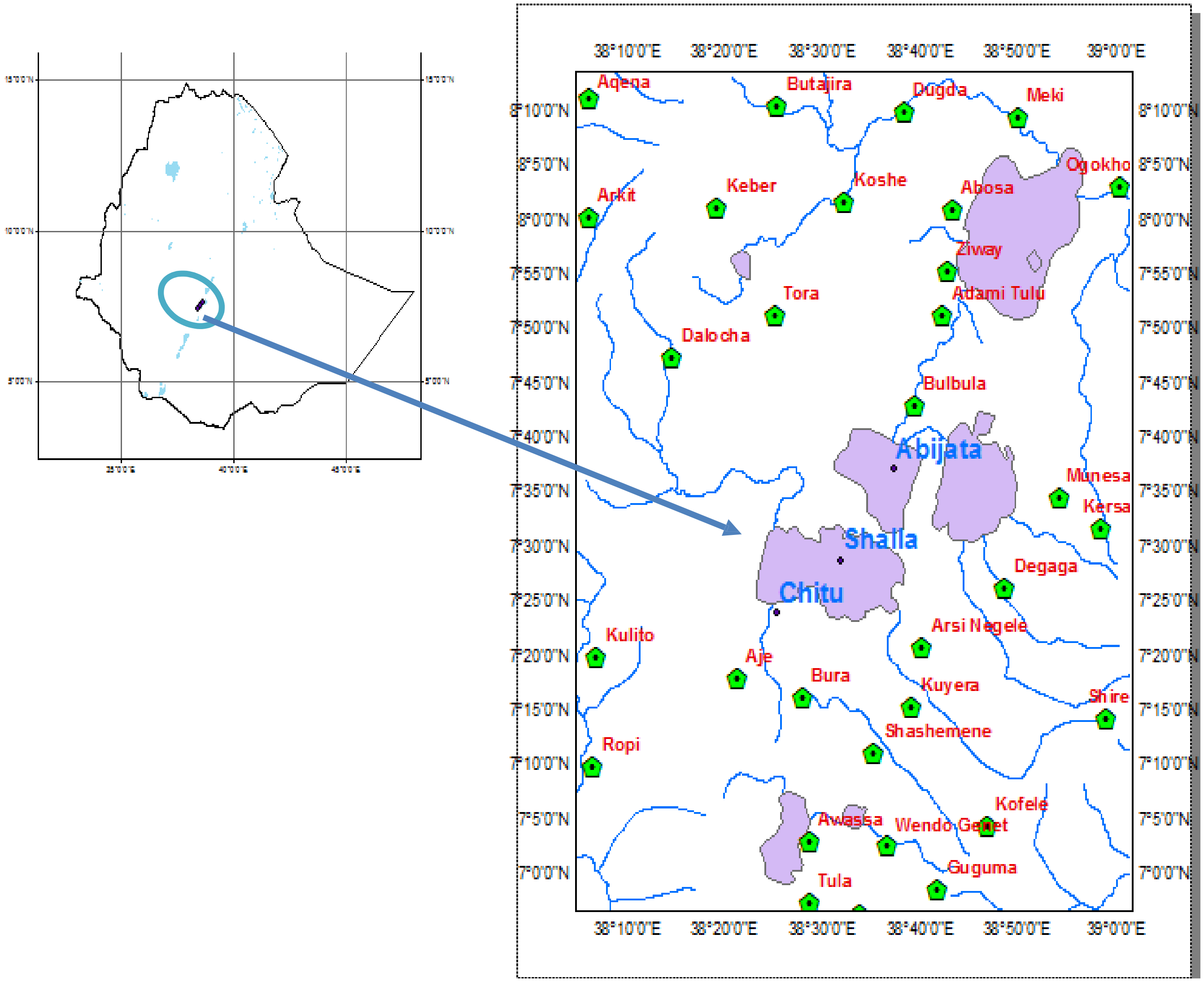


Figure 3.1: Map of Sampling Sites, Chitu, Shalla and Abijata Lakes in the Ethiopia Rift Valley (Assisted by Dr. Eng. Habtamu Itefa, Addis Ababa Science and Technology)

3.2 Chemicals/Reagents

All media and chemicals/reagents used for DNA extraction were analytical grade obtained from MERCK and SIGMA and EPICENTER international laboratory supplies unless stated otherwise. Restriction enzymes, T4 DNA ligase, DNA polymerase with10x reaction buffer, Proteinase K, Agarase and dNTPs were obtained from Fermentas to perform molecular biology experiments such as restriction digest, ligation and PCR for various DNA manipulation procedures. Wherever appropriate, microbial broth or growth media preparations were supplemented with aseptically prepared antibiotics. Filter sterilized (pore size, 0.22µm) ampicillin, streptomycin and chloramphenicol were added to autoclaved Luria-Bertani (LB) broth or agar media following cooling to ~ 55^0C at final concentrations of 100ug/ml, 100ug/ml, 12.5ug/ml, respectively.

3.3 Culture Media

3.3.1 Luria-Bertani (LB) broth and Agar

Luria-Bertani (LB) broth and LB agar (1.5% agar) was prepared from 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl (Miller, 1972). The components were mixed with high quality Millipore water and adjusted to pH 7.0 using NaOH before autoclaving.

3.3.2 X YT medium

2X YT medium was prepared from 1.6% (w/v) tryptone, 1% (w/v) yeast extract 0.25% (w/v) NaCl and mixed with sterile water. The components were mixed with high quality Millipore water and the pH adjusted to 7.0 using NaOH before autoclaving. The medum stored at 4^0C until required.

3.3.3 Protease Medium

Protease agar plates were prepared according to Lee *et al* (2007). The medium was prepared by mixing 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl and 1.5% bacteriological agar (section 2.2.1) with the addition of 2% (w/v) skim milk powder as source of protein. The pH of the LB agar and the skim milk powder mixtures was separately adjusted to 7.0, and later were mixed before cooling and supplemented with chloramphenicol antibiotics (12.5ug/ml) and poured into plates. Plates were stored at 4^0C until used.

3.3.4 Amaylase Medium

Amylase agar plates were also prepared according to Skerman (1969). The medium was prepared by preparing the mixture of LB agar as above (section 3.3.1) with the addition of 1% (w/v) soluble starch from potato, adjusted to pH 7.0, autoclaved, supplemented with chloramphenicol antibiotics (12.5ug/ml), and stored at 4^0C.

3.3.5 Cellulase Medium

Celullase agar plates were prepared according to Teather and Wood (1982). The medium was prepared by preparing the mixture of LB agar as above (section 3.3.1) with the addition of 0.2% (w/v) carboxymethylcellulose (CMC) after which the pH was adjusted to 7.0, supplemented with chloramphenicol antibiotics (12.5ug/ml), and stored at 4^0C.

3.3.4 Lipase/Esterase medium

Lipase agar plates were prepared according to Kouker and Jaeger (1987). The medium was prepared by preparing the mixture of LB agar as above (section 3.3.1) with the addition of 1% (v/v) olive oil, 0.001% rhodamine B and 0.1% gum Arabic after which the pH was adjusted to 7.0. Agar plates for esterase screening contained the same constituent as lipase except that 1% (v/v) tributyrin (v/v) was replaced by live oil and rhodamine B. The medium was supplemented with chloramphenicol antibiotics (12.5ug/ml) and stored at 4^0C.

3.4 Extraction of Total Community/Metagenomic DNA

Community DNA/Metagenomic DNA extraction from both Chitu and Shalla lake sediment samples was done using the methods developed by Zhou *et al* (1996) with modifications. Accordingly, lake sediment samples (10g) from each depth layer were suspended in 13.5 ml of extraction buffer (1% CTAB [w/v]; 100 mM Tris, pH 8.00; 100 mM NaH_2PO_4, pH 8.00; 100 mM EDTA; 1.5 M NaCl; 0.02% Protease K [w/v]) in 50ml sterile eppendorf tubes. The samples were then incubated with proteinase K (20 mg/ml) at 37 ^{0}C for 30 min on a horizontal shaking (150 RPM) followed by the addition of 1.5ml of 20% Sodium dodecyl sulphate (SDS) and further incubated at 65 ^{0}C for 2 hrs with occasional gentle inversion every 15 min and sediment residues were removed by centrifugation (Beckman) at 6,000 x g for 10 min at room temperature to remove the DNA in the supernatant.

The DNA was carefully transferred into new sterile centrifuge tubes. The DNA in the pooled supernatant was extracted with equal volume of phenol/ chloroform/ isoamyl alcohol (25: 24:1) at 16, 000 × g for 5 min at room temperature. The aqueous phase/supernatant was again transferred to sterile centrifuge tubes with the addition of an equal volume of chloroform. After careful mixing the tubes were centrifuged at 16, 000 × g for 5 min at room temperature to recover the DNA in supernatant. The DNA was allowed to finally precipitate for 4 hrs at room temperature by adding 0.6 volume of isopropanol, and centrifuged at 16, 000 × g for 10 min, washed with 70% ethanol and centrifuged at 16, 000 × g for 5 min. It was then air dried and finally dissolved in 10mM Tris buffer (pH 8.00).

3.5 Purification and Quantification of Metagenomic DNA

The crude metagenomic DNA obtained was electrophoresed (BioRad) on 0.7% low melting agarose gel at 40V for overnight at room temperature and gel was observed through UV transilluminator (BioRad). The high molecular size of the DNA was estimated using the 48 kb intact and *HindIII* digested lambda.

High molecular size of DNA band was aseptically excised and eluted from gel using agarase enzyme (New England BioLabs). The DNA containing agarose slice was transferred into a sterile tube and was molten by incubating at 65 ^{0}C for 10 min and cooled to 42 ^{0}C. After that, 200µl of 1% molten agarose was incubated with 1µl of ß-agarase I enzyme (1U) and ß-agarase I Buffer (1X) (New England BioLabs) at 42 ^{0}C for 1hr. For larger volumes, ß-agarase I enzyme was adjusted accordingly.

After end of the 1hr digestion, ß-agarase I was inactivated by incubation at 70°C for 10 minutes and the DNA was further purified and precipitated using 1/10 volume of 3M sodium acetate (pH 6.5) and 2.5 volumes of ethanol. The quality of DNA was also checked by running small fraction of the DNA sample on 0.7% low melting agarose gel at 40V for overnight at room temperature. Purified DNA from different extractions from each sample area was pooled and concentrated in a speed table centrifuge (3,000 rpm for 30 min) to a final volume of 40 µL. A Qubit® dsDNA HS (High Sensitivity). Assay Kit with its concentration reader, Qubit® 2.0 Fluorometer plateform (Life Technologies-invitrogen) was used to obtain an accurate quantification of the DNA. Sample reading was taken and calculated for 2 µL. The purified and quantified metagenomic DNA was stored at 4 ^{0}C until used.

3.6 Bacterial strains, Plasmids and Primers used for the Study

Table 3.2: Summary of the List and major Descriptions of the strains, plasmids and primers Used for various DNA studies.

Item	Features	Source
Strains:		
GeneHogs *E. coli*	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G *fhu*A::IS2 (confers phage T1 resistance)	Invitrogen
EPI300™-T1R Phage T1-resistant *E. coli*	[F– mcrA Δ(*mrr-hsd*RMS-*mcr*BC) (StrR) φ80dl*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G *trf*A *ton*A *dhfr*]	Epicentre
MaxPlax Lambda Packaging Extracts	Derived from lysogenic *E. coli* strains, BHB2690 and BHB2688, NM759* the preparation of the sonication extract, is a restriction-free K12-derived strain deficient in the production of l-phage capsid protein D.	Epicentre
Plasmids:		
pGEM-T Easy vector	*lac Z'*, Amp^R, 3015 bp	Promega
pCC1FOS™ Vector	Bacteriophage lambda *cos* site, Chl^R, 8139 bp	Epicentre
Primers (5'-3')		
E9F	*GAGTTTGATCCTGGCTCAG*	Farely *et al*. (1995)
U1510R	*GGTTACCTTGTTACGACTT*	Reysenbach *et al*. (1992)
S.C. Act- 235-S-20	*CGCGGCCTATCAGCTTGTTG*	Stach, *et al*. (2003a)
S.C. Act- 878-A-19	*CCGTACTCCCCAGGCGGGG*	Stach, *et al*. (2003a)
M13 F	*GTTTTCCCAGTCACGAC*	Promega
M13 R	*CAGGAAACAGCTATGAC*	Promega
V3F_341F	*CCTAYGGGRBGCASCAG*	Yu, *et al*. 2005
V4R_806R	*GGACTACNNGGGTATCTAAT*	Yu, *et al*. 2005
Pyrosequencing Adapters/Tags and Barcodes (5'-3')		
TitA_341F (Adapter A)	*CGTATCGCCTCCCTCGCGCCATCAG*	Yu, *et al*. 2005

TitB_806R (Adapter B)	*CTATGCGCCTTGCCAGCCCGCTCAG*	Yu, *et al.* 2005
Barcodes (9)		
(Chitu = MID51, MID52, MID 53): (Shalla = MID54, MID55, MID 56): (Abijata = MID 57, MID58, MID59):	*AGCTCACGTA, AGTATACATA, AGTCGAGAGA* *AGTGCTACGA, CGATCGTATA, CGCAGTACGA* *CGCGTATACA, CGTACAGTCA, CGTACTCAGA*	Yu, *et al.* 2005

3.7 PCR Amplification using bacterial and actinomycetes specific 16S rDNA primers

The pure DNA was then used as template to amplify a bacterial domain specific ca. 1500 bp 16S rDNA gene with the bacterial universal primers E9F (5‘– GAGTTTGATCCTGGCTCAG–3‘) and U1510R (5‘–GGTTACCTTGTTACGACTT–3‘) using Gene Amp® PCR system 2700 (Applied Biosystems). For actinomycetes specific ca. 646 bp 16S rDNA gene, the actinomycete specific universal primers, S-C-Act-235-a-S-20 (5‘-CGC GGCCTATCAGCTTGTTG-3‘) and S-C-Act-878-a-A-19 (5‘-CCGTACTCCCCA GGCGGGG-3‘) were used (Table 3.2).

The PCR cycling program for bacteria was as follows: initial denaturation at 95^{0}C for 4 min, 30 cycles of 95^{0}C for 30sec, annealing at 52^{0}C for 30 seconds, and elongation at 72^{0}C for 1 min, and a final extension of 5 min at 72^{0}C. Reagent composition of PCR reaction mixture (50 µL) contained genomic DNA extract (1µL), 10X Taq polymerase Buffer, dNTs mixture (2.5 pmol), each primer (20 pmol), and Taq DNA polymerase (2.5 U).

The PCR cycling program for actinomycetes was done using a ‗touchdown‘ protocol (Roux, 1995) as performed by Stach *et al.* (2003). It consisted of an initial denaturation at 95^{0}C for 4 min, followed by denaturation at 95^{0}C for 45 s, annealing at 72^{0}C for 45 s and extension at 72^{0}C for 1 min; 10 cycles in which the annealing temperature was decreased by 0.5^{0}C per cycle from the preceding cycle; and then 15 cycles of 95^{0}C for 45 s, 68^{0}C for 45 s and 72^{0}C for 1 min, with the last cycle followed by a 5 min extension at 72^{0}C. For both bacteria and actinomycetes, three independent PCR reactions were performed for each sample originating from each sample site and these three reactions were pooled following each PCR amplification. Finally, the 16S rDNA PCR amplicons were electrophoresd (BioRad) on 1% agarose gel and purified from gel using the

GFX™ DNA and gel band purification kit (GE Health Care) according to manufacturer‘s specifications.

3.8 Construction Bacterial and Actinomycetes specific16S rRNA Gene Libraries

The pGEM-T Easy vector (Promega, USA) cloning plasmid was used to construct the 16S rDNA metagenomic libraries of bacteria and actinomycetes. pGEM-T Easy vector was supplied as linearized vectors with a single 3′-terminal thymidine at both ends suitable for cloning DNA inserts with 3′A-tailed fragments generated by thermostable Taq polymerases.

The GFX™ gel purified 16S rDNA PCR amplicons were ligated according to the manufacturer‘s instructions and transformed into electrocompetent GeneHog *E. coli* cells (Invitrogen) by electroporation (Gene-Pulser, Bio-Rad) with the conditions: Voltage 1.8Kv, capacitor 25µF, and resistor 200 Ω and time constant 4.2-4.5 ms. The positive clones with insert were screened on X-Gal (5- bromo-4-chloro-3-indoly-b-D-galactopyranoside) –IPTG (isopropyl-b-D-thiogalactopyranoside) –ampicillin supplemented LB agar plates by blue/white color-based recombinant selection.

3.9 Preparation of Electrocompetent *E. coli* cells

Electrocompetent cells were prepared using the standard molecular biology manual by Sambrook and Russell (2001). A single colony of the *E. coli* strain was inoculated into 10 ml of LB-broth and incubated at 37°C with shaking until stationary phase. 5 ml of the culture was transferred to two aliquots of 250 ml of 2X YT medium (section 3.2.2) and incubated with shaking (200rpm, Beckman, Germany) at 37°C until mid-logarithmic phase (~ OD_{600} of 0.6). The flasks were rapidly cooled in ice-water for 20 min and the cells were collected in polypropylene tubes by centrifugation at 1000 × g for 10 min.

The supernatant was decanted and the cells were resuspended in equal volume ice-cold Millipore water and centrifuged as before. After harvesting the cells as above, the pellets were resuspended in 125 ml 10% glycerol, collected by centrifugation as before. The cell pellet was resuspended in

1 ml 2XYT medium containing 10% glycerol and the cells were adjusted to 2 - 3 × 10^{10} cells/ml, aliquotted into 50 µl volumes, and stored at -80°C in eppendorf tubes for further work.

3.9.1 Transformation of *E. coli* cells by Electroporation

Electroporation was done using standard molecular biology manual by Sambrook and Russell (2001). Fifty microlitters of electrocompetent cells was removed from -80°C and allowed to thaw on ice, to which 2µl of ligation mix was added and gently mixed. The mixture was returned to ice for ~ 1 min and pipetted into a pre-cooled 0.1 cm sterile electroporation cuvette (Bio-Rad Laboratories). Electroporation was performed using the following conditions: 1.25 – 1.8 kV, 25 µF, 200 Ω. Immediately after electroporation, 950 µl of a pre-warmed S.O.C medium was added to the cuvette and the cells were transferred to a 15 ml Falcon tube and incubated at 37°C for 1 h with agitation.

The cells were plated in aliquots of 5 to 50 µl onto LB-agar plates supplemented with the appropriate antibiotic. Where applicable, recombinant transformants were selected by blue/white colour selection based on insertional inactivation of the *lac*Z gene. For this purpose, the cells were spread together with 40 µl of X-gal (2% [v/v] stock solution) and 10 µl IPTG (100 mM stock solution) over the surface of LB-agar plates, supplemented with the appropriate antibiotic and were incubated overnight at 37°C.

3.9.2 M13 Colony PCR

Putative recombinant clones were screened by PCR amplification with pGME-T vector specific primer pairs, M13F and M13R (Stach *et al.*, 2003) (Table 3.1). Colonies were suspended in 50µl sterile water and lysed by boiling for 5 minutes. It was then centrifuged at **12,000** × g for 10 minutes and 5µl of the supernatant was used as template in each PCR amplification in 96-well plate format.

The DNA was amplified by PCR in 25µl volumes with 20pmol of each M13 primers and 1.25 units of Taq DNA Polymerase. After an initial denaturation of 2 min at 94°C, the amplification profile was 30 cycles of denaturation (94°C for 30 sec), annealing (52°C for 30 sec) and extension (72°C for 1 min); PCR was concluded with 1 cycle of 72°C for 7 min. Amplification products (5µl) were analyzed on a 1% agarose gel containing ethidium bromide.

3.10 ARDRA Patterning of 16S r DNA amplicons

The M13 colony PCR (section 3.7.2) products of positive recombinants with insert size were then digested with the restriction enzymes, *AluI*, *RsaI* and *HindIII* (according to the manufacturere's specifications, Fermentas). The digestion reaction contained the appropriate volume of 10X buffer supplied with the specific enzyme. Approximately, 4 U of each restriction enzyme was used to digest 4µg of the PCR product in a single reaction at 37°C for overnight in a sterile 0.6ml tubes. Digestion products were run on 3% agarose gels (containing ethidium bromide) at 100 V for 3 hrs and visualized with UV-Illumination.

3.11 Nucleotide Sequencing, Accession Number and Phylogenetic analysis

One representative clone from each ARDRA group was chosen for partial 16S rDNA sequencing (Sanger) using commercial Sanger sequencing facility at Stellenbosch University, South Africa. The vector sequences were recovered from sequence results to clean 16S rDNA sequences of $\geq$ 640 bp (for actinomycetes) and 850 pb (for bacteria). They were compiled and compared to the NCBI DNA sequence database using BLASTn to verify proximate phylogenetic positions.

The sequences were aligned using CLASTALx program. The phylogenetic tree of actinomycetes and bacterial communities of both Chitu and Shalla lakes were constructed by neighbour-joining method with 1,000 resampling bootstrap analysis using MEGA vesion4.0 (Tamura *et al.*, 2007). Phylogenetic analyses of clone sequences was performed based on ribosomal gene database program (RDP). The 16S rDNA sequences of actinomycetes and bacteria obtained from both Chitu and Shalla lakes were submitted to NCBI GeneBank Database.

3.12 16S rRNA V3-V4 amplicons 454 Pyrosequencing

Pyrosequencing of the 16S rRNA genes was performed on DNA extracted from sediment samples. A fragment approximately 466 bp V3-V4 region of the 16S rRNA gene was amplified with primers 341F (5‘-CCTAYGGGRBGCASCAG-3‘) and 806R (5‘-GGACTACNNGGGTATCTAAT-3‘) spanning the V3 to V4 hypervariable regions according to the methods of Neefs *et al* (1990) modified by Yu *et al.* (2005). PCR reactions were performed in 25 µL reaction volumes consisting of 1 µL of template DNA, 5x Phusion HF buffer (Finnzymes, Espoo, Finland), 0.2 mM dNTP mixture, 0.02 U/µl Phusion Hot Start DNA Polymerase (Finnzymes), and 0.4 µM of each primer (Sigma-Aldrich, St. Louis, Mo, USA).

The conditions for thermal cycling consisted of an initial denaturation of 98°C for 30 seconds, followed by 35 cycles of denaturation at 98°C for 5 seconds, annealing at 56°C for 20 seconds, extension at 72°C for 20 seconds, and a final extension at 72°C for 5 minutes in a Bio-Rad PTC-200 DNA Engine® Peltier Thermal Cycler. After PCR amplification, the samples were held at 70 °C for 3 min and then placed on ice until the products were analyzed on 1% (w/v) agarose gel with ethidium bromide and visualized with UV-Illumination.

Bands of the PCR products were cut from the gel and purified by the Montage Gel extraction kit (Millipore, Billerica, MA, USA), according to the manufacturer‘s specifications. After 20 rounds of amplification, a second PCR (Barcoding PCR) was done as described above to add primers with a 454 Titanium adapter (Adapter A and B) and 9 nucleotide barcodes/tags (Table 3.2) required for 454 pyrosequencing to specific ends of the amplified 16S rRNA fragments using similar PCR parameters except that PCR cycles were reduced to 15.

PCR products were again analyzed by agarose gel electrophoresis and purified from the gel as before. The amplicons with adapters and barcodes were quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) and stored at -20^{0}C. Barcoded and pooled amplicons were submitted to the laboratory of Molecular and Microbial Ecology, University of Copehhagen, Denmark to run on a Roche 454 GS FLX Titanium PicoTiterPlate using the GS FLX pyrosequencing system (Roche, Basel, Switzerland). The three sets of barcoded samples (Chitu, Shalla and Abijata lakes) were sequenced as part of a large run in triplicates across nine regions of a PicoTiter Plate.

3.13 Pyrosequencing Data processing

Sequences were cleaned with AmpliconNoise and representative sequences were processed and analysed with QIIME 1.4 software (quantitative insights into microbial ecology) pipeline default settings (Caporaso *et al*., 2010). *De novo* clustering was performed for estimating richness. Taxonomic analysis was done using the inbuilt RDP Taxon Assigner parameters of RDP classfier, version 2.2 (Wang *et al*., 2007; http://rdp.cme.msu.edu/classifier/classifier.jsp).

3.14 Construction of Metagenomic Library

A metagenomic DNA library was constructed using the CopyControl Fosmid Library production kit (Epicentre) according to manufacturer instructions. Accordingly, the metagenomic DNA was size selected and digested with ß-agarase I (as in section 3.3 and 3.4). The purified high molecular weight (HMW) ~ 40kb DNA was then end-repaired with End-Repair Enzyme Mix (Epicentre) and ligated directly into the CopyControl pCC1FOS vector with Fast-Link DNA ligase (Epicentre) at room temperature for 4 hours.

In vitro packing into lambda phage was performed by adding the supplied Lambda Packaging Extracts to the ligation mixture, and infection of EPI300-T1^R phage T1-resistant *E. coli* host cells was performed according to the manufacturer protocol (Epicentre). The transformed cells were plated on LB agar medium supplemented with 12.5 µg/ml chloramphenicol and grown at 37^{0}C for 16 h. Finally, the *E. coli* fosmid library clones were singly collected (picked manually by hand) into 96-well microplate format plates with LB liquid medium and incubated overnight at 37^{0}C and stored in duplicates at at -70°C by covering with glycerol (20% (v/v)).

3.15 Endonuclease Digestion of the metagenomic Fosmid library

To estimate the average insert size of the metagenomic library, eight randomly selected fosmid clones were inoculated in 20 ml LB medium supplemented with 12.5 µg/ml chloramphenicol. Induction of the fosmids to high copy number was done by inducing the cultures with 0.02% (v/v) 5000x CopyControl Fosmid Autoinduction Solution (Epicentre), 10mM $MgSO_4$ and 0.2% maltose and incubating at 37°C for 16 hrs under agitation at 200rpm (Beckman).

After induction, the cells were collected and the fosmid plasmid DNA was extracted using Plasmid Mini Ax (A & A Biotechnology, Poland) according to the manufacturer‘s instructions. Approximately, 2µg of each clone fosmid DNA was digested with *NotI* (1u/µl) restriction enzyme at 37^0C for 2 hrs and then analyzed by electrophoresis on 1.2% agarose gel with ethidium bromide.

3.16 Functional Screening of the metagenomic Fosmid library

Copies of the fosmid library stored in 96-well microtiter plates were thawed on ice and replica plated onto lipase/Esterase, protease, cellulase and amylase assay-specific medium supplemented with the appropriate substrate and antibiotics (Sections 3.2.3-3.2.6) according to Epicentre kit specifications. For detection of cellulose-hydrolyzing clones, 0.5% carboxymethyl-cellulose (CMC) and 0.01% Congo red were added to visualize positive clones as yellow colonies surrounded by a pale hydrolytic halo.

Protease-producing clones were detected for the presence of a clear hydrolysis zone on the medium containing 2% sckimmed milk. For amylolytic enzyme-producing clones, 1% starch was added to detect a clear hydrolytic halo which was highlighted by flooding plates with iodine/potassium iodide solution I_2/KI.

Lipolytic enzyme-producing clones were detected by addition of 1% emulsified Tributyrin to detect a clear hydrolysis halo against an opque background of the medium. All incubations were carried out overnight at 37 °C until colonies were visible followed by upto 3 additional days for screening the presence of the enzyme activities in the metagenomic library.

3.17 Illumina Sequencing and data processing

3.17.1 Illumina Sequencing

Pools of clones containing the enzymatic metagenomic fosmid library were inoculated in 20 ml LB medium supplemented with 12.5 µg/ml chloramphenicol. Induction of the fosmid plasmids to high copy number and extraction were made as before (section 3.13). The fosmid DNA sample was broken down to smaller fragments (400-600bp), suitable for illumina sequencing using an ultrasound water bath (sonicator).

Approximately 1µg of DNA sample in 1.5ml eppendorf tube was sonicated for 15 seconds at intervals of every 90 seconds for 7 cycles at 4^0C. The DNA fragments were end-repaired to blunt ends (NEBNext dA-Tailing) using NEBNext End Repair Enzyme mix (New Englands BioLabs) following the supplier‘s protocol. The end-repaired fragments were then ligated to NEBNext adaptor (PE adaptor oligomix) following the NEBNext Quick Ligation kit. The reaction was then purified using the MinElute reaction Cleanup kit (Qiagen) according the protocol from the same supplier.

In order to greatly multiply the DNA without adaptor, PCR was performed in a total of 50µl reaction volume using 10µl of primers, phusion polymerase, dNTPs and 5x F518 phusion buffer (HF buffer) from the same supplier. Thermal cycling conditions consisted of an initial denaturation of 98^0C for 1 min, followed by 12 cycles of denaturation at 98^0C for 10 seconds, annealing at 60^0C for 30 seconds, extension at 72^0C for 30 seconds, and a final extension at 72^0C for 7 minutes.

The PCR product was run on 1.2% agarose gel and purified from the gel with QIAXII Gel Extraction kit (Qiagen). Sizing, quantification and quality control of the PCR product was performed using the 2100 Bioanalyzer platform (Agilent Technologies) and submitted for sequencing using the Illumina Hi-seq 2000 paired-end sequencing.

3.17.2 Assembling, Annotatng and Metabolic Profiling of Functional Genes

The software AdapterRemoval (Lindgreen, 2012, http://code.google.com/p/adapterremoval/) was used for cleaning of sequencing reads with default parameters: minquality =15, minlength =25, and maxns = 2. After read preprocessing, Velvet (http://www.ebi.ac.uk/~zerbino/velvet/) was employed to assemble these short reads. Based on the contigs, prediction of protein-coding genes was performed through a combination of MetaGene according to Noguchi *et al*. (2006) and glimmer3 (Delcher *et al*., 2007).

For each predicted Open Reading Frame (ORF), observations were collected from similarity searches against the sequence databases NCBI-nr, Swiss-Prot, KEGG and genomesDB (Richter *et al*., 2008) and the Pfam (Bateman *et al*., 2004) and InterPro (Mulder *et al*., 2005) protein family databases. SignalP was used for signal peptide predictions (Bendtsen *et al*., 2004), and TMHMM was used for transmembrane helix-analysis (Krogh *et al*., 2001).

Predicted protein coding sequences were automatically annotated by the in-house software MicHanThi, that predicts gene functions based on similarity searches using the NCBI-nr (including Swiss-Prot) and InterPro databases (Quest, 2006). The annotation of proteins highlighted within the scope of this study was subject to manual inspection and was performed with GenDB, version 2.2 (Meyer *et al*., 2003), supplemented by the tool JCoast, version 1.7 (Richter *et al*., 2008).

For all observations regarding putative protein functions, an E-value cut-off of $1e^{-5}$ has been applied. To identify potential metabolic pathways, genes were searched for similarity against the KEGG (Kyoto Encyclopedia of Genes and Genomes, Kyoto, Japan) database. All occurring KO (KEGG Orthology) numbers were mapped against KEGG pathway functional hierarchies and were statistically analyzed. All predicted ORFs (Open Reading Frames) were also searched for similarity against the COG Cluster of Orthologous Groups) database (Tatusov *et al*., 2003). A match was counted if the similarity search was below E-value of 1e-5. Transfer RNA genes were identified using tRNAScan-SE (Lowe and Eddy, 1997), and ribosomal RNA genes were identified with Meta-RNA 1.0 (Huang *et al*., 2009).

CHAPTER FOUR:

16S rRNA Clone Library-based Bacterial Diversity in Chitu and Shalla Soda Lakes

4.1 Introduction

Chitu and Shalla Soda lakes, situated in the Ethiopian rift system, are one of the hyperalkaline extreme environments that contain high concentrations of sodium and carbon ions at high pH (Grant, 1992; Lanzen *et al*., 2013). They are crater lakes with maximum depth and surface area are 17m and 0.8 km^2 for Chitu, and 226m and 329km^2 for Shalla, respectivelly. Chitu is characterized with high salinity (60g/l) and pH value of 10.4. The salinity and pH values of Shalla lake are 10g/l and 9.7 respectively (unpublished data, Baye Sitotaw, 2014). These lakes were largely characterized in respect to geochemistry, limnology, fisheries and algal compositions (Ayenew, 2007; Kebede *et al*., 1994; Tadesse *et al*., 2003; Tilahun and Ahlgren, 2010) than in microbiological and biotechnological perspectives. Studies have been made on their microbilogical composition and biotechnological applications, several of which were based on culture-dependent techniques (Gessesse and Gashe, 1997; Gessesse et al., 2003; Haile and Gessesse, 2012).

Earlier, the gap between the number of cultivable microbes in the laboratory and those present in the natural environment was not well recognized until the ground-breaking studies of Woese (1987) and Pace (1997) based on 16S rRNA or DNA gene sequence analysis. From this time on, the development of molecular biology by direct PCR amplification and sequencing of diverse 16S rRNA genes of total community DNA from environmental samples together with cloning has generated a vast quantity of data. This helps to redefine prokaryotic diversity without requiring isolation and cultivation under laboratory conditions. This shows that diversity and composition of microbial communities still remain incomplete, and necessitates more microbial studies from different environments using earlier culture based and newly emerging molecular methods so as to reveal extensive microbial diversity. In view of that, the new metagenomic approach may help to discern and genetically describe unculturable microorganisms (Daniel, 2005). Metagenomics involves the use of 16S rRNA genes for the analyses of prokaryotic community composition and diversity as well as the determination of the functional profile of the whole microbial communities in an environment (Stewart *et al.*, 2010).

In this study, 16S rDNA clone library based phylgenetic analysis was employed to describe microbial communities from sediment samples of the two soda lakes, Chitu and Shalla lakes, located in the Ethiopian Rift valley, and the general objectives were;

- Isolation of pure, intact high molecular weight metagenomic DNA from both lake sediment samples
- Sequence analysis of 16S rDNA clone libraries and identification of unique 16S bacterial and actinomycete clones

4.2 Result

4.2.1 Cloning of the 16S rDNA PCR amplicons into pGEM-T vector

DNA extraction and purification methods were optimized to obtain high molecular metagenomic DNA from both soda lake sediments. DNA yield was significantly lower and averaged only 450 ng per gram of wet sediment sample used. In both lakes, the DNA also contained high levels of humic acids and other PCR inhibitors (Data not shown).

The universal bacterial and actinomycetes domain specific primers (Farelly *et al.,* 1995; Raysenbach *et al*., 1992) have amplified the conserved regions of the 16S rDNA genes at the 3‘ and 5‘ ends and yielded PCR products size around 1500 bp for bacteria and 646 bp for actinomycetes. The agarose gel of metagenomic DNA retrieved from the sediment samples of each lake is indicated in Fig. 4.1. Single bands of the 16S rDNA amplicon of 1500 bp (bacteria) and 646 bp (actinomycetes) were observed after gel electrophoresis (Fig. 4.2).

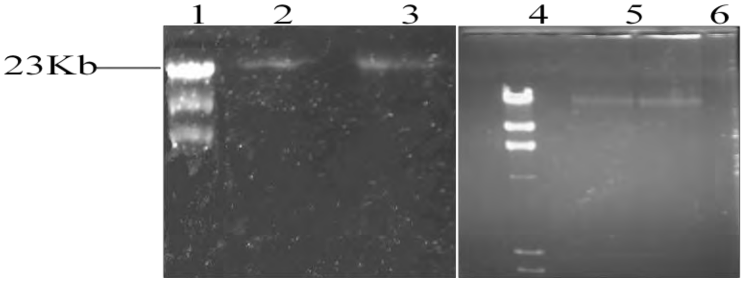


Figure 4.1: Agarose gel electrophoresis (0.8%) of gel purified DNA from Lake Sediments. Lanes 1 & 4: lambda DNA *HindIII* digested molecular weight marker. Lanes 2 & 3: Lake Chitu metagenomic DNA. Lanes 5 & 6: Lake Shalla metagenomic DNA.

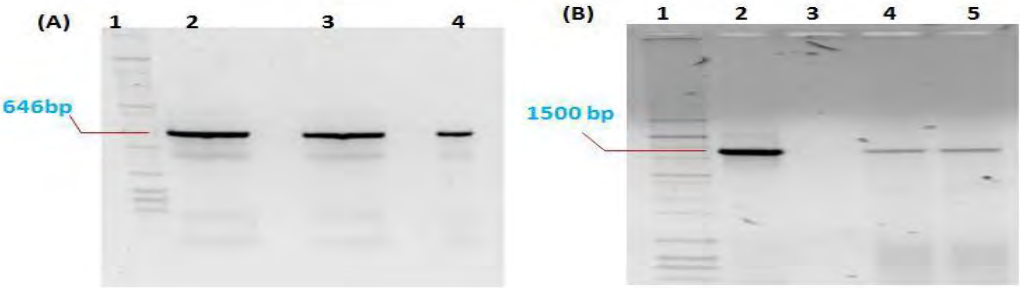


Figure 4.2: Lake Chitu and Shalla Bacterial 16S PCR Products. Lake Chitu and Shalla Actinomycetes (A) and Chitu and Shalla Bacteria (B) Lane 4 (A): positive control from known actinomycete genomic DNA. Lane 4 & 5 (B): Chitu and Shalla Bacteria respectively. Lane 2 (B): positive control from *E. coli* genomic DNA. Lane 3 (B): negative control without template DNA

M13 PCR products of bacterial and actinomycetes recombinant clones yielded expected sizes of 1500bp (for Bacteria) and 646 bP (for Actinomycetes) (Fig. 4.3) and their ARDRA analysis provided several ARDRA types or phylotypes (Fig. 4.4). Those recombinant clones showed false positive and incorrect size of M13 amplicons (E.g. lanes 7 & 17 in Fig. 4.3 A and lanes 3, 5, 16 & 54 in Fig. 4.3 B) were disregard from ERDRA analysis.

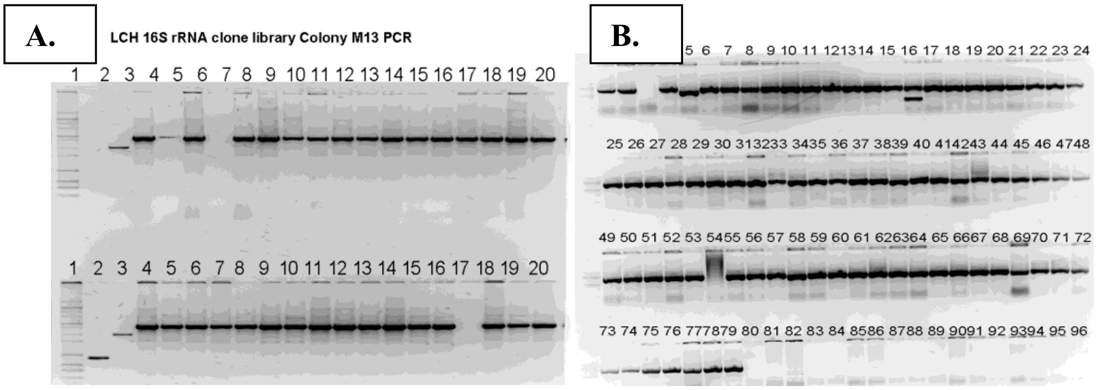


Figure 4.3: Representative M13 Colony PCR gel images for Chitu (A) and Shalla (B) Lakes.

4.2.2 ARDRA analysis of 16S rRNA genes

The number and size of each DNA fragments of the clones generated by restriction analysis showed diverse and unique ARDRA patterns (Fig. 4.4 and Fig 4.5). A total of 334 bacterial clones were analyzed for ARDRA patterns from which 51 unique 16S rDNA bacterial clones (ARDRA types) were from Lake Chitu and 44 bacterial clones were from Lake Shalla. They were identified and sequenced. Similarly, 12 unique 16S rDNA actinomycete clones from Lake Chitu and 13 unique 16S rDNA clones from Lake Shalla were identified and sequenced. In total, 98 actinomycetes clones, 46 from Chitu and 52 from Shalla were analyzed. According to ARDRA and sequence analysis, 28% of Chitu and 25% of Shalla bacterial clones; and 28% of Chitu and 23% of Shalla actinomycetes clones were unique (Table 4.1).

Table 4.1: Total Bacterial and Actinomycetes 16S clones Analysed and unique ARDRA types identified

Taxon/ Lake	Total 16S Clones	ARDRA Types identified	Unique clones
Bacteria	334		
• Lake Chitu (LCH)		51	28 %
• Lake Shalla (LSH)		44	25 %
Actinomycetes	98		
• Lake Chitu (LCH)		12	28%
• Lake Shalla (LSH)		13	23%

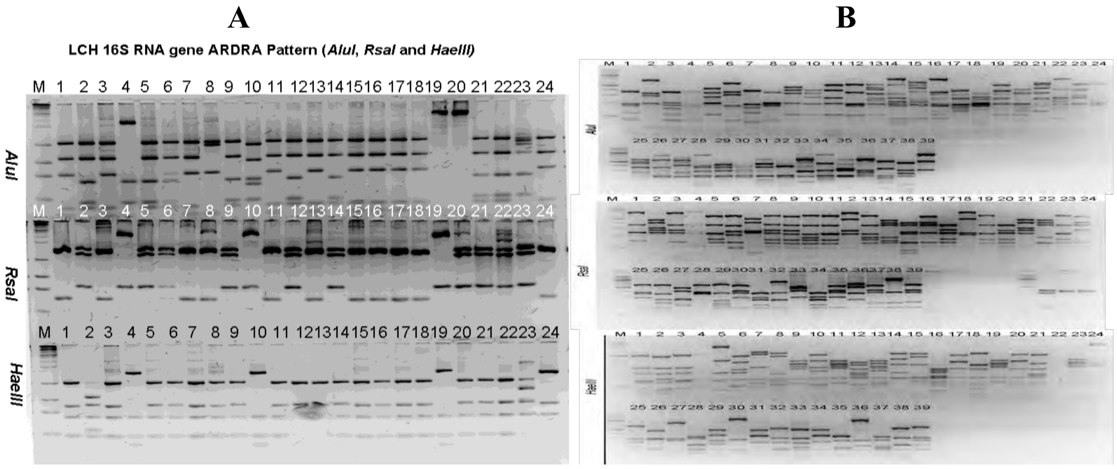


Figure 4.4: Representative gel of ARDRA Patterns for Chitu (A) and Shalla (B) Lakes associated bacterial 16S rDNA clones digested with *AluI*, *RsaI* and *HeaIII*

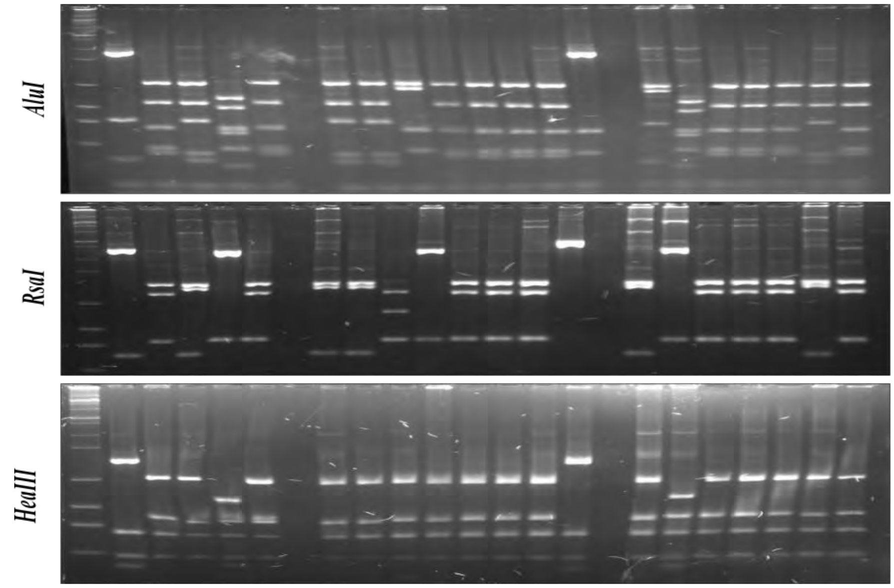


Figure 4.5: Representative gel of ARDRA Patterns for Shalla Lake associated bacterial 16S rDNA clones digested with *AluI*, *RsaI* and *HeaIII*

4.2.3 Sequencing and Phylogenetic data analysis of 16S r RNA Genes

The bacterial and actinomycete 16S rRNA gene libraries were constructed from Chitu and Shalla lakes sediment samples. Significant phylotypes were distinguished by Amplified Ribosomal DNA Restriction Analysis (ARDRA) and partial 16S rRNA sequencing. The total number of 16S rRNA gene sequences of bacterial and actinomycete clone libraries from each lake is shown in Table 4.2.

Table 4.2: Number of Bacteria and Actinomycetes 16S rRNA gene clone sequences from Chitu **and Shalla lakes.**

Lake	*Number of Bacterial clone sequences* **(Accession numbers)**	*Number of Actinomycetes clone sequences* **(Accession numbers)**
Lake Chitu (LCH)	**51** (HM582691-HM582741)	**12** (HM106292-HM106303)
Lake Shalla (LSH)	**44** (HM582642-HM582785)	**13** (HM106304-HM106316)
Total Sequences	**95**	**25**

The bacterial and actinomycetes 16S rDNA sequences obtained from both Chitu and Shalla lakes sediment samples were submitted to NCBI GeneBank Database. They were provided accession numbers as HM582691 to HM582785 for that of the bacteria (95 sequences) and HM106292 to HM106316 for actinomycetes (25 sequences) (Table 4.2). BLAST searches of the Gen-Bank database confirmed that most of the clones in the four libraries were closely related to the 16S rDNA sequences of the uncultured microbes (Table 4.3, 4.4, 4.5 & 4.6 under **Appindex III**).

Of the **51 sequences** retrieved from Lake Chitu sediment samples fell into six major lineages (Table 4.7) of the domain **Bacteria***:* γ-, α-, *ß*- and Δ -*Proteobacteria* (18%, 8% , 6%, and 2% respectively = **34%**), low-G+C-content gram-positive organisms, ***Firmicutes*** **(18% *Clostridium*, 2% *Bacillus*, and 6% *unclassified firmicutes*** = 26%**),** high-G+C-content gram-positive organisms, *Actinobacteria* (**2%**), *Bacteroidetes* **(4% *Bacteroidia*, 4% *Sphingobacteria, and* 14% *unclassified Bacteroidetes*** = **20 %**), and *Opitutae* in the phylum *Verrucomicrobia* (2%) with 33% of unclassified group.

Comparatively, most of the **44 sequences** retrieved from Lake Shalla sediment samples fell into six major lineages (Table 4.7) of the domain **Bacteria**: γ-, α-, *ß*- and Δ -*Proteobacteria* (30%, 9% , 2%, and 2% respectively = **43%**), low-G+C-content gram-positive organisms, ***Firmicutes* 16%** (***Bacillus,* 14%** and ***Clostridium* 2%**), high-G+C-content gram-positive organisms, *Actinobacteria* (**18%**), *Bacteroidetes* **(*flavobacteria and sphingobacteria;* 2%**), ***Plancomycetes*** (**2%**), and ***Acidobacteria*** (**3%**) with 12% of unclassified group.

Table 4.7: Major Bacterial Phyla in Chitu and Shalla Lakes Compared. The 16S rRNA sequences of both lakes were analysed using the online RDP program.

Taxa/Phylum		**Ribotypes Abundance in Lake sediments (%)**		**Remarks**
		LCH	**LSH**	
Protobacteria	γ- *Proteobacteria*	18	30	γ-***Proteobacteria are dominant in both lakes followed by*** α- ***Proteobacteria***
	α- *Proteobacteria*	8	9	
	ß- Proteobacteria	6	2	
	Δ -***Proteobacteria***	2	2	
Firmicutes	*Clostridium*	18	2	*Clostridia* are high in LCH while the *Bacilli* are high in LSH
	Bacillus	2	14	
	Unclassified Firmicutes	6	-	
Actinobacteria		2	18	Considerable abundance in LSH
Bacteriodetes	*Bacteroidia*	2	-	Significant unrepresented or unknown *Firmicutes* and *Bacteriodetes* in LCH
	Flavobacteria	-	2	
	Sphingobacteria	4	2	
	Unclassified Bacteriodetes	14	-	
Verrucomicrobia (Opitutae)		4	-	Not detected in LSH
Tenericutes (*Mollicutes*)		2	-	
Acidobacteria (*Gp3*)		-	3	Not detected in LCH
Plancomycetacia		-	2	Not detected in LCH
Unclassified (unrepresented or unknown)		**12**	**12**	

LCH * (Lake Chitu); LSH* (Lake Shalla)

A. Protobacteria

BLAST analysis showed that numerous clone sequences (approximately 34% for Chitu and 43% for Shalla) were affliated to the phylum *Protobacteria.* (Table 4.5 and 4.6 in Appendix III). The phylogenetic relationships of both lakes (Fig. 4.6 A & B) demonstrated that most of the protobacterial clone sequences were clustered with different sequences from uncultured bacteria derived from various environments including thermal springs, salaine soils and alkaline- saline lakes (Table 4.5 and 4.6 in Appendix III).

(i). *α- Proteobacteria*: a total of 4 clones from each of Chitu and Shalla samples were affiliated to members of the class α- *Proteobacteria*. From the phylogenetic analysis of Lake Chitu, the most common representatives of this group were distantly related (95-98% similarity) to sequences of the genera *Roseinatronobacter* where as the uncultured group of this class. Clone LCH_B139 (HM582728) and LCH_B142 (HM582723) were clustered with *Roseinatronobacter monicus* (DQ659236) (Fig. 4.6 A) with interclone similarity of 98% previously isolated from hypersaline soda Mono Lake California (Boldareva *et al.*, 2007).

The clones belonging to the class α- *Proteobacteria* from Lake Shalla have fromed four groups (Fig. 4.6 B). One of the clones, LSH_B26 (HM582758) was 99% similar (bootstrap 99) to *Rhodobaca bogoriensis* strain (NR 0255089) isolated from African Rift Valley soda lakes (Milford *et al.*, 2000). One group consisting of clone LSH_B25 and LSH_B31 (similarity 99% and bootstrap 98-99) formed a common lineage with two uncultured bacteria identified from alkaline saline soils (GeneBank description). The remaining one clone, LSH_B50 (HM582775) (similarity y 97% and bootstrap was 93) clustered with uncultured *Rhodospirillaceae bacterium* (AM936262) isolated from a hydrocarbon-contaminated soil.

(ii). *ß- Proteobacteria*: a total of 3 clones from Chitu and 1 clone from Shalla were affliated to members of the class *ß- Proteobacteria*. Phylogenetic analysis showed that two clone sequences from lake Chitu, LCH_B128 (HM582714) and LCH_137

(HM582720) were related (98% similarity and bootstrap 99) to *Azoarcus sp*. retrieved from hotspring areas.

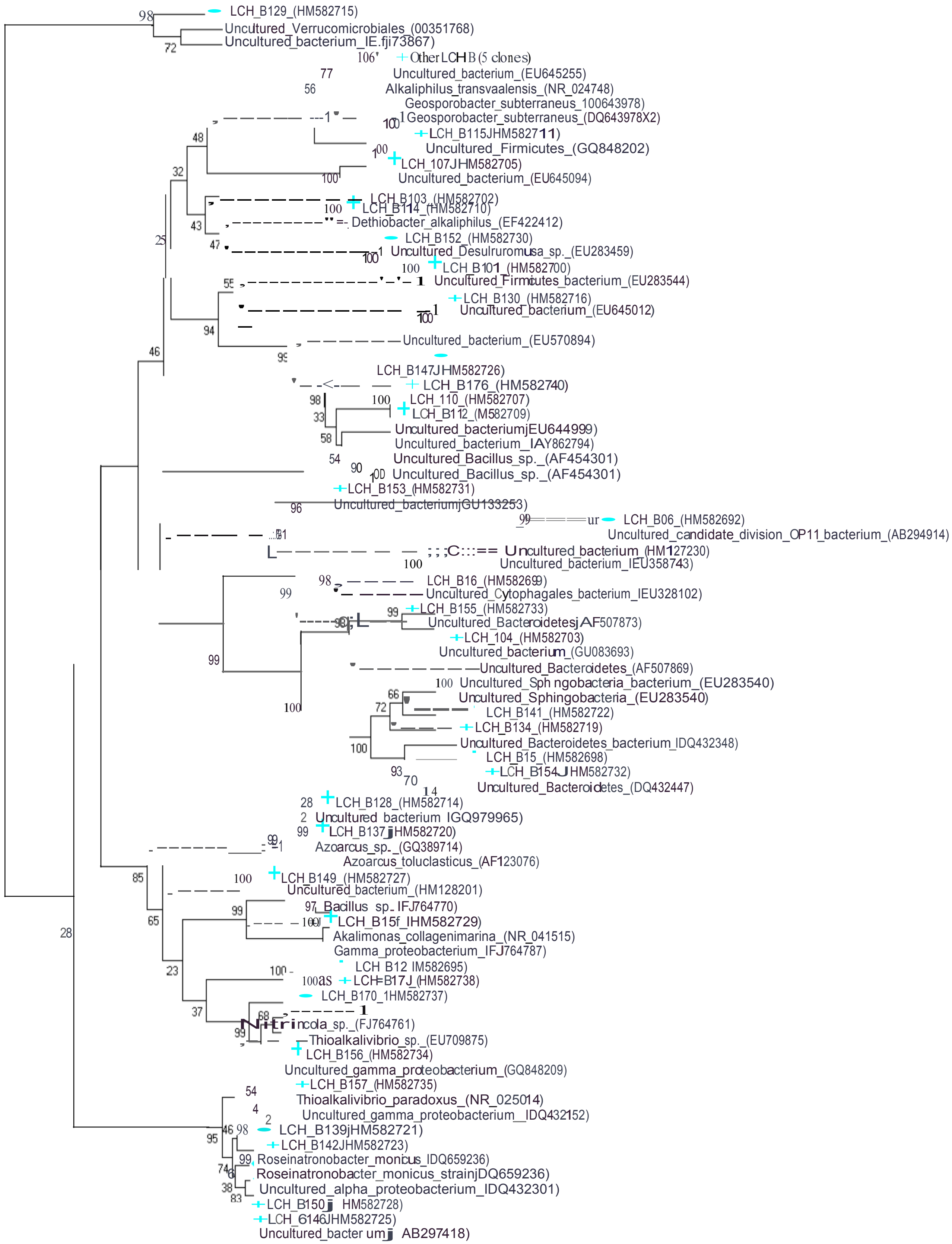
(iii). γ- *Proteobacteria*: BLAST analysis showed that γ- *Proteobacteria* dominated the clone libraries of both Chitu and Shalla lakes. Phylogenetic relationship analsis revealed that a total of twenty-two sequences (9 from Chitu and 13 from Shalla) were clustered with γ- *Proteobacteria* (Fig 4.6 A & B). The γ- *Proteobacteria* clone sequences from Lake Chitu including LCH_B151 (HM582729) and LCH_B170 (HM582737) formed lineages (98% similarity and bootstrap 100) with *Alkalimonas collagenimarina* and *Nitrinicola sp.* respectively identified from deep sea in Japan and Kenyan soda lake sediments respectively, and thus may perhaps belong to these taxa.

Similarly, two Chitu clones including, LCH_B156 and LCH_B157 were clustered with sulphur-oxidizing, alkalophilic bacterium of the genus *Thioalkalivibrio* recovered from soda lake sediments. One of the clones of Lake Shalla, LSH_B49 (HM582774) was separately clustered (99% similarity) with *Methylophaga lonarensis* strain (NR 025569) recovered from Lonar soda lake. The clone LSH_B58 (HM582783) was affliated to *Alkalimonas delamerensis* isolated from alkaline saline lake, Lake Elmenteita, Kenya. Several other clones from both Chitu and Shalla lakes were affiliated to as yet uncultured γ- *Proteobacteria* members of this subclass.

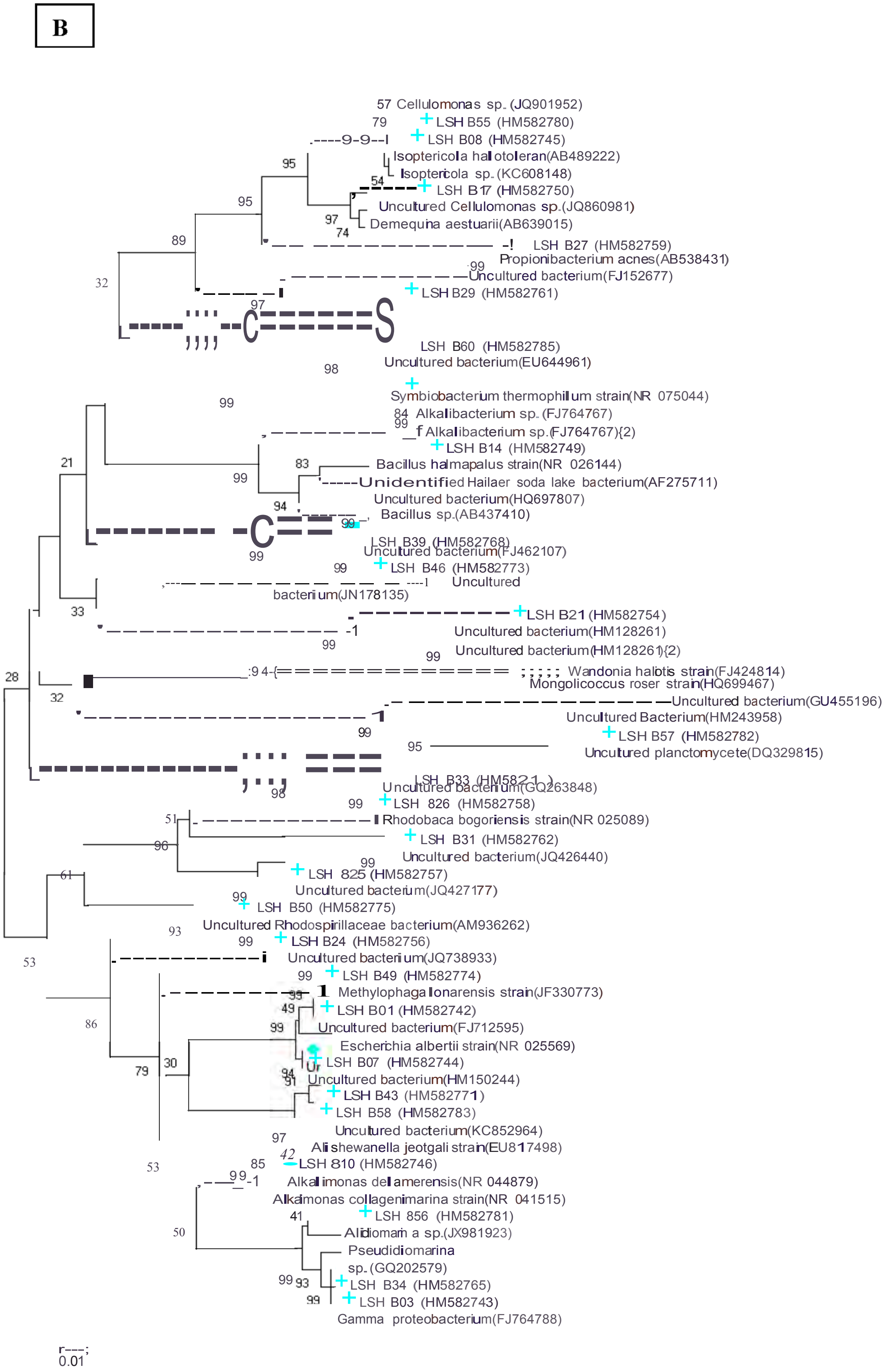
(iv). Δ-*Proteobacteria*: The Δ-*Proteobacteria* clone sequence was represented by one clone from Lake Chitu; LCH_B152 (HM582730) (98% similarity and bootstrap 100) was closely related touncultured *Desulfuromusa sp* (EU283459) which was identified from Anderson lake, USA (GeneBank description).

B. Low-G+C-content gram-positive bacteria (*Firmicutes*)

BLAST analysis showed that Low-G+C-content gram-positive clone sequences dominate bacterial clone libraries of Lake Chitu (18 clones = 35%) and Lake Shalla (8 clones = 20%) (Table 4.5 and 4.6 in Appendix III) next from the γ- *Protobacteria*. Of the 18 clones that fell into the phylum Firmicutes, majority of clones were affiliated to the *Clostridiales* in Lake Chitu; and form a separate cluster from the *Bacillales* (Fig 4.6 A). Four clone sequences (LCH_B147, LCH_B176, LCH_B110, and LCH_B112) from Chitu were distantly clustered with uncultured *Bacillus sp* derived from different soda lakes (Fig 4.6 A, Cluster 1) while nine other clones sequences were affliated with the *Clostridiales* clad (Fig 4.6 A, Cluster 2).


98
LCH_B129_(HM582715)
Uncultured_Verrucomicrobiales_(00351768)
Uncultured_bacterium_IE.fji73867)
72
106'
Other LCHB (5 clones)
77
Uncultured_bacterium_(EU645255)
56
Alkaliphilus_transvaalensis_(NR_024748)
Geosporobacter_subterraneus_100643978)
Geosporobacter_subterraneus_(DQ643978X2)
100
LCH_B115JHM582711)
48
100
Uncultured_Firmicutes_(GQ848202)
32
LCH_107JHM582705)
100
Uncultured_bacterium_(EU645094)
LCH_B103_(HM582702)
100
LCH_B114_(HM582710)
43
Dethiobacter_alkaliphilus_(EF422412)
25
LCH_B152_(HM582730)
47
Uncultured_Desulruromusa_sp._(EU283459)
100
100
LCH_B101_(HM582700)
55
Uncultured_Firmicutes_bacterium_(EU283544)
LCH_B130_(HM582716)
Uncultured_bacterium_(EU645012)
100
94
46
Uncultured_bacterium_(EU570894)
99
LCH_B147JHM582726)
LCH_B176_(HM582740)
98
100
LCH_110_(HM582707)
33
LCH_B112_(M582709)
58
Uncultured_bacteriumjEU644999)
Uncultured_bacterium_IAY862794)
54
Uncultured_Bacillus_sp._(AF454301)
90
100
Uncultured_Bacillus_sp._(AF454301)
LCH_B153_(HM582731)
96
Uncultured_bacteriumjGU133253)
99
LCH_B06_(HM582692)
Uncultured_candidate_division_OP11_bacterium_(AB294914)
81
Uncultured_bacterium_(HM127230)
100
Uncultured_bacterium_IEU358743)
98
LCH_B16_(HM582699)
99
Uncultured_Cytophagales_bacterium_IEU328102)
99
LCH_B155_(HM582733)
98
Uncultured_BacteroidetesjAF507873)
LCH_104_(HM582703)
99
Uncultured_bacterium_(GU083693)
Uncultured_Bacteroidetes_(AF507869)
100
100
Uncultured_Sphngobacteria_bacterium_(EU283540)
66
Uncultured_Sphingobacteria_(EU283540)
72
LCH_B141_(HM582722)
LCH_B134_(HM582719)
100
Uncultured_Bacteroidetes_bacterium_IDQ432348)
LCH_B15_(HM582698)
93
LCH_B154JHM582732)
70
14
Uncultured_Bacteroidetes_(DQ432447)
28
LCH_B128_(HM582714)
2
Uncultured_bacterium_IGQ979965)
99
LCH_B137jHM582720)
99
Azoarcus_sp._(GQ389714)
Azoarcus_toluclasticus_(AF123076)
85
100
LCH_B149_(HM582727)
Uncultured_bacterium_(HM128201)
97
Bacillus_sp._IFJ764770)
99
100
LCH_B15f_IHM582729)
65
Akalimonas_collagenimarina_(NR_041515)
Gamma_proteobacterium_IFJ764787)
23
100
LCH_B12_IM582695)
100
LCH_B17J_(HM582738)
LCH_B170_1HM582737)
37
68
Nitrincola_sp._(FJ764761)
99
Thioalkalivibrio_sp._(EU709875)
LCH_B156_(HM582734)
Uncultured_gamma_proteobacterium_(GQ848209)
28
LCH_B157_(HM582735)
54
Thioalkalivibrio_paradoxus_(NR_025014)
4
Uncultured_gamma_proteobacterium__IDQ432152)
2
46
98
LCH_B139jHM582721)
95
LCH_B142JHM582723)
99
Roseinatronobacter_monicus_IDQ659236)
74
6
Roseinatronobacter_monicus_strainjDQ659236)
38
Uncultured_alpha_proteobacterium_IDQ432301)
83
LCH_B150j HM582728)
LCH_6146JHM582725)
Uncultured_bacter umj AB297418)


0.01


B
57 Cellulomonas sp. (JQ901952)
79
LSH B55 (HM582780)
LSH B08 (HM582745)
99
Isoptericola halotolerans(AB489222)
Isoptericola sp. (KC608148)
95
54
LSH B17 (HM582750)
95
Uncultured Cellulomonas sp.(JQ860981)
97
74
Demequina aestuarii(AB639015)
89
LSH B27 (HM582759)
Propionibacterium acnes(AB538431)
99
Uncultured bacterium(FJ152677)
32
LSH B29 (HM582761)
97
LSH B60 (HM582785)
Uncultured bacterium(EU644961)
98
Symbiobacterium thermophilum strain(NR 075044)
99
84 Alkalibacterium sp. (FJ764767)
99
Alkalibacterium sp.(FJ764767){2)
LSH B14 (HM582749)
21
83
Bacillus halmapalus strain(NR 026144)
99
Unidentified Hailaer soda lake bacterium(AF275711)
Uncultured bacterium(HQ697807)
94
Bacillus sp.(AB437410)
99
LSH B39 (HM582768)
Uncultured bacterium(FJ462107)
99
99
LSH B46 (HM582773)
Uncultured
bacterium(JN178135)
33
LSH B21 (HM582754)
Uncultured bacterium(HM128261)
99
99
Uncultured bacterium(HM128261){2)
28
94
Wandonia haliotis strain(FJ424814)
Mongolicoccus roser strain(HQ699467)
32
Uncultured bacterium(GU455196)
Uncultured Bacterium(HM243958)
99
LSH B57 (HM582782)
95
Uncultured planctomycete(DQ329815)
LSH B33 (HM5821)
Uncultured bacterium(GQ263848)
98
99
LSH 826 (HM582758)
51
Rhodobaca bogoriensis strain(NR 025089)
96
LSH B31 (HM582762)
Uncultured bacterium(JQ426440)
99
LSH 825 (HM582757)
61
Uncultured bacterium(JQ427177)
99
LSH B50 (HM582775)
Uncultured Rhodospirillaceae bacterium(AM936262)
93
99
LSH B24 (HM582756)
53
Uncultured bacterium(JQ738933)
99
LSH B49 (HM582774)
99
Methylophaga lonarensis strain(JF330773)
86
49
LSH B01 (HM582742)
99
Uncultured bacterium(FJ712595)
Escherichia albertii strain(NR 025569)
79
30
LSH B07 (HM582744)
94
91
Uncultured bacterium(HM150244)
LSH B43 (HM582771)
LSH B58 (HM582783)
Uncultured bacterium(KC852964)
97
42
Alishewanella jeotgali strain(EU817498)
53
85
LSH 810 (HM582746)
99
Alkalimonas delamerensis(NR 044879)
Alkalimonas collagenimarina strain(NR 041515)
41
LSH 856 (HM582781)
50
Alidiomarina sp.(JX981923)
Pseudidiomarina
sp. (GQ202579)
99 93
LSH B34 (HM582765)
99
LSH B03 (HM582743)
Gamma proteobacterium(FJ764788)
0.01

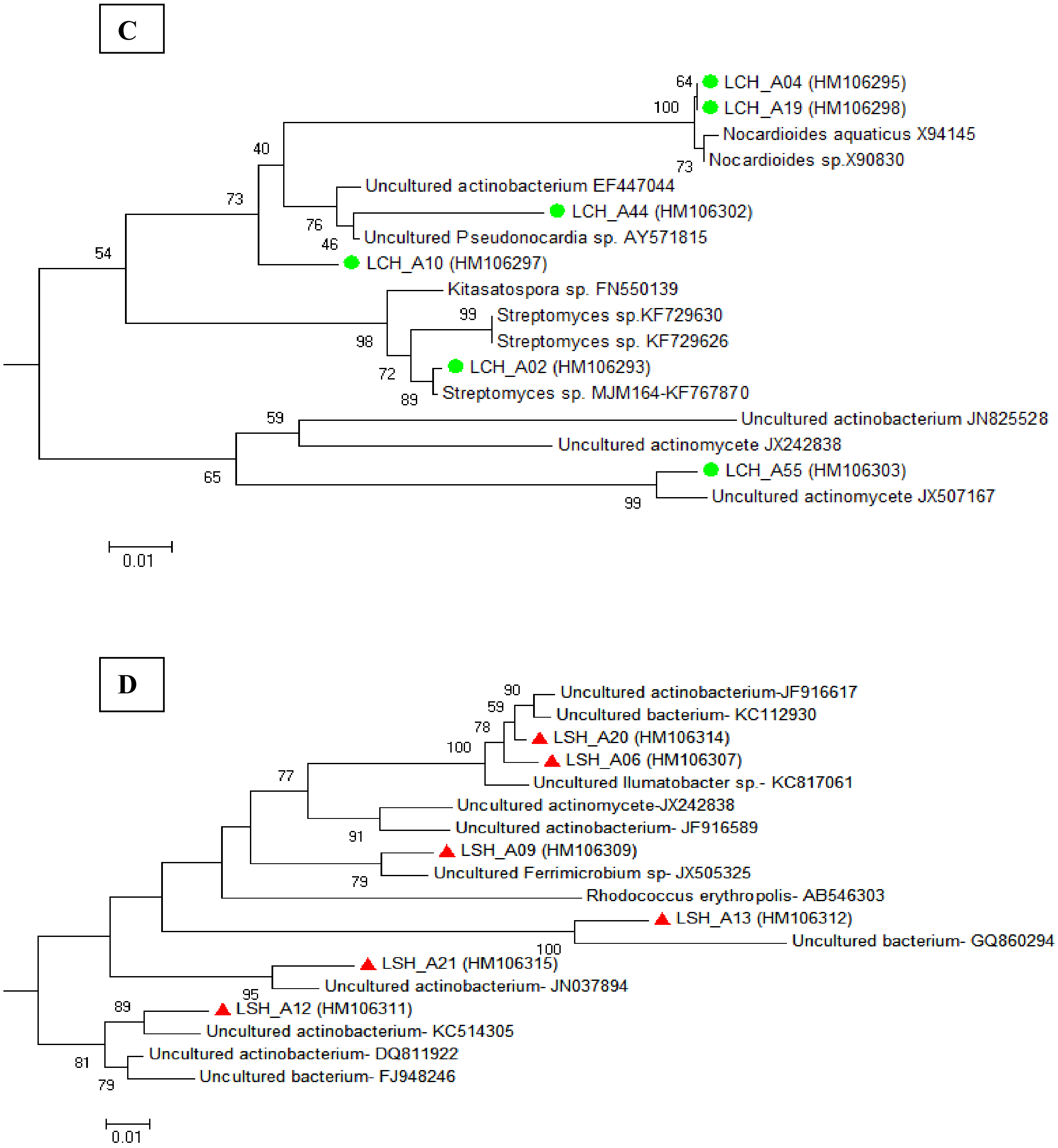


Figure 4.6 A-D: Phylogenetic Trees of Bacterial 16S for Chitu and Shalla Lakes. The trees were constructed using the Junkes and Cantor (1969) evolutionary distance matrix and neighbor-joining tree topology using MEGA 5.2. The values indicate the percentage of occurrence in 1000 boostrapped tree, and the scale bar represents 0.01 nucleotide substitution. A: Middle rooted phylogenetic tree of Lake Chitu Bacteria; B: Middle rooted phylogenetic tree of Lake Shalla Bacteria; C: Middle rooted phylogenetic tree of Lake Chitu Actinomycetes; D: Middle rooted phylogenetic tree of Lake Shalla Actinomycetes. Clone sequences characterized in this study are indicated with symbols and accession numbers are shown in parenthesis.

In Lake Chitu, LCH_B114 (HM582702) had maximum similarity (99% and bootstrap of 100) with *Dethiobacter alkaliphilus* (EF422412) isolated from Inner Mongolia, Xiarinur Soda Lake. In Lake Shalla, LSH_B60 (HM582785) had maximum similarity (85% and bootstrap value of 99) to *Symbiobacterium thermophilum* strain (NR075044) from culture collection (Database description). Besides, LSH_B14 (HM582749) showed close similarity (99% and bootstrap 99) with *Alkalibacterium sp.* (FJ764767) isolated from Lake Elmenteita, Kenya. Most of the remaining clones from both Chitu and Shalla were grouped together with uncultured bacteria associated with hypersaline lakes, soda lakes, and gold mine habitates.

C. High G+C-content gram-positive bacteria (*Actinobacteria*)

Several sequences affiliated to the phylum Actinobacteria were detected from Lake Shalla. The clone sequences were related to species of the genus *Cellulomonas*, *Isoptericola*, *Demequina*, and *Propionibacterium* and sequences associated with yet uncultured bacteria and *Cellulomonas sp*. (Fig. 4.6 B). Phylogenetic analysis of Actinomycetes of Lake Chitu revealed that three clone sequences (LCH_A04, LCH_A19, and LCH_A02) were similar with members of the genus *Nocardioides*, *Kitasatospora*, and *Streptomyces* (Fig. 4.6 C). The clone sequences LCH_A55, and LCH_A44 showed close similarity y to uncultured actinomycetes and *Pseudonocardia sp*. (Fig. 3.6 C). Similarly, in Lake Shalla, the clone sequences, LSH_A06 and LSH_A09 were related to uncultured species of the genus *llumatobacter* and *Ferrimicrobium,* respectively (Fig. 4.6 D). Four other clone sequences (LSH_A20, LSH_A13, LSH_A21, and LSH_A12) were affiliated to the sequences of uncultured bacterium or actinobacterium (Fig. 4.6 D). Thirteen (13) actinobacterial clone sequences (6 clones from Chitu and 7 clones from Shalla lake sediments) formed separate lineage and were not related to any previously documented 16S rRNA gene sequences (not shown in Fig. 4.6 C and D).

D. Cytophaga-Flavobacterium-Bacteroidetes (CFB)

The sequence analysis indicated that 10 clone sequences (20%) from Lake Chitu and 3 clonesequences (7%) from Lake Shalla were grouped within the members of the CFB clade derived from saline soils, Soap lake, Mono lake, and Inner Mongolia lake. In Lake Chitu, 3 clones (LCH_B16, LCH_B155, and LCH_B104) showed close identity with sequences of uncultured *Cytophagales* and *Bacteroidetes* (Fig 4.6A) recovered from saline lakes of Inner Mongolia, Chaina and Mono Lake, California. Four other clones (LCH_B141, LCH_B134, LCH_B15, and LCH_B154) showed weak association with sequences of uncultured *Shingobacteria* and *Bacteroidetes* recovered from the soda lakes of Wadi Natrun in Egypt and Mono Lake in California. From Lake Shalla, only one clone, LSH_B21 (HM582764), formed a separate taxon with the highest sequence similarity (87% and bootstrap value of 99) with uncultured bacterium (NCBI, HM128261) obtained from Lake Xiaochaidan in Tibet.

E. Other Bacterial divisions

One clone (LSH-B57) from Lake Shalla was grouped with uncultured *Planctomycete* (DQ329815) identified in hypersaline microbial mat. Another clone sequence from Lake Chitu was clustered with the uncultured *Verrucomicrobiales* (DQ351768) identified from agricultural soil. The clone sequence LCH_B06 (HM582692) from Lake Chitu was found to relate to the uncultured *Candidate division OP11 bacterium* (AB294914) detected from a hotspring.

From the sequence analysis of the bacterial domain 29 clone sequences (11 clones from Chitu and 18 clones from Shalla lake sediments) were affliated with various clades forming separate lineage that were not related to any previously documented 16S rRNA gene sequences (not shown in Fig. 4.6 A and B).

4.3 Discussion

The phylogenetic analysis of 16S rRNA clone libraries of Chitu and Shalla lakes revealed sequences affliated with low and high G+C Gram-positive bacteria (*Firmictes* and *Actinobacteria*), *Plancctomycetes*, CFB, and *Protobacteria* of the γ-, α-, *ß*- and Δ-*Proteobacteria* subdivisions. Most of the clones allied with members of the γ- subdivision of *Protobacteria* and *Firmicutes* (Fig 4.6 A & B). The results of this study suggest that both lakes harbor significant rich diversity and unique representatives of bacterial community.

This was exemplified by the fact that 57 clones (60%) isolated from both lakes were more closely affliated to uncultured members of the Domain bacteria. The remaining sequences were related to 16S rRNA gene sequences of cultivated organisms or to environmental sequences recovered from various sites. It is interesting to note that a few clone sequences in this study were found to be different from sequences previously retrieved from the East African soda lakes (Mwirichia *et al*., 2011; Rees *et al*., 2004). This indicates that a large proportion of the microbial population from these lakes has never been cultured, suggesting the need to describe the functional roles of these members of the microbial phylum.

The anoxygenic, phototrophic, alkalophilic purple nonsulfur bacterium, *bogoriensis* was detected from α--*Protobacteria* in Lake Shalla (LSH B26, HM582758) (Fig 4.6 B). *Rhodobaca bogoriencsis* was originally isolated from Lake Bogoria, Kenya and it is capable of both phototrophic and chemotrophic growth on wide variety of organic compounds (Milford *et al*., 2000).

Alkalimonas collagenimarina represented by clone LCH_B151 (HM582729) in Lake Chitu is an obligately alkaliphilic, collagenolytic enzyme- producing bacterium **(**Kurata *et al*., 2010) and found to be intimately affiliated to members of the genus *Alkalimonas.* This indicates that Chitu and Shalla lakes harbor biotechnologically significant representatives of microbes within the dominant group of the class **γ-**protobacteria. In the same genus, *Alkalimonas delamerensis***,** a strictly aerobic, heterotrophic, alkaliphilic, mesophilic bacteria which was isolated from Lake Elmenteita in Kenya (Ma *et al*., 2004a), was also represented in Lake Shalla and may play similar roles. *Nitrinicola sp* represented by LCH_B170 (HM582737) in Lake Chitu belonged to the member of Salt-tolerant rhizobacteria isolated from Kenyan soda lake sediments (GeneBank description). The species of the genus *Nitrinicola* has recently been used to induce tolerance and enhanced plant growth promotion of wheat in salt stressed soil (Tiwari *et al*., 2011).

Previous studies showed that several soda lakes contain members of bacterium of the genus *Thioalkalivibrio* associated sulfur oxidation in the soda lake as well as denitrification and thiocyanate utilization, several of which belonging to the **γ**- and α-*protobacteria* (Sorokin *et al*., 2001a; Sorokin and Kuenen, 2005b). Two clones, LCH_B156 and LCH_B157 represented the **a**lkalophilic bacterium of the genus *Thioalkalivibrio* in Chitu soda lake sediments may thus play similar roles in the oxidative part of sulfur cycle in the lake (Fig 4.6 A). The occurence of one the clones of Lake Chitu affliated closely to uncultured member of the genus *Desulfuromusa* may also confirm the the presence of yet uncultured sulfate-reducing halo-alkaliphiles as main actors in sulfur cycle in this lake.

Interestingly, the genus *Desulfuromusa* was not previously reported at least from geographically nearby soda lakes of East Africa. The Clone LSH_B49 (HM582774) from Lake Shalla was separately clustered closely with *Methylophaga lonarensis* strain. *Methylophaga lonarensis* strain is a haloalkaliphilic methylotroph (non-methane utilizing) isolated recently from Lonar Lake, a saline and alkaline lake in Maharashtra, India (Antony *et al*., 2012a). It utilizes methanol as its sole carbon and energy source (Antony *et al*., 2012a).

BLAST analysis showed that many of the clones generated from Chitu and Shalla lakes were phylogenetically related to sequences from Low-G+C-content gram-positive bacteria (*Firmicutes*) next to the **γ**- *protobacteria* (Table 4.5 and 4.6 in Appendix III). Several of the bacterial sequences showed similarity to sequences from the group of *Bacillus* and *Clostridium*.

A few clone were also identified with *Alkaliphilus*, *Alkalibacterium*, and *Thermophilum* group (Fig. 4.8 A and B). *Dethiobacter alkaliphilus* was also found in Lake Chitu and this bacterium was recently identified as a member of the class *Clostridia* within the *Firmicutes* as sulfate-reducing bacteria (Poster, *et al*., 2013). A number of sequences related to this low-G+C-content gram-positive bacterial clad playing similar ecological role have been identified from sulfide-rich, anoxic marine or deep sea sediments (Bowman *et al*., 2000b; Li *et al*., 1999) as well as Lonar Soda Lake in India (Wani *et al*., 2006).

A few CFB (Cytophaga-Flavobacterium-Bacteroidetes) phylotypes were identified in Shalla and Chitu Lakes and showed weak identity to known sequences (Fig. 4.6 A & B), which may represent new taxa. Previous studies showed that members of the CFB including *Cytophagales* and *Bacteroidales* are associated with mineralization of complex organic matters in several soda lakes (Humayoun *et al*., 2003; Mwirichia *et al*., 2011; Wani *et al*., 2006).

4.4 Conclusion

The phylogenetic compositions of bacterial communities of Chitu and Shalla Lakes were studied by sequencing 16S rRNA gene clones using Bacterial and Actinomycetes domain specific primers. Four 16S rRNA clone libraries were constructed (Bacterial and actinomycetes clone libraries for each lake). The data revealed that sequences were mostly affliated with low and high G+C Gram-positive bacteria (*Firmictes* and *Actinobacteria*), CFB, and *Protobacteria* of the **γ**-, α-, *ß*- and Δ- ***Proteobacteria*** subdivisions. From the study, it can be concluded thatt both lakes harbor significant higher rich diversity and unique representatives of bacterial overall diversity and community composition. In both lakes, most of the clones were phylogenetically allied with members of the **γ**- subdivision of *Protobacteria* (accounted for 34–43% of sequences)*,* followed by *Firmicutes* (affiliated sequences at 16–26%).

CHAPTER FIVE

Analysis of Prokaryotic Communities of Chitu, Shalla and Abijata Soda Lakes Using Barcode-Tagged 16S rRNA gene pyrosequencing

5.1 Introduction

Chitu, Shalla, and Abijata are part of the East African Riftvalley soda lakes characterized by unique environments with double extremes of high pH and salinity and yet, support vastly productive and diverse functional novel genera and species of autotrophic and heterotrophic microorganisms (Antony, *et al*., 2013; Grant and Sorokin, 2011; Lanzen *et al.,* 2013). Previous works on these lakes were mostly limited on geochemical, limnological, fisheries and algal composition studies (Ayenew, 2007; Kebede *et al*., 1994; Tadesse *et al*., 2003; Tilahun and Ahlgren, 2010). However, only a few works have been reported with regads to their microbilogical composition and biotechnological applications based on cultivation techniques (Gessesse and Gashe, 1997; Gessesse et al., 2003; Haile and Gessesse, 2012; Lanzen *et al.,* 2013).

Currently, metagenomic studies of microbial diversity with much deeper analysis of microbial composition and structure could be made with the emergence and rapid improvement of high-throughput sequencing technologies; notably, Roche 454 and Illumina sequencing (Metzker 2010). The use of pyrosequencing-based 16S rRNA technique has emerged as a powerful strategy to profile and characterize microbial communities from diverse and mixed ecological samples (Sogin *et al.,* 2006; Huber *et al*., 2007; Humblot and Guyot, 2009; Dadheech *et al*., 2013).

Principally, the bar-code-tag primer-based pyrosequencing technology enables deeper community analysis of environmental prokaryotic diversity by generating large number of sequence data for pooled multiple samples in a single-run sequencing plateform (Parameswaran *et al*., 2007; Tringe and Hugenholtz 2008; Yu *et al*., 2005). This method is more robust and cost-effective compared to traditional Sanger sequencing and can reveal the microbial community structure from the genus to the species level.

In this study, the microbiome community structure and diversity of naturally occurring and inadequately studied sediment samples of soda lakes of Chitu (LCH), Shalla (LSH) and Abijata (ABJ) were analyzed using pyrosequencing-based barcode-tagged 16S rRNA genes. QIIME algorithm pipeline was used to analyze the hypervariable pyrosequences of the barcoded 16S rDNA V3–V4 regions for the identification of prokaryotic communities essentially at the phylum, class, family, and genus levels (Caporaso *et al.*, 2010).

The objectives of the study were:

- Isolation of pure, intact high molecular weight DNA from Lake sediment samples
- PCR amplification of the V3-V4 regions of 16S rRNA genes
- High throughput 454 pyrosequencing of tagged 16S rRNA gene amplicons
- Pyrosequence analysis of tagged 16S rRNA genes

5.2 Result

5.2.1 Pyrosequence tagged 16S rRNA gene amplicons and analysis

Metagenomic DNA extraction and purification were performed from the three soda lake sediments according to the methods described in Chapter two, section 2.3. A 466-nucleotide sequence of the V3-V4 region of the 16S rRNA genes was successfully amplified by PCR (Fig. 5.1) in triplicate for each of the three lake samples and a second PCR or barcoding PCR was performed to add 9 nucleotides tags (Fig. 5.2) then purified from gel (Fig. 5.3).

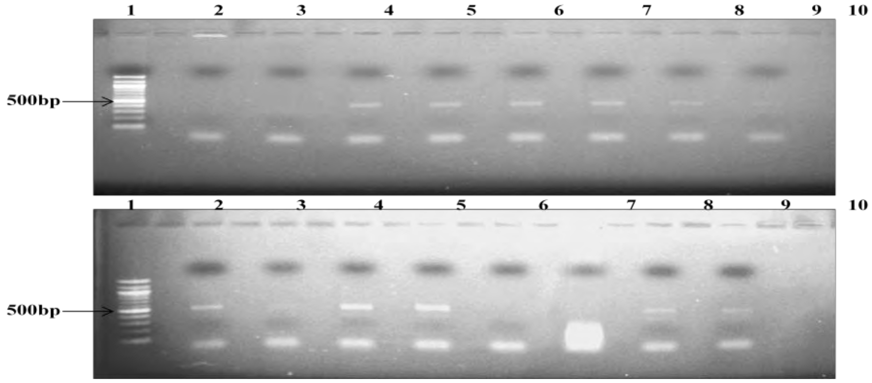


Figure 5.1: PCR PCR of LCH, LSH and ABJ V3-V4 regions of the 16S rRNA genes

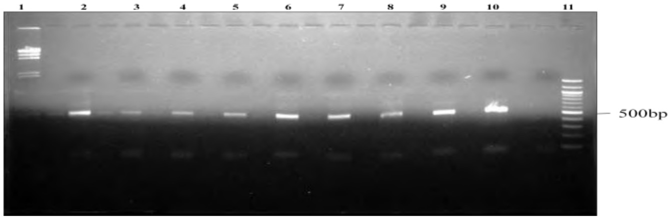


Figure 5.2: Bar-coding PCR products (Second PCR) of LCH (Lanes 2-4), LSH (Lanes 5-7) and ABJ (Lanes 8-10) with addition of Adapter and A & B. Lane1 & 11 are molecular DNA ladders.

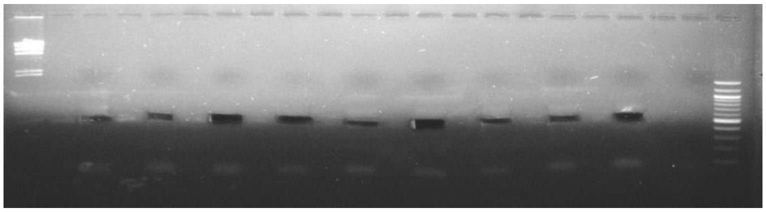


Figure 5.3: Gel excision of Products of 16S rRNA genes LCH, LSH and ABJ PCR as in Fig.4.2 for Pyrosequencing.

QIIME algorithm pipeline was used to analyze the hypervariable pyrosequences of the barcoded amplicons of the 16S rDNA V3–V4 regions generated from the three soda lakes. From the three soda lake samples, pyrosequencing of metagenomic DNA of LCH, LSH and ABJ generated 151, 289 total sequencing reads (61, 977 sequences for LCH, 42, 988 for LSH and 46, 324 for ABJ). These included 3, 835 quality unique sequence reads (OTUs), with mean lengths of 369 bp (taken from raw data of QIIME pyrosequence analysis pipline).

5.2.2 Microbial community composition and distribution

The results from 454 pyrosequence data showed that bacterial pyrosequence reads represented more than 90% of the prokaryotic sequencing reads of all the three sampling lakes were affiliated to 28 microbial phyla and approximately 199 genera (taken from raw data of QIIME algorithm pipeline, see Appendix V. Fig 5.4 for representative phyla). Of the total sequence reads of all the three lakes, the most dominant phylum was *Protobacteria* represented by (52.3 %) followed by the *Firmicutes* (12.7%) (Table 5.1). Other prominant phyla greater than 1% of all sequence reads included the *Cyanaobacteria* (3.1%), *Bacteriodetes* (2.6%), *Euryarchaeota* (2.1%), *Chloroflexi* (1.7%), *Deinococcus-Thermus* (1.4%) *Actinobacteria* (1.1%), *Acidobacteria* (0.9%), and unclassified taxa (0.5%) (Table 5.1).

The other phyla and candidate divisions represented by less than 0.5% of the total prokaryotic communities in the lakes were *Verrucomicrobia*, *Thermotogae*, *Tenericutes*, *Spirochaetes*, *Planctomycetes*, *Crenarchaeota*, and several candidate divisions (BRC1, OD1, OP10, SR1, TM7, and WS3) (Table 5.1). Among the *Proteobacteria*, the *ß-proteobacteria* (32.2%) were represented by the dominant family *Comamonadacaea and Alcaligenacea* followed by *Gamma-Protobacteria* (9.9%), *Alpha-protobacteria* (8.7%) and *Delta-protobacteria* (1.3%) in the total sequence reads of the three lakes (Table 5.3).

Table 1 5.1 Prokaryotic Community Distribution in Chitu, Shalla and Abijata Lakes Sediment Samples. The table shows the relative % of pyrosequence reads (triplicates for each sample) of different phyla in each Lake as in Fig. 5.4. (MID51- MID59 is sample ID of the lakes). Phyla with less than 0.0% of total sequence reads were omitted from the table.

Taxonomy	**Total Seq. reads** (%)	**MID51**	**MID53**	**MID52**	**MID57**	**MID58**	**MID59**	**MID54**	**MID55**	**MID56**
		LCH (%)	**LCH** (%)	**LCH** (%)	**ABJ** (%)	**ABJ** (%)	**ABJ** (%)	**LSH** (%)	**LSH** (%)	**LSH** (%)
Archaea;Crenarchaeota	**0.4%**	**0.1%**	**0.2%**	**0.1%**	**0.6%**	**0.9%**	**1.0%**	**0.0%**	**0.1%**	**0.2%**
Archaea;Euryarchaeota	**2.1%**	**0.3%**	**0.4%**	**0.3%**	**3.5%**	**4.0%**	**5.4%**	**1.0%**	**1.3%**	**2.7%**
Archaea;Other	0.5%	0.1%	0.0%	0.0%	1.1%	0.6%	1.3%	0.2%	0.2%	0.5%
Bacteria;Acidobacteria	**0.9%**	**0.6%**	**0.7%**	**1.0%**	**0.4%**	**1.0%**	**0.6%**	**1.2%**	**1.7%**	**1.2%**
Bacteria;Actinobacteria	**1.1%**	**0.7%**	**0.7%**	**0.7%**	**1.1%**	**1.1%**	**1.1%**	**1.1%**	**1.8%**	**1.6%**
Bacteria;Bacteroidetes	**2.6%**	**0.8%**	**0.6%**	**0.7%**	**5.5%**	**3.8%**	**3.5%**	**2.9%**	**2.7%**	**3.0%**
Bacteria;Chloroflexi	**1.7%**	**0.4%**	**0.4%**	**0.3%**	**3.5%**	**4.4%**	**4.7%**	**0.5%**	**0.7%**	**0.6%**
Bacteria;Cyanobacteria	**3.1%**	**0.3%**	**0.2%**	**0.1%**	**9.5%**	**8.8%**	**9.1%**	**0.1%**	**0.0%**	**0.2%**
Bacteria;Deinococcus-Thermus	**1.4%**	**1.3%**	**1.5%**	**1.3%**	**1.2%**	**1.1%**	**1.1%**	**1.4%**	**1.6%**	**1.9%**
Bacteria;Firmicutes	**12.7%**	**20.1%**	**19.5%**	**19.8%**	**6.3%**	**7.5%**	**6.9%**	**11.7%**	**11.8%**	**10.8%**
Bacteria;Other	**19.2%**	**6.2%**	**6.1%**	**7.1%**	**40.3%**	**44.5%**	**45.8%**	**6.4%**	**8.5%**	**8.2%**
Bacteria;Planctomycetes	**0.1%**	0.0%	0.0%	0.0%	**0.1%**	**0.2%**	**0.1%**	0.0%	**0.1%**	0.0%
Bacteria;Proteobacteria	**52.3%**	**68.7%**	**69.4%**	**68.1%**	**23.6%**	**18.8%**	**16.6%**	**71.6%**	**67.4%**	**66.7%**
Bacteria;Spirochaetes	**0.2%**	0.1%	0.1%	0.0%	0.3%	0.3%	0.3%	0.1%	0.2%	0.0%
Bacteria;TM7	**0.1%**	0.0%	0.0%	0.0%	0.1%	0.2%	0.1%	0.2%	0.1%	0.0%
Bacteria;Tenericutes	**0.1%**	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%	0.1%	0.0%
Bacteria;Thermotogae	**0.2%**	0.1%	0.1%	0.2%	0.3%	0.6%	0.5%	0.1%	0.0%	0.0%
Bacteria;Verrucomicrobia	**0.5%**	0.0%	0.1%	0.1%	1.1%	1.0%	0.8%	0.4%	0.5%	0.5%
Bacteria;WS3	**0.2%**	0.0%	0.0%	0.0%	0.3%	0.5%	0.3%	0.2%	0.2%	0.3%
Unclassified;Other	**0.5%**	**0.1%**	**0.0%**	**0.0%**	**0.8%**	**0.5%**	**0.5%**	**0.7%**	**0.7%**	**1.4%**

In view of the relative microbial abundance in the lakes, the *Proteobacteria* was the dominant phylum with 69% of the total sequences of each LCH and LSH sediment samples (Table.5.2) of which 20% of the sequences obtained from Lake ABJ sediment samples were represented by *ß-proteobacteria*. The second most abundant phylum in both LCH and LSH was the *Firmicutes*, represented by approximately 11% and 20% of the sequenses respectively. In lake ABJ, each of the phylum *Bacteriodetes* and *Chloroflexi* largely represented by more than 4% of the total sequences and nearly 9% of the sequences were in the *Cyanobacteria* phylum (Table.5.2). Approximately 0.5% -1.5% of the sequences from each sample remained unclassified and 19% of the sequences represented unusual microbial groups (from raw data of QIIME algorithm pipeline).

Table 5.2: Major Dominant Phyla Represented in Chitu, Shalla and Abijata lakes. Numbers show the average % of pyrosequence reads in each lake.

Major Microbial Phyla	Chitu (LCH)	Shalla (LSH)	Abijata (ABJ)
Actinobacteria	0.70	1.50	1.10
Bacteroidetes	0.70	2.87	4.27
Chloroflexi	0.37	0.60	4.20
Cyanobacteria	0.20	0.10	9.13
Deinococcus-thermos	1.37	1.63	1.13
Firmicutes	19.80	11.43	6.90
Protobacteria	68.73	68.57	19.00
Unclassified	0.03	0.93	0.60

The most abundant genera in all the three soda lakes included the *Delftia* (28.5%) from Beta-proteobacteria followed by *Anoxybacillus* (9.3%) from Firmicutes, *Brevundimonas* (4.4%) from Alpha-protobacteria, and *Acinetobacter* (3.8%) species from Gamma-Protobacteria. Other genera included *Thioalkalivibrio* (1.1%), *Achromobacter* (1.1%), *Ralstonia* (1.1%), *Massilia* (0.8%), *Truepera* (0.7%) and *Meiothermus* (0.7%) of all sequences. Among others, the species of the genera *Delftia* (45%) and *Brevundimonas* (7%) were the most dominant genera from the *Betaprotobacteria* and *Alphaprotobacteria,* respectively in both Chitu and Shalla lakes.

The species of the genera *Acinetobacter* (10%) *Anoxybacillus* (19%) were largely detected in Chitu Lake, representing the *Gamaprotobacteria* and *Firmicutes*, respectively. Compared to other classes of the *Protobateria,* Abijata lake contained much of the *Gammaprotobacteria* dominantely represented by the species of the genera *Thioalkalivibrio* and *Marinosprillum* (Table 5.3). In addition, the species of the genera *Desulfurivibrio* represented the *Deltaprotobacteria* in ABJ with 1.5% of the sequences.

The primers used for the 454 pyrosequencing in this work also amplify the 16S rRNA gene from the *Archaea.* In Lake Abjjata, the *Archaea* encompassed majority of sequences (6%) with phylum *Euryarchaeota* representing a fairly significant proportion (4.27%) of the total sequences followed by LSH (2.5%) and Chitu (0.33%) (Table 5.2). The *Crenarchaeota* was the second largely detected phylum in Lake ABJ with 0.83% of all sequences containing a large proportion of microbes in the Class *Thermoprotei.* Lake Chitu contained relatively higher sequences of the species of the genus *Methanocalculus* (0.1%) in the phylum *Euryarchaeota.*

Table 5.3: Major Mcrobial Taxa Represented in Chitu, Shalla and Abijata derived from 16S rRNA genes pyrosequencing profile. The symbol, * stands for ―very rare or no ovbvious detection under this study”

Major Microbial Taxa	**Main Representatives of Class /Order/Family/Genera**	**Average of Total Pyro-sequences (%)**		
		LCH	**LSH**	**ABJ**
Archaea		**0.50**	**2.07**	**6.13**
	Euryarchaeota; other	0.33	2.50	**4.27**
	Euryarchaeota;Methanomicrobia; Methanomicrobiales ***Methanocalculus***	0.10	*	*
	*Crenarchaeota;**Thermoprotei***	0.10	0.10	**0.83**
Acidobacteria		**0.77**	**1.37**	**0.67**
	Acidobacteria_Gp1; ***Gp1***	0.3	0.57	0.23
	Acidobacteria_Gp2; ***Gp2***	0.2	0.23	0.17
	Acidobacteria_Gp3; ***Gp3***	0.1	0.27	0.17
	Acidobacteria_Gp4; ***Gp4***	0.1	0.1	*
Actinobacteria		**0.70**	**1.50**	**1.10**
	*Nitriliruptorales;Nitriliruptoraceae;**Nitriliruptor***	0.1	0.5	0.1
	*Actinomycetales;Microbacteriaceae;**Microbacterium***	0.2	0.1	*
	*Actinomycetales;Mycobacteriaceae;**Mycobacterium***	0.1	0.1	*
Bacteroidetes		**0.70**	**2.87**	**4.27**
	Bacteroidetes;Other;Other;Other;Other	0.33	1.9	**2.9**
	flavobacteria**;Flavobacteriales;**Cryomorphaceae	*	*	0.13
	Sphingobacteria;Sphingobacteriales;Chitinophagaceae;Gracilimonas	0.10	0.20	0.67
	*Sphingobacteria;Sphingobacteriales;**Chitinophagaceae**;Other*	0.10	0.30	0.33

	*Sphingobacteria;Sphingobacteriales;**Cyclobacteriaceae**;Other*	*	0.20	0.10
Chloroflexi		**0.37**	**0.60**	**4.20**
	*Anaerolineae;Anaerolineales;**Anaerolineaceae**;Other*	0.13	0.57	1.93
	*Dehalococcoidetes;**Dehalogenimonas**;Other*	0.3	0.1	1.63
	Thermomicrobia;Sphaerobacterales;Sphaerobacteraceae; Sphaerobacter	*	*	0.43
Cyanobacteria		**0.20**	**0.10**	**9.13**
	Cyanobacteria;Chloroplast;Streptophyta;Other	0.17	0.03	0.03
	*Cyanobacteria;Family II;**GpIIa**;Other*	*	*	**5.57**
	Cyanobacteria;Family IV;GpIV;Other	*	*	0.87
	Cyanobacteria;Family IX;GpIX;Other	*	*	0.80
	Cyanobacteria;Other;Other;Other	*	*	1.80
Deinococcus-Thermus		**1.37**	**1.63**	**1.13**
	Deinococci;Deinococcales;Trueperaceae;Truepera	0.13	0.83	1.10
	Deinococci;Thermales;Thermaceae;Meiothermus	1.23	0.80	*
Firmicutes		**19.80**	**11.43**	**6.90**
	***Bacilli**;Bacillales;Bacillaceae;Anoxybacillus*	**18.89**	**8.90**	0.10
	*Bacilli;Bacillales;Bacillaceae;**Halolactibacillus***	*	*	0.1
	Bacilli;Bacillales;Bacillaceae;Bacillus	*	0.10	0.20
	Clostridia;Clostridiales;Clostridiaceae;Alkaliphilus	*	*	0.23
	Clostridia;Clostridiales;Clostridiaceae;Tindallia	*	0.20	0.10
	Clostridia;Natranaerobiales;Natranaerobiaceae;Dethiobacter	*	0.10	0.27
	*Clostridia;Thermoanaerobacterales;Thermoanaerobacteraceae;**Gelria***	*	*	0.10
	***Clostridia**;Other;Other;Other*	0.53	0.97	**3.90**
Protobacteria:		**68.73**	**68.57**	**19.69**
Beta-protobacteria		**48.13**	**47.10**	**1.37**
	Burkholderiales;Comamonadaceae;Delftia	**42.83**	**41.53**	0.77
	*Burkholderiales;Comamonadaceae;**Pelomonas***	0.17	0.37	*
	Burkholderiales;Alcaligenaceae;Achromobacter	1.37	1.90	0.10
	Burkholderiales;Burkholderiaceae;Ralstonia	1.13	1.77	*
	*Burkholderiales;Burkholderiales_incertae_sedis;**Tepidimonas***	0.10	0.10	*
	*Burkholderiales;Burkholderiales_incertae_sedis;**Aquabacterium***	0.20	0.20	*
	Burkholderiales;Oxalobacteraceae;Massilia	1.70	0.67	*
Gamma-protobacteria		**11.03**	**6.83**	**11.9**
	Pseudomonadales;Moraxellaceae;Acinetobacter	**8.57**	**2.70**	0.10
	Chromatiales;Ectothiorhodospiraceae;Thioalkalivibrio	0.13	0.03	**3.17**
	Chromatiales;Ectothiorhodospiraceae ; Thioalkalispira	*	*	**0.13**
	Chromatiales;Ectothiorhodospiraceae;Halorhodospira	0.1	*	**0.3**
	Oceanospirillales;Oceanospirillaceae;Marinospirillum	*	*	**1.03**
	Pseudomonadales;Pseudomonadaceae;Pseudomonas	0.17	0.30	*
	*anthomonadales;Xanthomonadaceae;**Stenotrophomonas***	0.10	0.10	*
	Xanthomonadales;Xanthomonadaceae;Other	**1.43**	**1.10**	0.07
	*Enterobacteriales;Enterobacteriaceae;**Escherichia/Shigella***	0.10	0.10	*
	*Methylococcales;Methylococcaceae;**Methylomicrobium***	0.20	0.10	*
Alpha-protobacteria		**9.30**	**13.63**	**3.30**
	Caulobacterales;Caulobacteraceae;Brevundimonas	**5.67**	**6.93**	0.37
	Caulobacterales;Caulobacteraceae;Phenylobacterium	0.20	0.17	*
	*Caulobacterales;Hyphomonadaceae;**Oceanicaulis***	*	*	0.10
	*Rhizobiales;Phyllobacteriaceae;**Phyllobacterium***	*	*	0.13
	*Rhizobiales;**Bradyrhizobiaceae***	0.23	0.17	0.07
	*Rhodobacterales;Rhodobacteraceae;**Rhodobaca***	0.10	0.27	0.20
	*Rhodobacterales;Rhodobacteraceae;**Paracoccus***	*	0.10	0.10
	***Sphingomonadales**;Other;Othe*	1.90	**4.20**	0.17
	Sphingomonadales;Sphingomonadaceae;Sphingomonas	0.30	0.20	*
Delta-protobacteria		**0.23**	**0.80**	**2.80**

	Desulfobacterales;Desulfobacteraceae;Desulfobacterium	*	0.10	0.10
	*Desulfobacterales;Desulfobacteraceae;**Desulfosarcina***	*	0.10	0.10
	Desulfobacterales;Desulfobulbaceae;Desulfurivibrio	*	0.10	**1.50**
	Desulfovibrionales;Desulfohalobiaceae;Desulfonatronovibrio	*	*	0.13
	*Desulfovibrionales;Desulfohalobiaceae;**Desulfonatronospira***	*	*	0.10
	Desulfovibrionales;Desulfonatronaceae;Desulfonatronum	0.10	0.10	**0.37**
	Desulfuromonadales;Desulfuromonadaceae;Other	0.10	0.10	0.10
	Myxococcales;Other;Other	*	*	0.10
Epsilonproteobacteria	*Campylobacterales;Helicobacteraceae;**Sulfurimonas***	*	0.10	*
Other Bacteria		**6.47**	**7.7**	**43.53**
Nitrospira	*Nitrospira;Nitrospirales;Nitrospiraceae;**Nitrospira***	*	0.10	0.10
Planctomycetes	*Planctomycetacia;Planctomycetales;Planctomycetaceae*	*	0.1	0.1
Spirochaetes	*Spirochaetales;Spirochaetaceae;**Spirochaeta***	0.10	0.10	**0.30**
Thermotogae	*Thermotogales;Thermotogaceae;**Kosmotoga***	0.17	0.10	0.47
Verrucomicrobia	*Opitutae;Puniceicoccales;Puniceicoccaceae;**Coraliomargarita***	*	*	0.10
	*Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae;**Luteolibacter***	*	0.10	*
Tenercutes	***Tenercutes***	*	0.10	0.10
BRC1	***BRC1***	*	*	0.10
OD1	***OD1***	*	0.10	0.10
TM7	***TM7***	*	0.10	0.17
WS3	***WS3***	*	0.23	0.37
Unclassified		**0.03**	**0.93**	**0.60**

The prokaryotic community richness indices (non-parametric Chao1) and diversity indices (Shannon estimator) were analyzed for the three lakes. Chao1 analysis (Fig. 5.6) shows that the microbial richness was highest in Abijata lake and the lowest in Chitu lake. The differences between the lakes microbial richness is statistically significant. Shannon analysis (Fig. 5.7) shows that the microbial diversity was highest in Abijata lake followed by Shalla lake and the lowest in Chitu lake. The differences between the lakes microbial diversity is statistically significant. The PCoA bi-plot analysis of the three lakes (Abijata, Chitu, Shalla) sediments gave three clusters based on OTUs detected at 97% similarity in the three replicates obtained from the three lakes samples, where the replicates from each lake were tightly clustered together (Fig. 5.8). PC1 explains 96% of the variation in the microbial community structure, whereas, PC2 explains only 2% of the variation.

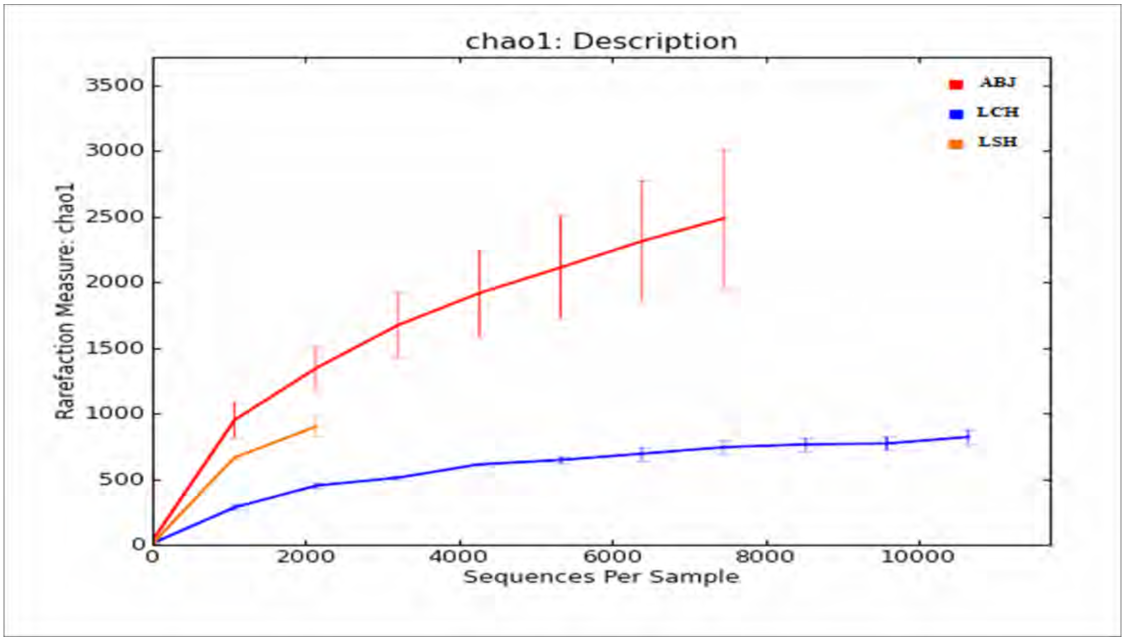


Figure 5.5: Chao 1 Analysis of LCH, LSH and ABJ based on 16S rRNA gene Pyrosequences. OTUs detected at 97% similarity.

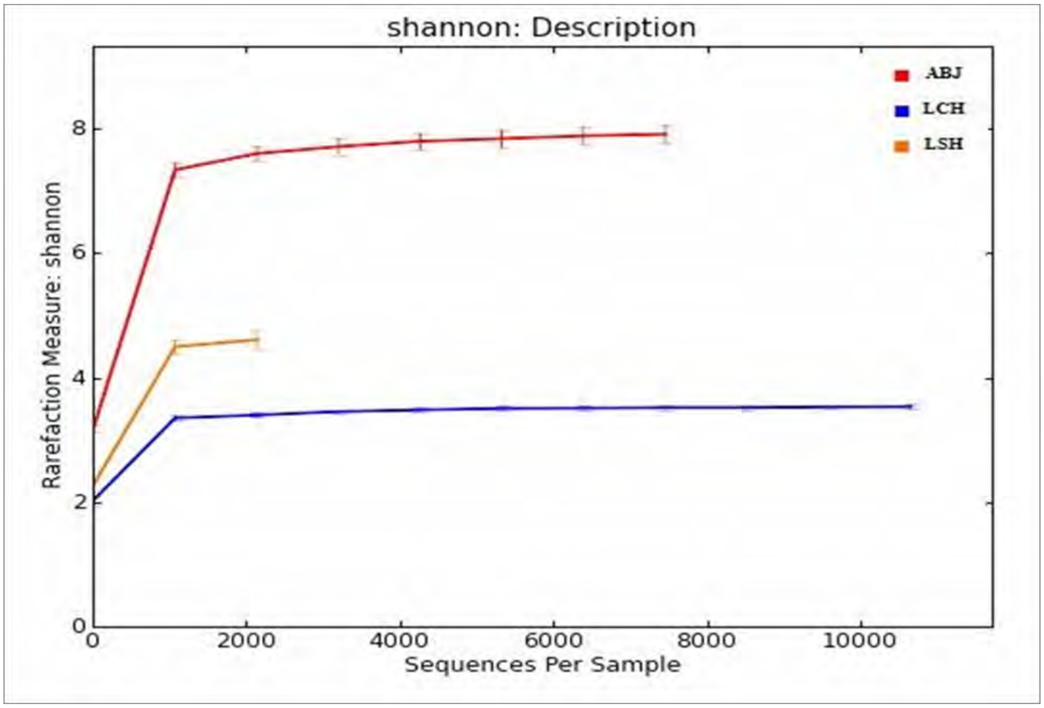


Figure 5.6: Shannon analysis of LCH, LSH and ABJ based on 16S rRNA gene Pyrosequences OTUs detected at 97% similarity.

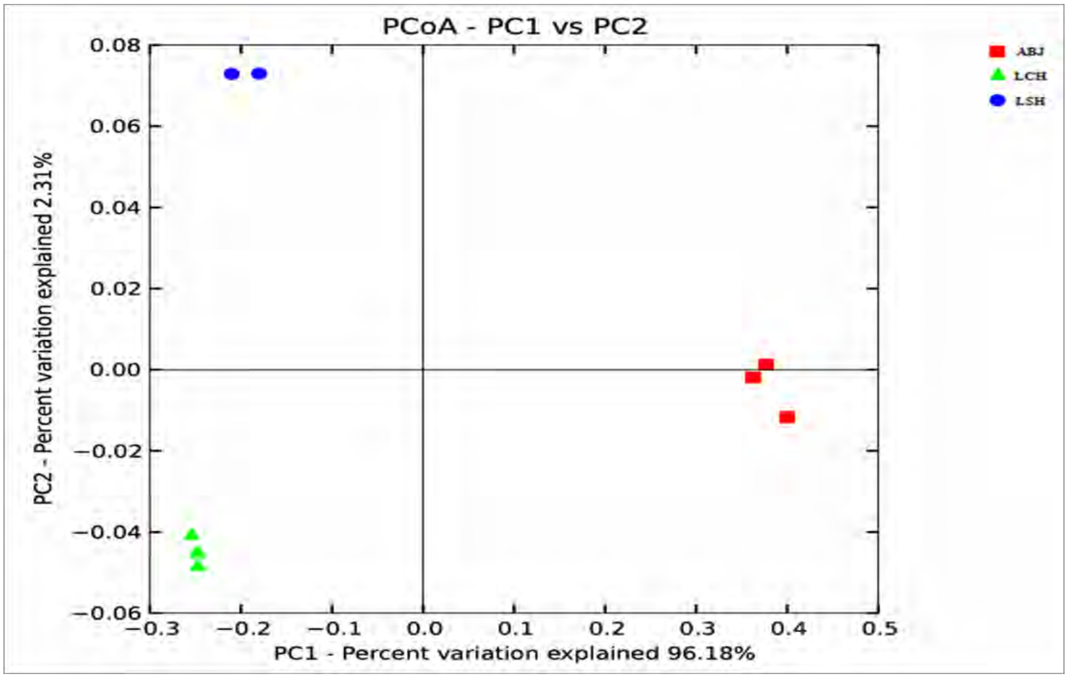


Figure 5.7: PCoA analysis of LCH, LSH and ABJ based on 16S rRNA gene pyroseqquences. OTUs detected at 97% similarity

5.3 Discussion

In this study pyrosequencing of barcode-tagged 16S rRNA gene amplicons was used to decipher the microbiome of the three Soda Lakes (Chitu, Shalla and Abijata) located in the Rift Valley of Ethiopia. From metagenomic data of all the sampling lakes indicated that bacterial sequencing reads were represented by more than 90% compared to that of 10% archaea, suggesting that bacteria were absolutely dominant in the prokaryotic communities in all the three soda lakes. Although microbial cell densities might be much lower to recover DNA from sediment samples, the results of their 16S rRNA pyrosequence data remarkably revealed high microbial diversity. Predominant pyrosequence types of prokaryitic communities were identified from the three lakes along with their potential ecological implications.

The metagenomic data of the three lakes contained a high proportion of sequences related to *Protobacteria* (Table 5.3). Notably, *Betaproteobacteria* within the metabolically versatile order *Burkholderiales* (including the families: *Burkholderiaceae, Comamonadaceae*, *Acaligenaceae* and *Oxalobacteraceae*) appeared to harbor the largest groups in Chitu and Shalla lakes (Table 5.3). The predominance of *Proteobateria* (essentially orders *Burkholderiales*) may be due to high amount of organic matter in deeper layers of soda lake sediments sampled, which also increases the production and consumption rates of methane and organic matter degradation by genera belonging to α-*Proteobacteria* and that use a great diversity of C1 substances (Amann *et al*., 2000).

These groups include strictly aerobic and facultative anaerobic chemoorganotrophs, obligate and facultative chemolithotrophs, nitrogen-fixing organisms, as well as plant, animal and human pathogens (Garrity *et al*., 2005); therefore constituting metabolic and ecologically diverse bacterial lineage. A variety of bacteria belonging to this order have extensive biodegradative abilities (including Chlorinated and aromatic compounds as well as antibiotics) and they belong to the second most dominant of the genome database following *Enterobacteriales* (Pérez-Pantoja *et al*., 2012).

Euryarchaeota and *Crenarchaeota* sequences were significantly represented in the three lakes studied. Interestingly, the species of the genera *Methanocalculus* in the phylum *Euryarchaeota* were found to relatively dominate the Chitu lake sediments, which showed the existence of potential methanogenic *Archaea* or methanogens in this lake. Earlier culture-based studies performed by Kevbrin *et al* (1998) and direct molecular analysis by Sukarasi *et al* (2007) revealed the isolate of *Methanocalculus sp.* from Lake Magadi in Kenya and Lonar lake in India, respectively. Comparably, Lanzen *et al* (2013) detected abundant species of Methanogens, *Methanocalculus sp., Methanolobus* and *Methanoseata* from Arenguadi (at 30 m) and Chitu lakes.

The haloalkaliphilic methanogens isolated so far from soda lakes were considered to be mainly obligately methylotrophic that produce methane through the utilization of complex compounds such as methanol and methylamine (Oren, 1999). Current studies also describe that they may also produce methane by using acetic acid, H_2, and CO_2 (revised in Grant and Sorokin, 2011).

Interestingly, the crenarchaeal hyperthermophiles that belong to the class *Thermoprotei* were also detected in the three lakes studied, relatively predominating in Lake Abijta (Table 5.3). Chrenarchaea are best known as hyperthermophiles that dwell in the hottest environments on Earth, including volcanic hot springs and thermal vents (Stetter, 2002).

In this study, the results obtained from 16S rRNA pyrosequence also demonstrated the occurrence of *Methylomicrobium sp* in Chitu and Shalla lakes, indicating the existence and significance of aerobic methane-oxidizing bacteria in the sediments of both lakes studied, which was not previously reported by Lanzen *et al* (2013). Molecular probing of the methanotrophic population in soda lakes of Central Asia and enrichment culture from Rift Valley Soda Lake in Kenyan as well as study from Lunar creter lake in Egypt demonstrated ubiquitous domination of type I methanotrophy belonging to the genus *Methylomicrobium* in the *Gammaproteobacteria* (Li *et al*., 2004; Kalyuzhnaya *et al*., 2008; Surakasi *et al*., 2010).

Predominat number and diverse groups of cyanobacteria were observed in Lake Abijata (approximately 9% of the sequences) with the majority of the family II, clad *GpIIa* as well as a few others including family IV (clad *GpIV*) and family IX (clad *GpIX*) (Table 5.3), indicating the potential of photosynthetic primary productivity of this lake. It was described that the cyanobacteria are one of the most diverse prokaryotes in various habitats and survivers of extreme environments, including haloalkaline lakes (Jones *et al*., 1998). They play vital roles (particularly *Spirulina spp*.) as fixers of CO_2 and N_2 and producers of O_2 (Jones and Grant, 1999). Previous studies (Jones *et al*., 1998) have also shown that photosynthetic primary productivity in East African soda lakes are attributable mainly to *Arthrospira spp*., as well as seasonal blooms of other cyanobacterial species related to *Cyanospira*, *Synechococcus* and *Chroococcus*.

In addition to the cyanobacteria, anoxygenic phototrophic purple sulfur and nonsulfur bacteria were also readly detected in the lakes under this study. The photosynthetic purple bacteria belonging to *Halorhodospira* within the *Gammaprotobacteria* were detected in Lake Abijata and Chitu sediments (Table 5.2). The hypersaline soda lake Magadi in Kenya and Wadi Natrun in Egypt were also fond to harbor the species of the genus *Ectothiorhodospira* and *Halorhodospira* contributing to similar ecological roles (Grant and Tindall, 1986; Grant 1992; Imhoff *et al*., 1979; Kompantseva *et al*. 2005).

Lanzen *et al* (2013) conducted RNA transcript analysis of the anoxic Chitu and Abijata lakes water bodies and found to harbor *Ectothiorhodospira* species. The absence or presence of *Halorhodospira* or *Ectothiorhodospira* species in Lake Chitu and Abijata in this study and that of Lanzen *et al* (2013) might be due to the biases that could arise from type of sample used, sampling, and/or downstream nucleic acid processing procedures. The position of the genus *Ectothiorhodospira* in the re-cycling of sulphur is also imperative. The non-sulfur purple bacterial genera, *Rhodobaca* from the family *Rhodobaceraceae* were another important anoxygenic photosynthetic bacterial species detected in the three lakes studied, Chitu, Shalla and Abijata (Table 5.3). The study of Lanzen *et al* (2013) also recorded the existence of *Rhodobaca,* including three other genera (*Rhodobacter*, *Pseudorhodobacter* and *Roseibacter*) from Abijta and Shalla lakes. *Rhodobaca* sp. was also isolated from Soda Lake of the Barguzin Valley in Eastern Siberia (Boldareva *et al*. 2008). However, in this study, only *Rhodobaca* was detected that may be due to the preference of the other genera to the surface water of the lakes.

In this study, the genus of *Thioalkalivibrio sp.* in the family *Ectothiorhodospiraceae* were abundantly detected in Lake Abijata and also a relatively few reads represented in Chitu and Shalla lakes. Besides, pyrosequences of 16S rRNA genes relevant to the genus *Thioalkalispira* were also detected in Lake Abijata. Four culturable forms of Sulfur oxidizing bacteria (SOB) that include *Thioalkalispira, Thioalkalibacter*, *Thioalkalivibrio*, *Thioalkalimicrobium* were also recently described in haloalkaliphilic soda lakes (Grant and Sorokin, 2011). In this study, the detection of SOB types were abundantly detected in the Na_2CO_3 /$NaHCO_3$- soda rich Lake Abijata (Table 5.3) that could explain SOB types prefer a sodium carbonate/bicarbonate-based (soda–dominated) environment than NaCl-dominated brines (Sorokin *et al.,* 2011).

Notably, the sulfur-oxidizing *Epsilonproteobacteria*, especially those of the widespread members of the genus *Sulfurimonas*, were detected in Lake Shalla (Table 5.3). Epsilonproteobacteria are chemolithoautotrophs or mixotrophs and they have been identified from vast array of sulfidogenic environments including deep sea hydrothermal vents (Campbell, 2006). The plentiful hydrothermal springs of different teperatures feeding Lake Shalla may probabily be one of the key sources for relative occurrence of *Sulfurimonas* in the lake.

In the soda lakes currently studied, several SOB in the *Deltaprotobacteria* that belong to two major Orders embracing three genera affliated to the Order *Desulfobacterales* (*Desulfobacterium, Desulfosarcina, and Desulfurivibrio*) and other three genera allied to the Order *Desulfovibrionales* (*Desulfonatronovibrio, Desulfonatronospira, and Desulfonatronum*) were detected (Table 5.3). Four of the genera detected in the lakes are common to other genera found in other soda lakes elsewhere except the two species of the genera *Desulfobacterium* and *Desulfosarcina* which were not reported normally from other soda lakes. The species of the genus *Desulfonatronum* was common to all the three lakes studied and yet it was the only SOB genus within the *Deltaprotobacteria* readily represented in Lake Chitu compared to Lake Shalla (see Table 5.3). Recently, a new haloalkaliphilic SRB bacterium, *Dethiobacter alkaliphilus*, was identified as a member of the class *Clostridia* within the *Firmicutes* to be able to grow by disproportionation of elemental sulfur as one of the sulfate-reducing bacteria; SRBs (Poster, *et al*., 2013).

Recent microbiological and molecular analysis has also ellucidated a domination of important groups of haloalkaliphilic sulfate-reducing bacteria (SRB) belonging to the orders *Desulfovibrionales* and *Desulfobacterales* (Foti *et al*., 2007) in Siberian, Egyptian, American and Kenyan soda lakes. Several lithoautotrophic SBR belong to the species of the genera *Desulfononatronospira* (Sorokin *et al*., 2008a), *Desulfonatronum* (Pikuta *et al*., 1998; Zhilina *et al*., 2005), *Desulfonatronovibrio* (Zhilina *et al*. 1997), culture isolates of *Dethiobacter* (Gram positive, *Clostridia*) and *Desulfurivibrio spp*. (Sorokin *et al*, 2008b), and several heterotrophic SRB allied to *Desulfobotulus alkaliphilus*, *Desulfobacteraceae*, *‗Desulfobulbus alkaliphilus*‘ and *Synthrophobacteraceae* have been isolated from soda lakes in Kulunda Steppe (Sorokin *et al*., 2010b).

The heterotrophic, Gram-positive and high/low G + C bacteria within the phyla of *Firmicutes* and *Actinobacteria* were also well represented in the soda lakes under this study (Table 5.3). Remarkably, the alkali-tolerant thermophilic species of the genus *Anoxybacillus* among the Firmicutes was highly represented in Lake Chitu and also in Lake Shalla (Table 5.3). This bacterium was recently proposed as a new genus of the family *Bacillaceae*; the name coined based on its anaerobic property (Pikuta *et al*., 2009). So far, a total of 19 species have been reported from this new genus, and several hydrolytic enzymes, mostly related to the metabolism of carbohydrates (such as starch and lignocellulosic biomasses) have been characterized from this genus (Goh *et al*., 2013). Compared to other *Bacilli* and *Clostridia* species, *Anoxybacilli* are seemingly more abundant relatively in Lake Chitu.

A novel alpha-amylase enzyme producing gene from alkalitolerant moderate thermophilic *anoxybacillus sp* DT3-1 was recently described for the first time (Goh *et al*., 2013). The enzymatic screening of fosmid clones from Lake Chitu has also evidenced that Chitu may harbor such specialized species of microbes that have an arsenal of potential extracellular alkalophilic hydrolytic enzymes (see Chapter six).

The sequences of a diverse *Clostrida* species, low G+C Gram-positive bacteria that belong to the genus *Alkaliphilus*, *Tindallia*, *Dethiobacter*, *Gelria, and* others unknown were encoutered in Lake Abijata and a few of the group in Lake Shalla (Table 5.3). One of the Orders among many fermentative and anaerobic alkalophilic respires from Soda lakes belongs to *Clostridiales* (Sorokin *et al*., 2012). Several other alkaliphilic soda lake acetogenic bacteria were identified *Tindallia* from Lake Texcoco and others within the *Clostridium* spectrum from Lake Magadi in Kenya, Lake Owens in the USA, and Verkhnee soda lake in Russia (Sorokin and Grant, 2011).

The species of the genus *Tindallia* previously identified from Lake Magadi, Kenya was described as ammonifier bacteria which ferment the amino acids arginine and ornithine, producing acetate, propionine, and ammonia (Kevbrin, *et al*., 1998). *Dethiobacter alkaliphilus*, was newly identified member of the *Clostridia* from a soda lake and it is also capable of growing by disproportionation of elemental sulfur as one of the sulfate-reducing bacteria; SRBs (Poster, *et al*., 2013).

The pyrosequences data of this study showed that the chemolithoautotrophic nitrite-oxidizing bacteria (NOB), *Nitrospira* were presented in Shalla and Abijata Lakes (Table 5.3). In addition, the occurrences of the pyroreads for *Paracoccus* within *Rhodobacteraceae* and the *Pseudomonas* were evidenced in the soda lakes studied (Table 5.3). Several corresponding nitrifiers and denitrifiers have been described from Soda lakes in African Rift Vally and other habitates (Antony *et al*., 2013).

The high-G+C-content gram-positive bacteria, *Actinobacteria* that comprise the species of the genus *Nitriliruptor*, *Mycobacterium*, and *Microbacterium* were occurred in Lake Chitu and Shalla, fairely dominating in Lake Shalla (Table 5.3). This phylum has unique adaptations to the extreme soda lake environments and presents a great biochemical and morphological diversity, as well as different forms of metabolic capability, generally correlated with synthesis of novel types of bioactive compounds/secondary metabolites (Hayakawa *et al*., 2000; Zhao *et al*., 2008).

Close to 4.3% of the total sequence reads of prokaryotes was represented by *Bacteroidetes* in Lake Abijata (Table 5.3). The sequences were similar to the order *Sphingobacteria*, strict anaerobic bacteria usually associated with the mineralization of complex polymers, and are frequently found where there is deposition of organic matter (Schauer *et al*., 2006). The phylum *Deinococcus-Thermus* that have been previously detected from geothermal springs (Brock *et al*. 1972) were also represented in the three lakes analysed. The cellular components *Deinococcus-Thermus* are of biotechnological interest with potential applications in bioremediation or molecular biology (e.g sources of thermostabe enzymes).

About 0.1% of sequence reads of *Planctomycetes* were recovered in Abijata and Shalla Lakes (Table 5.3) indicating the presence of bacterial taxon that are well-known for anaerobic ammonia oxidation (anammox) (Kirkpatrick *et al*., 2006). *Spirochetes* were also well represented in the three lakes studied. Alkaliphilic anearobic *Spirochaeta* species were previously isolated from Lake Magadi (Kenya) and Lake Khatyn (Central Asia) (Zhilina *et al*., 1996) and have been characterized to utilize a wide variety of carbohydrates anearobically producing acetate, lactate, ethanol and H_2 (Jones and Grant, 1999). Finally, the pyrosequence data revealed that unclassified archaea and bacteria were found in all the three lakes (Table 5.3), indicating that a great part of the archaeal and bacterial diversity is not known in thsese soda lakes.

5.4 Conclusion

This study is the first metagenomic analysis using Barcode- tagged pyrosequencing to characterize a more comprehensive overview of the prokaryotic community of Chitu, Shalla, and Abijata soda lakes located within the Rift Valley of Ethiopia. The 454 pyrosequencing of the barcoded amplicons of the 16S rDNA V3–V4 hypervariable regions revealed much wider diversity of the soda lakes compared to the 16Sr RNA clone library sequences employed in chapter 4 of this study. The 454 pyrosequencing in the three soda lakes detected diverse major microbial communities associated with primary producers (Mostly cyanobacteria and anoxygenic phototrophs), aerobic heterotrophs, fermentative and respiratory anaerobes (sulfate-reducing bacteria, SRB), acetogens, and methanogens. Accordingly, a high proportion of sequences related to *Protobacteria* notably, *Betaproteobacteria* within the metabolically versatile order *Burkholderiales* were found in the three lakes. Remarkably, the alkali-tolerant thermophilic species of the genus *Anoxybacillus* among the Firmicutes was highly represented in Lake Chitu and also in Lake Shalla. More interestingly, the existence *Sulfurimonas* in Lake Shalla adds to the knowledge about soda lake microbial diversity and it also designates the substantial contribution to the oxidative portion and function of these bacteria in soda lake sulfur cycle.

CHAPTER SIX

Large-insert Fosmid Metagenomic DNA Library Construction and Analysis of Enzymatic Potential of Chitu Soda Lake

6.1 Introduction

The 16S rRNA-based traditional and/or highthrouput methods can‘t characterize the functional component of environmental microorganisms. In metagenomics, the direct cloning of environmental DNA, allows the identification and isolation of the vast and yet untouched genetic pool, novel products and pathways of microbial communities living in most diverse natural habitats, including, environments of extremes of pH, salinity, pressure and temperatures (Streit and Schmitz, 2004; Riesenfeld *et al*., 2004).

Enzymes are significantly needed for industrial, pharmaceutical and biotechnological processes (Sanchez and Demain, 2011). The total market for industrial enzymes reached $3.3 billion in 2010 and it is anticipated to escalate to a value of 4.4 billion by 2015 (BBC Research Report, 2011). Metagenome-derived enzymes have the potential to be used for various biotechnological applications (Steele *et al*., 2009). Recently, microorganisms from soda lakes have attracted attention as a possible source of novel enzymes and metabolites for use in biotechnology for various applications such as medicine, food, and reagents (Grant and Heaphy, 2010).

Enzymes derived from extremophilic microorganisms are endowed with unique functional properties and catalytic power to be used in several commercial interests and novel applications of industrial processes (reviwed in Adrio and Demain, 2014). These enzymes are stable at varied ranges of harsh reaction conditions of pH, temperatures, salinity etc. (Kumar *et al*., 2011). Screening of new source of novel and industrially useful enzymes is yet a foremost research focus of many researchers. Microbes that possess adaptations to high salinity and alkaline pH of soda lakes including Rift Valley soda lakes in East Africa, have also attracted a lot of attention as potential sources of enzymes with novel activities and applications (Grant *et al*., 1990). Particularly, alkaphilic polmer-degrading enzymes such as proteases, lipases and cellulases are most frequently isolated from *Bacillus* or related species from soda lakes, from which two different cellulases were marketed for use in laundary and textile processes (Grant and Heaphy, 2010).

With the use of metagenomics that is helpful in constructing a genomic library from environmental DNA, it would be possible to reveal novel genes or open reading frames that potentially encode putative enzymes from soda lake microbes for applications (Gilbert and Dupont, 2011; Steele *et al.*, 2009). In this procedure large-insert metagenomic DNA is commonly cloned into commercial vectors (e.g. Fosmid, pCC1FOS, from Epicentre) packed into λ phage heads and transfer to E. *coli* host cells by infection (Park *et al.*, 2008). Next, functional enzyme screening from from the clone libraries would be made using various methods, among which phenotypic detection is possible using plate assays on growth agar containing substrates and indicator substances would be employed.

To this end, extensive study on microbial diversity and functions of microbes from haloalkaline lakes in the Rift Valley was not undertaken. Therefore, the study of metagenome-derived enzymatic potential of haloalkalophiles from soda lakes of Ethiopia is very vital. The present study used metagenomic approaches to (i) construct a large- insert fosmid (pCC1FOS vector) metagenomic DNA library of Chitu soda lake (ii) identify clone expressing enzyme activities (especially amylases, proteases, cellulases and lipases/estrases) from the fosmid clone libraries using functional /activity-based metagenomic screening techniques.

6.2 Result

6.2.1 Cloning of Metagenomic DNA

Intact high molecular weight metagenomic DNA was extracted from LCH sediment samples by employing the modified Zhou protocol as described in section 3.3. This DNA extraction technique was preferred because it yielded very little or no shearing of DNA compared to other various trial techniques performed.

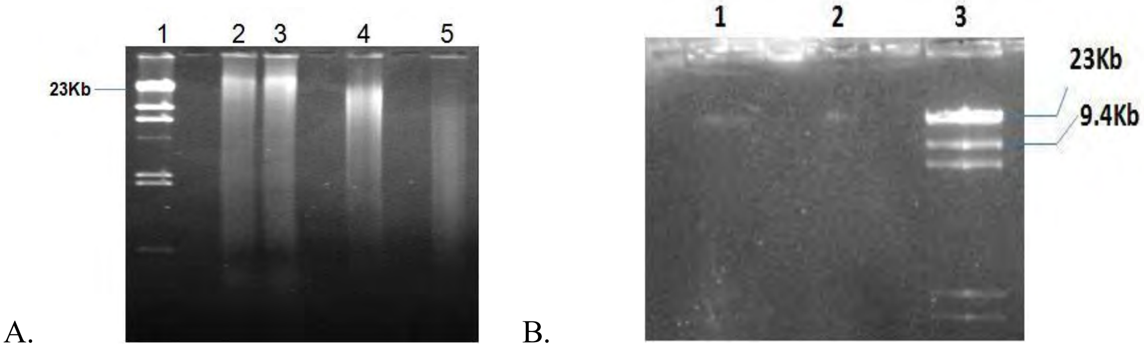


Figure 6.1: Metagenomic Crude DNA from Lake.
A: The Zhou protocol for Metagenomic crude DNA retrieved from Lake Sediment Samples. Lanes 1: Lambda DNA *HindIII* digested molecular weight marker. Lanes 2 & 3: Lake Chitu metagenomic DNA. Lanes 4: Lake Shalla metagenomic DNA. Lane 5: Lake Abijata metagenomic DNA.

B: Gel Purified Metagenomic DNA retrieved from Lake Chitu. Lanes 1 & 2: Lake Chitu metagenomic DNA. Lanes 3: Lambda DNA *HindIII* digested molecular weight marker.

The Zhou DNA retrieval method could not yield high-molecular intact DNA for LSH and ABJ sediments samples suitable for cloning into the Fosmid vector as part of this study (Fig. 6.1A, Lane 4 & 5). This prompted the use of Zhou-extracted DNA of LCH sediment for Fosmid library construction in this study. The crude LCH metagenomic DNA was purified from the most difficult inhibitory substances and huge amount of humic acids contaminates by using gel electrophoresis with low melting point agarose (LMA). Fig. 6.1 B shows the preparation of size-selected, gel purified, high- molecular and intact LCH metagenomic DNA for Fosmid library construction in this work.

6.2.2 Library Size and Coverage

A total of 10 µg of crude metagenomic DNA of LCH sediment samples was recovered after pooling different extractions and electrophoresed on LMA. About 3 µg of DNA was recovered after size selection and purification steps, and 500 ng of of this was finally recovered from end–repairing procedure. The metagenomic library roughly containing 250 ng of the end-repaired, intact and high molecular DNA form a single 10 µl ligation reaction was finally acquired in a Fosmid system (pCC1FOS) after transformation into phage resistant T1 *E. coli* EPI300 cells (Fig. 6.2). The fosmid library generated contained 7, 200 clones which were individually picked into 96-well plates (Fig 6.3).

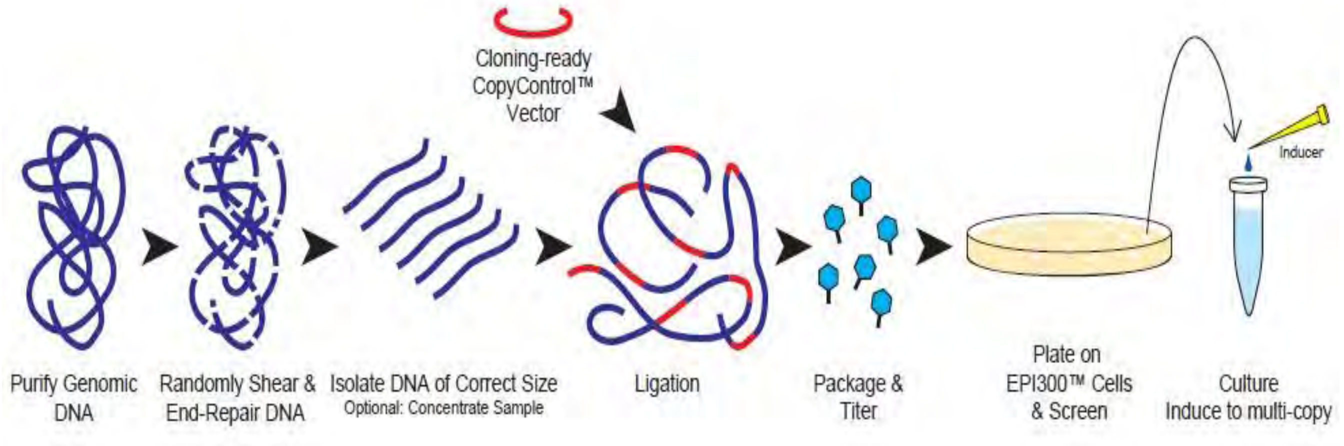


Figure 6.2: Schematic Diagram for Production of Fosmid Libraries. A CopyControl™ Fosmid library and subsequent induction of clones to high-copy number. (Adapted from Epicenter Product Protocol for Fosmid pCC1FOS system).

Figure 6.2: Photo showing the laborious hand Collection of 7, 200 Fosmid Clone library of LCH from Petri-dishes singly into 96-well plate format (Photo: MME lab, University of Copenhagen, Denmark)

To estimate the average insert size of the metagenomic library, restriction endonuclease enzyme digestion was performed with *NotI*. Accordingly, *Not I* digestion of fosmid DNA isolated from eight randomly chosen clones verified a 40-Kb insert size (Fig. 6.3). In the same figure, this could interestingly be confirmed that fosmid clone in lane 7 carried approximately a 40 Kb DNA insert size.

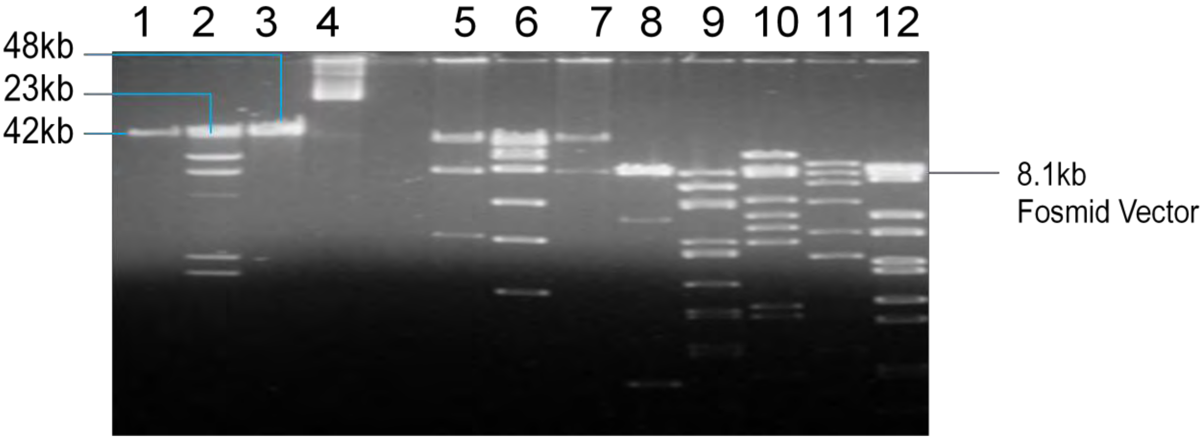


Figure 6.3: Analysis of Fosmid Clones with NotI Digestion: Lane 1 = Lambda DNA; Lane 2 = Lambda HindIII; Lane 3 = 42 Kb Fosmid Control insert DNA; Lane 4 = Undigested Fosmid DNA; Lane 5-12= *Not I*- digested Fosmid DNA

The total amount of genetic information (library size or covarage) covered in the 7, 200 Chitu Fosmid clone library was calculated to be 288 Mb of DNA, which is approximately equivalent to 62X, 68X, and 32X the complete genomes of *E. coli* (4.7 Mb), *Bacillus licheniformis* (4.2 Mb), and *Streptomyces coelicolor* (9.02 Mb), respectively.

6.2.3 Functional Primary Screening of Fosmid Clones

Primary screening for different enzymatic activities involving Cellulases, Proteases, Amaylases and Esterase/lipases was done by replicating the clone library stored in 96-well format with a 96-pin array onto agar indicator plates containing the appropriate assay-specific basic medium and Chloraphenicol antibiotic. Accordingly, activity-based primary screening of 7, 200 Chitu fosmid clone library showed 0 tributyrin-, 4 cellulose-, 22 milk- and 48 starch-hydrolyzing clones (Table 6.1), representing 0%, 0.06%, 0.31% and 0.67% of the total clones screened for each activity, respectively (Fig 6.4A).

Table 6.1: Activity-Basede Primary Screening of Lake Chitu Sediment DNA Fosmid Library Similarly, 0% of esterase/lipase, 0.06% cellulase, 0.31% protease and 0.67% amylase activities were detected per Mb of the total DNA coverage of the library, respectively (Fig 6.4B).

Function/ Activity	Library Type	Average insert size (Kb)	Number of Clones Screened	Library size (Mb)	Substrate	Number of hits	Hit rate (hit per Mb)
Amylase	**Fosmid**	**40**	**7, 200**	**288**	**Starch**	**48**	**1/6**
Protease	**Fosmid**	**40**	**7, 200**	**288**	**Skimmed milk**	**22**	**1/13**
Cellulase	**Fosmid**	**40**	**7, 200**	**288**	**Carboxymethyl-cellulose (CMC)**	**4**	**1/72**
Esterase /Lipase	**Fosmid**	**40**	**7, 200**	**288**	**Tributyrin**	**0**	**0**

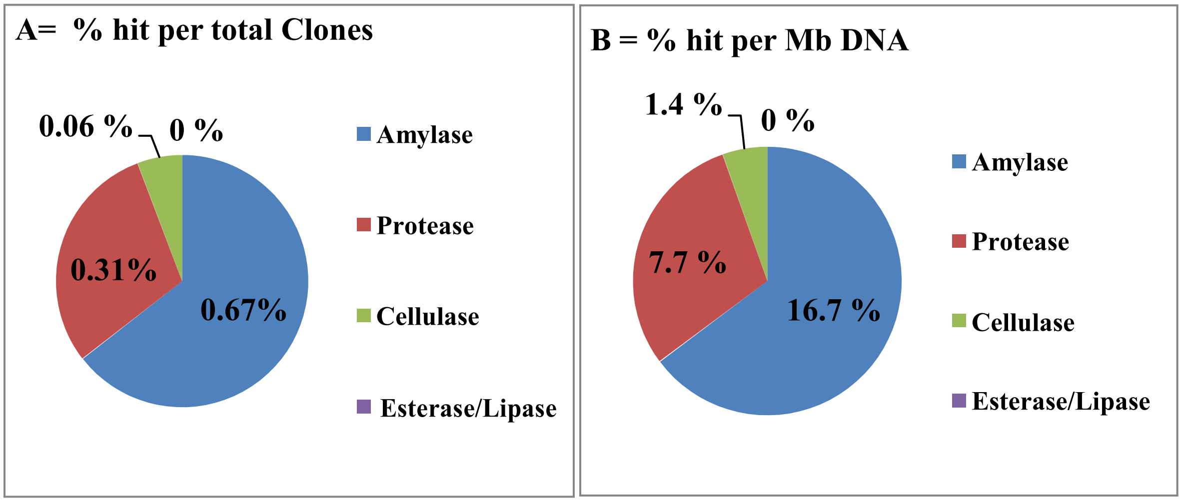
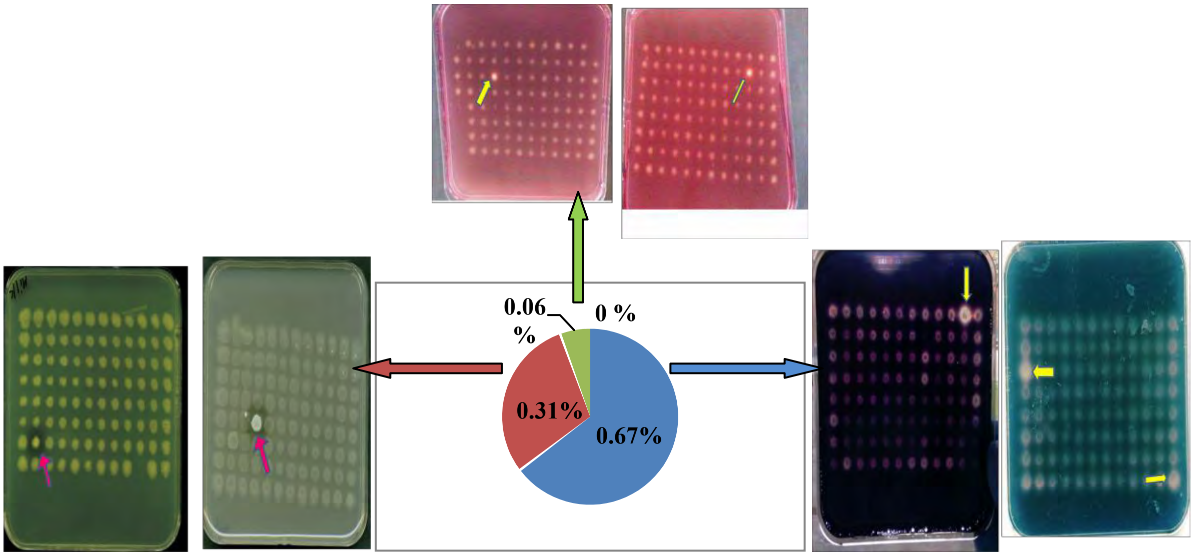


Figure 6.4: Hit rate of Enzyme Activity: Percentage hit rate per total clones (A) and Percentage hit rate per Mb DNA (B).



CELLULASE PLATES
4 positive clones
Substrate: 1% tributyrin
Detection: Hydrolysis halo

PROTEASE PLATES
22 positive clones
Substrate: 1% skimmed milk
Detection: Hydrolysis halo

AMYLASE PLATES
48 positive clones
Substrate: 1% starch + I_2/KI
Detection:Iodine decoloration

Figure 6.5: The Number and Percentage of Active Chitu Fosmid Clones: all detection plates contained activity specific substrates supplemented with 12.5µg/ml chloramphenicol. The plates shown here are representatives of other screening plates.

6.3 Discussion

One of the recent developments to uncover the genetic resurces of enzymatic significance of haloalkaliphilic microbes from extreme soda lake environments involves fosmid DNA library construction and activity screening (Kennedy *et al*., 2011). However, high molecular weight DNA preparation is one the important steps as well as the basis for constructing a high quality Fosmid library. In this study, the construction of large-insert metagenomic libraries was tough due to low biomass of the extreme habitates and large amount of co-extracted contaminants such as salts and humic acid compounds. In addition, during the optimization of extraction method (modified Zhou *et al*., 1996), it was difficult to obtain intact high molecular weight metagenomic

DNA from lake Shalla and Abijata sediments. Satisfactory amount and quality of DNA retrival was relatively possible from Lake Chitu sediments for Fosmid library construction, but with careful, time-consuming and cumbersome extraction and purification steps (see Methodology in Chapter 2).

In this study, fosmid library was constructed and functional metagenomic approach was applied so as to access enzymatic gene expression from the prokaryotic community of Lake Chitu sediments. The Fosmid vector (CopyControl pCC1FOS from Epicentre) used for this purpose was a large–insert cloning vector with the capability of holding upto 40Kb (Kim *et al*., 1992). Fosmid library is sensitive that do not take very short DNA fragments or size less than the fosmid carrying capacity up to 40 kb. During the library construction, shearing was applied to achieve a 40 kb length DNA fragment if the high molecular weight metagenomic DNA is too long or above the carrying capacity of the fosmid vector. The frequency of shearing treatments also should be checked; the less shearing treatment might generate long DNA fragments, while the more shearing action will produce undersized DNA fragments.

Analysis of both the insert size of the DNA fragments and the library coverage of the genome are crucial for characterizing the quality of the Fosmid library constructed. In this st, it was observed from one of the fosmid 8 clones analysed using *NotI* restriction enzyme, it could be possible to successfully place a 40kb size Chitu metagenomic DNA in the pCC1FOS vector (Fig. 6.3). The fosmid DNA vector has two *Not I* sites, either side of the multiple cloning site, and hence if the cloned fragment didn‘t contain *Not I* site then the digest would be expected to reveal two fragments (fosmid backbone and insert). One of the fosmid clones (lane 7 of Fig. 6.3,) was found to have the required large 40Kb DNA insert. The Metagenomic DNA insert had diverse *NotI* restriction sites.

In this study 7, 200 fosmid clnes were picked up and stored with an average insert size of 40Kb. The library coverage was calculated using the formula, N = ln (1-P) / ln (1-f). Assuming an average insert size of **40 kb**, the 7,200-clone library represents **288 Mb** of DNA, which was sufficient for approximately **a coverage of sixty two 4.7** Mb of genomes (*E-coli*). A one-fold coverage indicates that the chance of finding a particular genomic sequence in a single genome was approximately 99%. Accordingly, ***7, 200 clones = ln (1 – 0.99) / ln (1 – [X bases / 4.7 x 10^6 bases])***. So, 7, 200 clones are used to ensure a 99% probability of a given DNA sequence of a bacteria (genome = 4.7 Mb) being contained within the built Fosmid library composed of 40-kb inserts. The construction of large-insert Fosmid DNA library for Lake Chitu might be the first in the endeavor of metagenomic studies of the East African Rift Valley soad lakes.

In previous studies, although many were based on short- insert DNA samples, metagenome libraries have been similarly constructed for various environmental samples from soils and soda lakes using various vector systems (including Plasmid, Lambda, Cosmid and Fosmid) to screen important enzymes (Li *et al*., 2009; Rees *et al*., 2003).

This study demonstrated the construction of fosmid library of Chitu soda lake sediment large-insert DNA and expression screening of enzymatically active clones. Among the whole fosmid clone libraries generated, only four selected hydrolase enzyme producing fosmid clones were screened by plate assay on starch, skimmed milk, CMC, and tributyrin-containing agar plates for amylases, proteases, cellulases, and lipase/estrases respectively (table 6.1). Functional screening of Chitu Fosmid DNA library study has indicated the wide occurrence and secretion of extracellular enzymes, protease, amylase, and cellulase enzymes.

In this study, fosmid clones that showed clear zone or halos were identified as enzyme producing bacteria (Fig. 6.5). Amylase, Protease, and Cellulase activities from Lake Chitu Fosmid library was detected as a result of the degradation of the substrates suplimented onto the medium by extracellular enzymes secreted from the colony (Ekkers *et al.,* 2012).

In the current metagenomic study of Lake Chitu, 22 protease positive fosmid clones were detected out of a total of 7, 000 clones with a hit rate of 7.7 % of the total 288 Mb DNA library size (~ 22.2Mb DNA) (Fig 6.4). The number of hits obtained for Lake Chitu was high compared to previous reports made on functional plate screening of protease from other sources such as soils, Valley deserts, deep-sea sediments (Lee *et al*., 2007; Neveu *et al*., 2011; Waschkowitz *et* *al*., 2009) and pure culture studies (Pailin *et al*. 2001). Rondon *et al* (2000) reported 29 clones exhibited hemolytic activity out of a total of 25,000 clones from soil DNA libraries. In related study, a metagenomic study in which 100,000 plasmid clones were screened for proteolytic activity, only one was found positive, which was determined to be novel by sequencing analysis (Gupta *et al*., 2002). Most of the previous studies of the screening were made using skmmed milk. However, recent studies involved lactose free-milk or the pure protein, casein itself to be more robust and minimize false positivity (a background perplexity of the hydrolysis of milk sugar by glycoside hydrolases) for screening of extracellular microbial protease in metagenomic libraries and pure cultures (Jones *et al*., 2007; Morris *et al*., 2012).

Screening of 288 Mb of DNA of Chitu Fosmid libraries on starch agar paltes, 48 clones out of 7, 000 were positive for amylolytic activity (starch degradation) , which was 16.7% of the total Mb DNA of the library (Fig 6.4). Delavat *et al* (2012) screened roughly 80,000 plasmid clones from Acid Mine Drainages and found 28 positive clones (0.035%) for amylolytic activity on LB plates supplemented with 1% starch.

Functional screening of Chitu fosmid DNA library using tributyrin agar plate also showed 4 cellulase activities out of 7, 000 total fosmid clones (Table 6.1). Previously, short metagenomic DNA library screening from alkaline East African soda lake environment samples (Rees *et al*., 2003), and Egyptian soda lake enrichments (Grant *et al*., 2004) resulted in clones having cellulase or lipase gene activity were isolated with a frequency of between 1 in 15,000 and 1 in 100,000.

In this study of fosmid library screening, there was scarcity of cellulase active clones and a total lack for lipase/esterase expressing clones. Similarily, previous works described that metagenomic clones in enzyme activity screening often showed low expression. For example, in the libraries derived from soil DNA, only 1 in 730,000 clones showed activity for lipases (Henne *et al*., 2000). Screening of about 37, 000 to 130,000 Lambda clone libraries constructed from soda lake sediment and water samples has conferred only one cellulase enzyme- expressing clone (Rees *et al*., 2003; Sorokin *et al*., 2004). In the functional analysis by Kim *et al* (2006), fosmids generated from large-insert soil DNA screening for esterase enzyme provided only one acive fosmid clone out of 60,000. Litthauer *et al* (2010) recently showed that tributyrin agar plate screening of metagenomic clone libraries showed lipolytic activity of hydrolysis halo. However, upon sequencing the genes of those active clones, most of them did not have any correlation with liplytic activity, indicating the limitations of screening with trybutyrin.

The unsuccessful attempts in identification of lipase/esterase genes from Chitu metagenomic Fosmid library could be attributed to the problems associated with the expression of cloned gene in the heterologous host (Handelsman, 2004) and low frequency of target sequence in the metagenomic library (Henne *et al*., 1999). Limitations associated to protein expression and substrate specificity in plate screening may thus explain the absence of lipase/esterase degrading activities as well as the limited detection of cellualse activities in this study. This may reqiure for the development of robust and efficient functional screens and selection methods as well as new activities or molecules (Lorenz *et al*., 2002, Streit *et al*., 2004, Uchiyama *et al*., 2005, Litthauer *et al*., 2010; Nyyssonen *et al*., 2013).

The enzymatic activity screens in this study didn‘t include the prominent *Bacillus sp*. although this bacterial species are generally well-known for their various sources of enzymes (Ito *et al*., 1989; Rao *et al*., 1998). From the diversity study in Chapter five, pyrosequencing revealed that *Bacillus spp*. were rare, while *Anoxybacillus* species dominated Lake Chitu (Table 5.3). Illumina sequences of enzymatically active fosmid clones of Lake Chitu (Chapter seven) again revealed that a total of six ORF nucleotide sequences that belonged to peptidases/proteases, and different enzyme families were mostly associated with genus ***Delftia***, ***Comamonas***, ***Bordetella***, ***Achromobacter***, ***Gallionella*****,** **and** ***Brevundimonas*** (Appendix VII).

Thus, the fosmid DNA library constructed for Lake Chitu might not have covered enough DNA of fewer biomass of *Bacillus sp* from sediment samples of the lake. This might explain the reason that the enzymatic activity screens in this study didn‘t include the prominent *Bacillus sp*. A novel alpha-amylase enzyme from *anoxybacillus sp* was recently described by Goh *et al*

(2013) as source of amylase enzymes, indicating the potential of Lake Chitu sediments for amaylase enzyme families from *anoxybacillus sp*. that are abundant in this lake.

6.4 Conclusion

In this study, a metagenomic DNA Fosmid library was constructed from Lake Chitu sediments in Ethiopian Rift Valley area. The results of this study indicated that only a few fosmid clone libraries were generated for Lake Chitu (a total of 288 Mb DNA library size from 7, 200 fosmid clones) due to the difficulties in recovering enough DNA from salt and humic acid-rich alkaline sediment samples. The Fosmid library was screened for proteolytic, amylolytic, cellulosic and lipolytic activities. The functional screens for 7, 200 total fosmid clones of Lake Chitu were low in frequency for cellulosic enzymes and nil for lipolytic expression. However, the number of hits for protease and amylase expressing fosmid clones for Lake Chitu was relatively high compared to previous reports made on functional plate screenings of the enzymes from other places. The fosmid screens from Lake Chitu were limited to a few genera and the common *Bacillus sp* was‘t demestrated, indicating the requiremet for Fosmid libraries that include functional screens of wider microbial genera.

This study generally indicated that Lake Chitu microorganisms have a great potential for enzymes. The constructed high-molecular Fosmid DNA library of Lake Chitu could be used as a useful material to hunt for novel several other useful genes and gene products. Finally, this study showed that interesting novel enzymes could be detected from Lake Chitu if appropriate biochemical tests and highthrouput/ improved functional screening techniques are applied.

CHAPTER SEVEN

Genes Conferring Microbial Activities in Fosmid DNA Library of Chitu Soda Lake

7.1 Introduction

Microorganisms dominate extreme environments including the soda lakes in terms of total biomass and metabolic activities, playing significant roles in primary productivity as well as in biogeochemical transformations or energy recycling (Antony *et al*., 2013; Grant and Sorokin, 2011). The various environmental challenges such as extremes or flactuation of pH, salinity, nutrient scarcity, temperatures, pressure etc within the soda lake environment have directed the microorganisms to adopt diverse survival and growth strategies, resulting in novel biochemistry encoded within their genetic resources (Grant *et al*., 1990).

Sequence–based metagenomics detects the metabolic potentials of a microbial community and discloses ―functional" genes specific for certain adaptive metabolic processes that play key roles for their survival in a given habitat (Bohannon, 2007; Thomas *et al*. 2012). Consequently, such genes can be used as indicators for the presence or absence of certain groups of microorganisms or of certain biochemical pathways in that particular habitat. A study on Mono Lake, California, revealed genes responsible for ribulose bisphosphate carboxylase/oxygenase involved in autotrophic fixation of carbon dioxide (Giri *et al*., 2004). Ward *et al*. (2000) identified genes coding for ammonia monooxygenase, the key enzyme of autotrophic oxidation of ammonia to nitrite. Similarily, the existence of a bisulfate reductase marker genes that perform dissimilatory sulfate reduction and involved in methanogenic pathway were identified (Scholten *et al*., 2005).

In this study, a small metagenome sequence dataset was generated from selected enzymatically active Fosmid clones of Lake Chitu using the Illumina sequencing technology in order to obtain important preliminary insights into the metabolic potential and functional genes of microbial communities in Lake Chitu. COG distribution and KEGG metabolic maps of Chitu fosmid reads were used for functional and pathway analysis.

7.2 Results

7.2.1 Illumina Sequence Assembly and Gene annotation

The Illumina sequences of Chitu enzymatic fosmid clone resulted 221, 310 short reads with average length of 98 bases, which equals to a total of 22 Mb (result from Illumina sequence of fosmid clones). The total cleaned appropriate sequence was about 20 Mb. The short sequence reqads were assembled and a total of 1439 sequnces of the short reads were identified for protein coding genes (ORFs). Approximately, 242 contigs (>500 bp) of predicted ORF sequences were annotated (Table 7.1).

Table 7.1: Illumina Sequence Assembly Resul of Lake Chitu Fosmids

No. of Contig (>500 bp)	Size of all contigs (bp)	N50 (bp)	MaxLen (bp)
242	**322637**	**1606**	**20692**

7.2.2 COG and KEGG Categories Identified in Lake Chitu Sediments

The metagenomic fosmid sequences of Lake Chitu were searched against Clusters of Orthologous Groups (COG) of the NCBI to look at characteristics associated with some functions (Table 7.2). Accordingly, the genes engaged in COG clusters detected in Lake Chitu Fosmid library demonstrated charactirstics linked with functions related to ‗energy production and conversion', ‗inorganic ion transport and metabolism', ‗aminoacid metabolism and transport', ‗cellwall/memberane/envelope biogenesis', ‗lipid metabolism', ‗Charbohydrate and nucleotide metabolism', ‗coenzyme metabolism', ‗replication and repair', and ‗defense mechanisms' (Fig. 7.1). Lower hits of other COG functional categories largely represnted functions specific to eukaryotes (Fig. 7.1). Approximately, 30% of the predicted ORF sequence reads obtained matches in 20 COG categories (Fig. 7.1) and 27% of the predicted ORF sequence reads aqiured matches in 15 KEGG categories (Fig. 7.2).

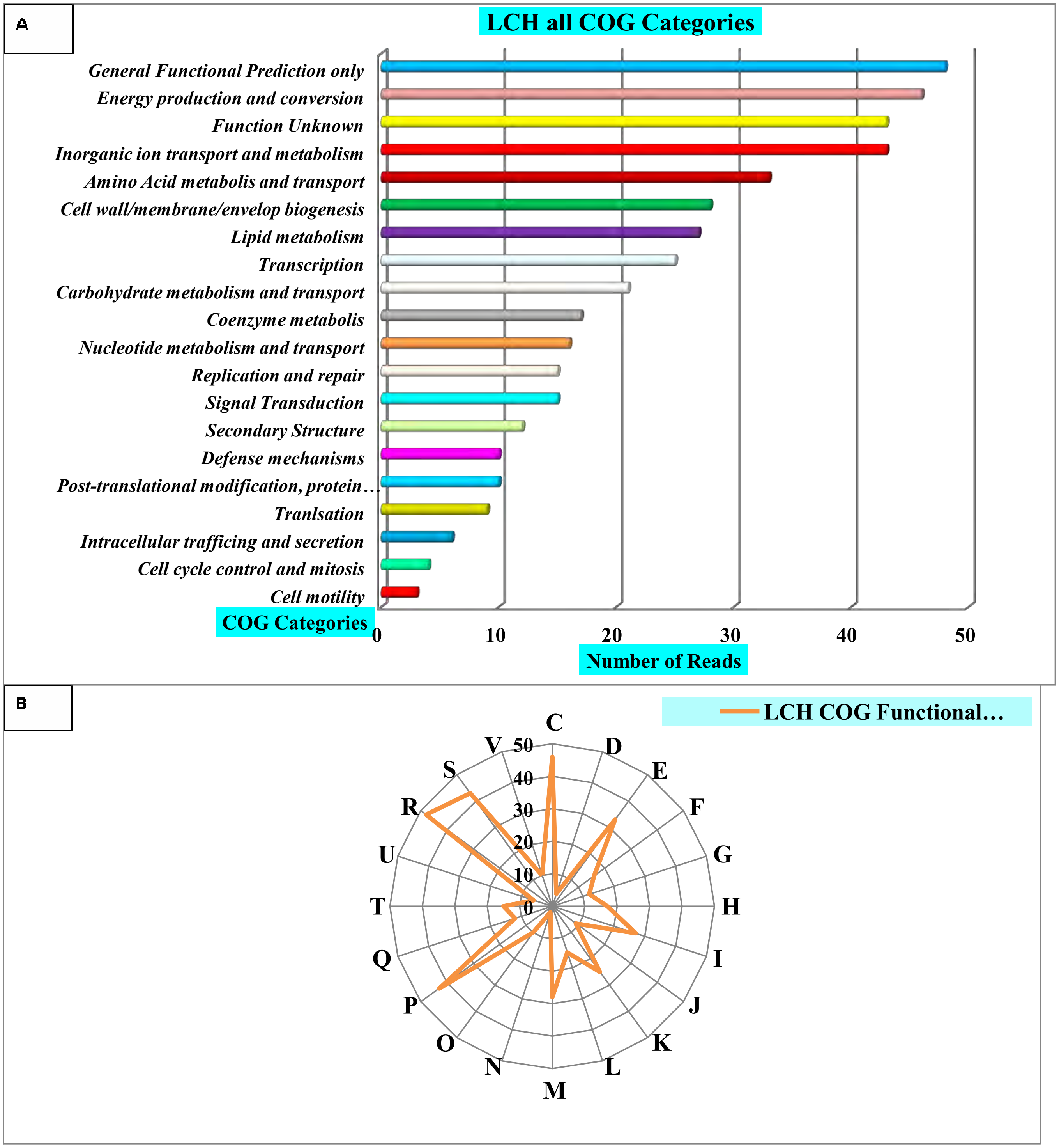


Figure 7.1: Functional Assignment of Chitu Fosmid Library Illumina Reads**.** BLASTX search for similarirty against the COGs database. Read numbers were assigned to specific COG functional categories [**See Appendix VI and VII** for COG categories (A) and Codes (B) respectively].

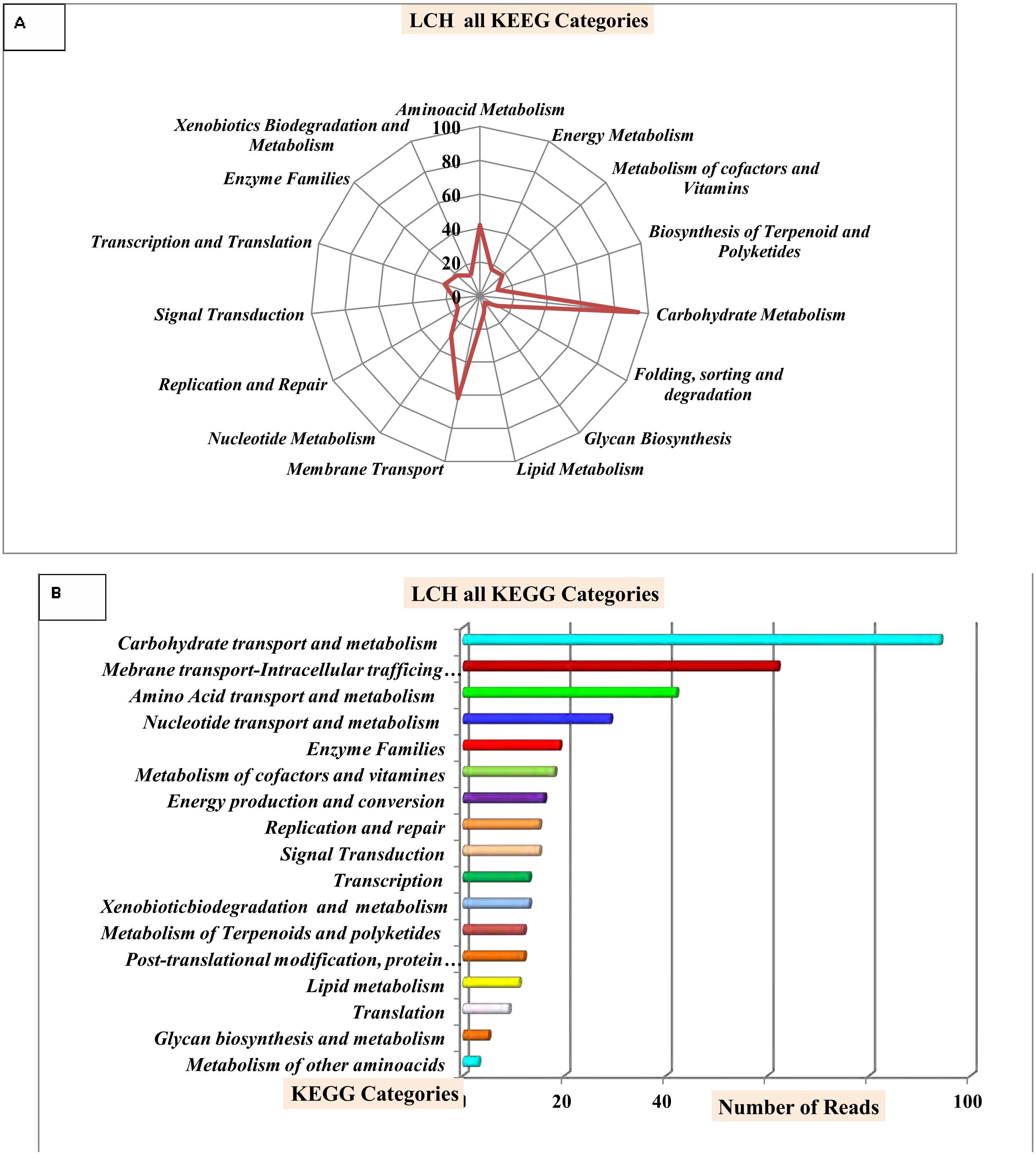


Figure 7.2: Functional Assignment of Chitu Fosmid Library. BLASTX analysis against the KEGG database. Read numbers were assigned to specific KEGG identifiers.

Abundant hits for KEGG categories were also obtained for Lake Chitu in carbohydrate metabolism and transport, membrane transport, and aminoacid metabolism and transport (Fig. 7.2). Major COGs and KEGG categories of Fosmids of Lake Chitu sediments were assigned to functions in cellular processes/signaling and metabolism while the remaining majority of other sequences were either found to be poorly characterized (Table 7.2) or not assigned to any COG functional categories. The detailed list of all COGs and KEGG categories as well as metabolic pathways are provided in appendix VII.

Table 7.2: Major COG and KEGG Categories Identified in Lake Chitu Fosmid Library

Major COG Category	**KEGG Category**	**# of Sequence hits/reads**
Cellular processes and signaling	Mebrane transport-Intracellular trafficking and secretion	62
	Signal Transduction	15
	Post-translational modification, protein turnover, chaperone functions	12
Information storage and processing	Replication and repair	15
	Transcription	13
	Translation	9
Metabolism	Carbohydrate transport and metabolism	94
	Amino Acid transport and metabolism	42
	Nucleotide transport and metabolism	29
	Enzyme Families	19
	Metabolism of cofactors and vitamines	18
	Energy production and conversion	16
	Xenobioticbiodegradation and metabolism	13
	Metabolism of Terpenoids and polyketides	12
	Lipid metabolism	11
	Glycan biosynthesis and metabolism	5
	Metabolism of other aminoacids	3
Poorly Characterized	General Function predicton only	48
	Function Unknown	43

7.2.3 Metabolic Profile

Some of the hits obtained from the current fosmid sequence of Chitu in the KEGG category could be used to describe the capability and involvement of Chitu sediment miceobes in important biogeochemical transformations or nutrient recycling. Major energy metabolism constituted Methane (Carbon), Nitrogen, and Sulfur metabolism chiefly allied to the family *Comamonadaceae* and *Alcaligenaceae* (Table 7.3). One hit for a gene responsible for Oxidative phosphorylation that expresses Cytochrome O ubiquinol oxidase protein was also observed.

Table 7.3: Major Energy Metabolism Categories of Chitu Sediments

Energy Metabolism Function/Path way	**KEEG Orthology**	**Predicted Protein/gene Product**	**Taxonomic affiliation**
Methane metabolism [PATH:ko00680] (4 hits)	K03520	Carbon monoxide dehydrogenase	*Betaproteobacteria, Burkholderiales,* ***Comamonadaceae***
Methane metabolism [PATH:ko00680] (1 hit)	K03518	2Fe-2S iron-sulfur cluster-binding domain-containing protein	*Betaproteobacteria, Burkholderiales,* ***Comamonadaceae***
Methane metabolism [PATH:ko00680] (3 hits)	K03781	Catalase	*Betaproteobacteria, Burkholderiales,* ***Comamonadaceae, Acidovorax***
Nitrogen metabolism [PATH:ko00910] (1 hit)	K01673	Carbonate dehydratase (Hydro-lyases)- CO_2 forming	*Betaproteobacteria, Burkholderiales,* ***Alcaligenaceae***
Nitrogen metabolism PATH:ko00910] (2 hits)	K01745	Histidine ammonia-lyase –NH_3 forming	*Betaproteobacteria, Burkholderiaceae* ***Comamonadaceae***
Nitrogen metabolism PATH:ko00910] (2 hits)	K01672	LysR family Transcriptional regulator	*Betaproteobacteria, Burkholderiales,Burkholderia* ***Alcaligenaceae** ,* ***Achromobacter***
Sulfur metabolism [PATH:ko00920] (2 hits)	K00641	Sulfatases and Homoserine O-acetyltransferase	*Betaproteobacteria, Burkholderiales,* ***Comamonadaceae*** *Delftia acidovorans SPH-1*
Oxidative phosphorylation [PATH:ko00190] (1 hit)	K02298	Cytochrome O ubiquinol oxidase protein	*Betaproteobacteria, Burkholderiales,* ***Comamonadaceae***

7.3 Discussion

In this study, several COGs and KEGG categories with metabolic pathways were ditected in Lake Chitu (see **Appendix VII)**. COG cluster associated with ‗cell wall/memberane/envelope biogenesis' was detected in Lake Chitu which indirectly indicates the presence of enzymes such glycosyltransferases and phosphate isomerases. They were also previously detected in deep- sea samples from high-pressure-loving bacterium *Photobacterium profundum* (El-Hajj *et al*., 2009). Because Chitu Lake sediment samples were sampled ranging from the shore upto depth of 16 meters, genes associated with hydrostatic pressure ‗cell wall/memberane/envelope biogenesis would be relevant to organism from deep sediments. The presence of these types of genes is probably due to the fact that microbes under deep and high pressure sediments may need them for keeping their cell wall integrity (Shi *et al*., 2012).

Genes related to signal transduction mechanisms such as genes encoding histidine kinas, NLPA lipoprotein for two component systems, K^+- transporting ATPase subunits, and Chemotaxis protein methyltrasnferase cheR were also detected in Lake Chitu. Histidine kinases were reported as significantly essential proteins for chemotaxis and quorum sensing in micoorganisms Wolanin *et al* (2002). Similarly, several of these types of genes and other related genes have previously been detected in the metagenomic data from deep-sea water samples playing related significant roles associated to coping with changing environmental factors (Wu *et al*., 2013).

Several genes encoding for functions related to repair and replication regulation (ATP depenedant DNA ligase, Ku domain-containing protein, Exonuclease ABC subunit B, Helicase), Transcription-repair coupling factor, Stress response DNA-binding protein (Dps), and Exonuclease VII small subunit for DNA mismatch repair pathway were detected in Lake Chitu metagenome (Appendix VII). As it has been reported earlier, replication, recombination, and repair genes are strongly associated with adaptation to and survival in harsh environments for prokaryotic communities of deep-sea water samples (Nakai *et al*., 2011; Wu *et al*., 2013). Simlarly, the predominance of DNA repair and replication regulation protein coding genes in Lake Chitu, most likely, show the capability of DNA repair potential of the resident prokaryotic communities to survive and shield themselves from probable DNA damage in the harsh soda lake habitats.

Genes encoding a number of extracellular membrane ligands/receptors of various functions associated with mebrane transport or intracellular trafficking and secretion were detected in Lake Chitu fosmid library reads. Among these the most dominant functions include, solute-binding protein families, ATPase component ABC-type nitrate/sulfonate/bicarbonate transport system, Fe (3^{+})-chelate uptake/transporting ATPase, permease ABC-type multidrug transporter systems, type I secretion system, ATPase-K/Mg/**Cd**/**Cu**/**Zn**/Na/Ca/H-transporter, molybdenum ABC transporter protein, glutamate/aspartate ABC superfamily, D-galactonate transporter, and lipid ABC transporter ATPase. As described above, the genes encoding the function of membrane transilocation of heavy metals such as Cd, Cu, and Zn wee detected in Lake Chitu sediments. It has been reported that some microorganisms surviving in hydrothermal fluids have the potential to remove heavy metals from the fluid (Jannasch, 1995; Wu *et al*., 2013). Lake Chitu microorganisms are expected to acqured such unique abilities because of the current or earlier influence from thermal hotsprings feeding the lake from nearby tectonic areas.

Indications of the metabolic potential of Lake Chitu prokaryotic community could be generated using the functional annotation of KEGG orthology. Most of the KEGG pathway identified predominantly harbored genes assigned to carbohydrate metabolism pathways (Table 7.2), indicating the presence of metabolic pathways such as glycolysis /gluconogenesis, pyruvate metabolism, and the TCA (tricarboxylic acid) cycle. The presence of the Glycolysis/Glycogenesis path way in Lake Chitu metagenome with genes responsible for NAD-dependent aldehyde dehydrogenase, Acetyl-CoA Synthase and betaine aldehyde dehydrogenase were confirmed with considerable affliation to both *Betaprotobacteria* and *alphaprotobacteria* (Appendix VII). Previous studies showed that the sequences for carbonmonoxide dehydrogenase [KEGG: K03520] and acetyl-CoA synthase enzyme [KEGG: K01895] genes show the presence of the reductive CoA pathway (Wood- jungdahl pathway) that is characteristic for acetogenesis by acetogenic bacteria (Nyyssonen *et al*., 2013). Pyruvate metabolism was exhibited with the detection of various genes encoding Acetyl CoA carboxylase, malic enzyme, phosphoenolpyruvate carboxylase, carbamoyl phosphate synthase, and proteins containing phosphate acetyl/butarryl transferase in Lake Chitu.

The pathways for the xenobiotics biodegradation and metabolism of organic compounds such as carbazole (gene encoding AMP-dependent synthase/ligase) [KEGG: K09461], benzoate (genes encoding 3-oxoadipate enol-lactone hydrolase and 3-oxoacid CoA transferase) [KEGG: K01055], and 2, 4-Dichlorobenzoate (genes encoding 4-hydroxybenzoate 3-monoxygenase) [KEGG: K00481] were found to exist in the community of Lake Chitu sediments (Appendix VII). Most of these genes were found to be associated with the species of the genera *Bordetella*, *Achromobacter*, *Delftia* and *Ralstonia* within the *Betaprotobacteria.* Such related microbial functional genes involved in organic contaminant degradation were detected from oil-contaminated fields (Liang *et al*., 2011).

Methane metabolism was confirmed by the presence of the genes encoding carbon monoxide dehydrogenase [KEGG: K03520], catalase [KEGG: K03781], and 2Fe-2S iron-sulfur cluster-binding domain-containing protein [KEGG: K03518] affliated to the group of *Burkholderiales* within the *Betaproteobacteria* (Table 7.3, Appendix VII). The presence of carbon fixation pathways via CO_2 fixation in Lake Chitu was also anaylsed from the fosmid sequences. Indication for anearobic carbon fixation via methane metabolism pathway (PATH: ko00680) was evidenced by the presence of the gene for the key enzyme, carbon monoxide dehydrogenase /acetyl-CoA synthase. The catabolism of carbon compounds via carbon loss was evidenced from the detection of the cytochrome oxidase gene [KEGG: K02298] which is responsible in oxidative phosphorylation [PATH: Ko00190] (Appendix VII).

The detection of assimilatory nitrogen metabolic components, such as histidine ammonia-lyase (NH_3 forming) [KEGG: K01745], carbonate dehydratase (CO forming) [KEGG: K01673], and LysR family transcriptional regulator [KEGG: K01672] in addition to an ABC-type nitrate/sulfonate/bicarbonate transport system shows the existence of nitrogen cycle in Lake Chitu with closest association to *Achromobacter* and *Comamonadaceae* in the *Betaprotobacteria* (Appendix VII).

Also putative sulfatases with closest relative to NCBI locus YP_001562739 and YP_001562746 and homoserine O-acetyltransferases [KEGG: 00641] with closest similarity to *Delftia acidovorans SPH-1* within the *Betaprotobacteria* in Lake Chitu metagenome further indicates the roles of these enzymes in sulfur cycle in this lake. Although major marker genes for sulfur reduction and oxidation were not detected from the current study of Chitu fosmid illumina sequences, several SOB and SRB were represented in Lake Chitu (see Table 5.2 in Chapter 5), demonstrating the importance of such metabolism in this lake.

In silico illumina sequence analysis of the enzymatically active Lake Chitu fosmid clones gave a total of six ORF nucleotide sequences that belong to different peptidase/protease domains-ATP-dependent Clp protease (proteolytic subunit ClpP), endopeptidase La, Peptidase C56 (PfpI), protein containing Peptidase M23 domain, LexA repressor, and Xaa-Pro dipeptidase (KEGG: K01262, K01356, K08259, K05520, K01338, and K01358) with closest affiliation to *Gallionellaceae*, *Caulobactereraceae*, and *Burkholderiales/Comamonadaceae* within the *Betaprotobacteria* (Appendix VII). However, the existence of amylolitic and cellulosic enzyme families were not detected from the current Chitu Fosmid sequences. More interestingly, 2 hits for esterase-like enzyme gene sequences, including glycerophosphoryl diester phosphodiesterases [KEGG: K01126] (Appendix VII), resspnsible for glycerophospholipid (Lipid) metabolism were detected although enzyme that belong to eaterases or lipases were never shown up during the functional/activity plate screening.

7.4 Conclusion

This work represents a first attempt to reveal an overview of Lake Chitu potential metabolic genes and pathways using Fosmid metagenomic DNA library constructed from sediment samples. The short illumina sequence of Lake Chitu Fosmid library gave a preliminary genetic insight to highlight the biogeochemical transformations that occur in this lake. Functional analysis using COG distribution revealed that Lake Chitu prokaryotes have adapted more energy production and conversion, inorganic ion transport and metabolism, amino acid metabolism, and Cell wall/membrane/envelop biogenesis life styles in the harsh soda lake environments. Pathway analysis using KEGG metabolic maps of the Fosmid reads indicated that the microbial population in Lake Chitu play roles in many ways of the nitrogen cycle, sulfur cycle as well as those, in many forms, repair, use and recycle carbon through metanogenesis, acetogenesis, glycolysis, TCA cycle, and oxidative phosprylation associated largely with *Burkholderiales, Delftia acidovorans SPH-1*, *Achromobacter*, *Comamonadaceae, Ralstonia, Bordetella*, within the *Betaprotobacteria*.

CHAPTER EIGHT: Conclusion and Future Prespectives

8.1 Summary Conclusion

It is estimated that less than 1% of all known environmental bacteria are culturable due to the difficulty to accurately mimic specific needs and the conditions suitable for microbial growth in the laboratory. It is currently believed that the uses of genetic approaches including metagenomics can bypass culturing of microorganisms and allow the identification of novel gene sequences and complete metabolic pathways of unculturable microorganisms from the environment.

In this study, 16S- rRNA-based metagenomics was applied to Chitu, Shalla and Abijata soda lakes in the Ethiopian Rift Valley to investigate thier microbial community composition and phylogenetic releationships. Here, four 16S- rRNA-clone libraries were generated for Chitu and Shalla lakes by using metagenomic DNA from the sediment samples, and published bacterial and actinomycetes domain specific 16S rRNA primers. In addition, 454-pyrosequencing was used to infere microbial phylogeny and compositions of Chitu, Shalla and Abijata soda lakes using the V3-V4 hypervariable specific regions of the bacterial 16S-rRNA genes with barcoding strategy. Moreover, metagenomic approaches were used to construct a large- insert fosmid (pCC1FOS vector) metagenomic DNA library of Chitu Soda Lake to identify potential enzyme activities using functional /activity-based metagenomic screening techniques. Further, potential metabolic genes and pathways of Lake Chitu were infered from short illumina sequences of selected fosmids.

The 16S rRNA sequences generated for Chitu and Shalla lakes using the Bacterial and actinobacterial universal pimers were related to low and high G+C Gram-positive bacteria (*Firmictes* and *Actinobacteria*), CFB, and *Protobacteria* of the γ-, α-, *ß*- and Δ- ***Proteobacteria*** subdivisions. Most of the clones were phylogenetically allied with members of the γ- subdivision of *Protobacteria* (accounted for 34–43% of sequences), followed by *Firmicutes* (affiliated sequences at 16–26%) in both lakes. They were found to harbor biotechnologically and ecologically significant representatives of microbes that belong to the members of the group *Rhodobaca*, *Alkalimonas*, *Nitrinicola*, *Thioalkalivibrio*, *Methylophaga*, *Desulfuromusa*, *Bacillus*, *Clostridium*, *Dethiobacter*, *Cytophagales*, *Actinobacteria*, and *Bacteroidetes*.

Barcode- tagged 454 pyrosequencing of prokaryotic community of Chitu, Shalla, and Abijata soda lakes revealed much wider diversity of the soda lakes microbial communities that include primary producers (Mostly cyanobacteria and anoxygenic phototrophs), aerobic heterotrophs, fermentative and respiratory anaerobes (sulfate-reducing bacteria, SRB), acetogens, and methanogens. These majorily included the methanogenic archaea (*Methanocalculus*), methane-oxidizing (*Methylomicrobium sp.*), Cyanobacteria, photosynthetic purple bacteria (*Halorhodospira*), non-sulfur purple bacterial genera (*Rhodobaca)*, SOB (*Thioalkalivibrio*, *Thioalkalispira, Sulfurimonas*), SRB (*Desulfobacterium*, *Desulfosarcina*, *Desulfurivibrio*, *Desulfonatronovibrio*), *Actonobacteria*, *Anoxybacillus*, *Bacillus*, *Clostridia*, *Nitrospira*, *Paracoccus*, *Rhodobacteraceae*, *Pseudomonas*, *Bacteroidetes*, *Deinococcus-Thermus*, *Planctomycetes*, and *Spirochetes*. The results from sequence analysis suggest that the soda lakes under study harbor significant higher rich diversity and unique representatives of bacterial overall diversity and community composition.

A total of 288 Mb metagenomic DNA Fosmid library size was constructed from Lake Chitu sediments in Ethiopian Rift Valley area. The screening of 7, 200 of total Fosmid clone library of the lake sediments resulted significant proteolytic, amylolytic, cellulosic enzyme activities. The frequency of protease and amylase expressing fosmid clones for Lake Chitu was relatively high compared to previous reports made on functional plate screenings of the enzymes from other places. However, the fosmid screens were low for cellulosic enzymes and nil for lipolytic expression. The constructed high-molecular Fosmid DNA library of Lake Chitu could serve as a useful reference material for future screening of other novel useful genes and gene products of biotechnological implications.

Functional analysis of Lake Chitu Fosmid DNA using short illumina sequence provided that Lake Chitu prokaryotes have adapted more energy production and conversion, inorganic ion transport and metabolism, amino acid metabolism, and Cell wall/membrane/envelop biogenesis life styles in the harsh soda lake environments. Pathway analysis using KEGG metabolic maps of the Fosmid reads indicated that the microbial population in Lake Chitu play roles in many ways of the nitrogen cycle, sulfur cycle as well as those, in many forms, repair, use and recycle carbon through metanogenesis, acetogenesis, glycolysis, TCA cycle, and oxidative phosprylation associated largely with *Burkholderiales, Delftia acidovorans SPH-1*, *Achromobacter*, *Comamonadaceae, Ralstonia, Bordetella*, within the *Betaprotobacteria*.

8.2 Recommendations

- ✓ DNA retrieval and purification from the soda lake sediments are generally difficult due to large amount of salt and humic acid contaminants. In this study, different available methods have been optimized, and the Zhou method was finally selected. The quantity of the DNA extracted was not sufficient after long steps of purification. Other novel environmental DNA extraction and purification of methods could be useful to minimize biases in diversity and functional studies as a result of insufficient, degraded, and contaminated DNA samples.

- ✓ Variations were observed among dominanat bacterial phyla between 16S-rRNA clone library- and bar-coded 454 pyrosequencing- based microbial diversity studies due to biases in sampling techniques. The lake sediment samples used for 16S-rRNA clone library were collected from shore sites and up to a maximum depth of only 0.5m. For pyrosequencing, lake sediments were pooled from different depths from shore sites to a maximum bottom depth of 16m.

- ✓ Generally, PCR-based techniques are known to impart significant biases to bacterial diversity and compositions studies. Non-PCR techniques such as Shotgun DNA Sequencing of soda lake sediment samples could be applied for yet deep information regarding microbial composition and functions.

✓ The pyrosequencing and fosmid library data of this study have identified unique microbial functions and metabolic genes and gene products from the soda lakes. Detailed culture-based techniques should still be applied to these lakes so as to capture biotechnologically and ecologically very vital and novel bacteria and archaea species. The present pyrosequence study could serve as a groundwork taxonomic inventory for further detailed functional and culture-dependent analysis of the soda lakes studied. Hope, future researches in this field will obviously amend the technical pitfalls that might have associated with the current metagenomic investigations and under or over-representation of the coverage of microbial diversity of the three soda lakes.

✓ Functional/activity plate screening of Lake Chitu Fosmid library was limited to a few enzymes, and lipase/esterase enzyme was not ditected in this study. High throughput functional screening techniques are recommended to identify yet undetected or unique/novel enzymes and other industrially important products from the fosmid clones.

✓ In parallele to metagenomics, future studies should also involve a wider study on micromial community of the soda lakes using other functional genomics such as transcriptomics, metabolomics, and proteinomics.

✓ The metabolic profile of Lake Chitu optained from fosmid sequences was not comprehensive. Recently, no studies were made regarding the availability of heavy hydrocarbon compounds in Lake Chitu, but the presence of these pathways may indicate a current or long-ago associations with these hydrocarbon chemicals. Thus, further studies on the analysis and detection of organic compounds in the sediments of Lake Chitu would strengthen the current findings. Detailed study of the soda lake environments including the nearby hostsprings that feed them as well as chemical composition of water and sediment samples of the lake would provide broader map of and insight into the microbial diversity and geobiochemical processes in Lake Chitu.

- ✓ Further advanced bioinformatics tools should be used to interprate the huge sequence data generated from clone libraries, pyrosequencing and illumina in this PhD study.

- ✓ In the present study, the sequences from pyrosequencing and illumina revealed that the lakes are under huge antropogenic influences, and the sequences hits for very important genes and gene products are little. These suggest that proper management and conservation procedures should be considered for unique microbial and genetic resources of Ethiopian soda lakes.

8.3 Potential Manuscripts

1. 16S rRNA Clone Library-based Bacterial Diversity in Chitu and Shalla Soda Lakes of Ethiopia (**in preparation**)

2. Analysis of Prokaryotic Communities of Ethiopian Soda Lakes Using BarcodeTagged 16S rRNA gene pyrosequencing (**Under Co-aouthors review**)

3. Large-insert Fosmid Metagenomic DNA Library Construction and Analysis of Enzymatic Potential of Chitu Soda Lake, Ethiopia (**in preparation**)

4. Genes Conferring Microbial Activities in Fosmid DNA Library of Chitu Soda Lake, Ethiopia (**in preparation**)

REFERENCES

Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V., and Polz, M.F (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple operons. *J Bacteriol,* **186:** 2629–2635.

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F (2005). PCR induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and environmental microbiology,* **71**: 8966-8969.

Adrio, J.L, and Denain, A.L (2014). Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules*, **4:** 117-139.

Alemayehu, T., Ayenew, T. and Kebede, S (2006). Hydrogeochemical and lake level changes in the Ethiopian Rift. *Journal of Hydrology,* **316:** 290-300.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215:** 403–410.

Amann, R.I., Ludwig, W., and Schleifer K.H (1995). Phylogenetic identifica-tion and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev*, **59:**143–169.

Amann, G., Stetter K.O., Llobet-Brossa E., Amann R., and Anton J (2000). Direct proof for the presence and expression of two different 16S rRNA genes in individual cells of *Haloarcula marismortui*, *Extremophiles*, **4:** 373–376.

Angelov, A., Mientus, M., Liebl, S., & Liebl, W. (2009). A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Systematic and Applied Microbiology*, **32:** 177–185.

Anton, J., Oren, A., Benlloch, S., Rodriguez-Valera, F., Amann, R., and Rossello-Mora, R (2002). ―Salinibacter ruber gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds." *Int J Syst Evol Microbiol*. **52** (Pt 2): 485-91.

Antony, C.P., Doronina, N.V., Boden, R., Trotsenko, Y.A., Shouche, Y.S., Murrell, J.C (2012a). Methylophaga lonarensis, a novel moderately haloalkaliphilic methylotroph isolated from the soda lake sediments of a meteorite impact crater. *Int J Syst Evol Microbiol*., **62:** 1613–1618.

Antony C. P., Kumaresan D., Hunger, S., Drake, H. L., Murrell J.C and Shouche Y. S (2013). Microbiology of Lonar Lake and other soda lakes, *The ISME Journal*, **7:** 468–476; doi:10.1038 /ismej.2012.137.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nature Genetics*, **25:** 25–29.

Auguet, J.C., Barberan, A., Casamayor, E.O (2010). Global ecological patterns in uncultured Archaea. *The ISME Journal*, **4:** 182–190.

Ayenew, T. (2007). Water management problems in the Ethiopian rift: Challenges for development. *Journal of African Earth Sciences*, **48:** 222-236.

Babcock, D. A., Wawrik, B., Paul, J. H., McGuinness, L. amd Kerkhof, L. J. (2007). Rapid screening of a large insert BAC library for specific 16S rRNA genes using TRFLP. *Journal of Microbiological Methods*, **71:** 156–161.

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, **28:** 45–48.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., and ***et al.*** (2004). The Pfam protein families database. *Nucleic Acids Res*, **32:** D138–D141.

Baye et al (2014). Microbial diversity of two Ethiopiam soda lakes having constrasting physiochemical features, PhD theisis.

BBC Research. Report BIO030 F (2011). Enzymes in Industrial Applications: Global Markets; BBC Research: Wellesley, MA, USA.

Beales, N (2004). Adaptation of microorganisms to cold temperatures, weak acid preservatives, low pH, and osmotic stress: A review. *Comp. Rev. Food Sci. Food Safety*, **3:** 1–20.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, **340:** 783–795.

Binga, E. K., Lasken, R. S., and Neufeld, J. D. (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME Journal*, **2:** 233–241.

Bivin D.B. and Stoeckenius W. (1986). Photoactive retinal pigments in Haloalkaliphilic bacteria. *J Gen Microbiol*., **132:** 2167-77.

Bohannon, J. (2007). METAGENOMICS: Ocean Study Yields a Tidal Wave of Microbial DNA. *Science*, **315 (5818):** 1486–1487.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R. et al (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods,* **10**: 57-59.

Boldareva, E.N., Akimov, V.N., Boychenko, V.A., Stadnichuk, I.N., Moskalenko, A.A., Makhneva, Z.K., Gorlenko, V.M (2008). Rhodobaca barguzinensis sp nov., a new alkaliphilic purple nonsulfur bacterium isolated from a soda lake of the Barguzin Valley (Buryat Republic, Eastern Siberia). *Microbiology,* **77:** 206–218.

Boldareva, E.N., Moskalenko, A.A., Makhneva, Z.K., Tourova, T.P., Kolganova, V.M., Gorlenko, T.V (2009). Rubribacterium polymorphum gen. nov., sp nov., a novel alkaliphilic nonsulfur purple bacterium from an Eastern Siberian soda lake. *Microbiology*, **78:**732–740.

Bouchet, H. and Huot, R. **(**2008). Goldstein Molecular genetic basis of ribotyping. *Clin Microbiol Rev*, **21 :** 262–273

Bowman, J.P., Rea, S.M., McCammon, S.A., and McMeekin,T.A. (2000b). Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. *Environmental Microbiology* **2:** 227-237.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, **99:** 14250–14255.

Brennan, Y., Callen, W.N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernández, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G.P., Mathur, E.J., Short, J.M., Robertson, D.E., and Steer, B.A (2004). Unusual microbial xylanases from insect guts. *Appl Environ Microbio.,* **70:** 3609–3617.

Brock, T.D., Brock, K.M., Belly, R.T and Weiss, R.L (1972). Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Microbiol.,* **84**: 54–68.

Bryantseva,I., Gorlenko,V.M., Kompantseva,E.I, ImhoffJ.F, SulingJ & Mityushina (1999a). *Thiorhodosp irasibirica* gen. nov., sp. nov., a new alkaliphilic purple sulfur bacterium from a Siberian soda lake. *Int J Syst Bacteriol* **49**: 697–703.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K (2006). The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4:**458–468.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I, Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, **7:** 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N. et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J,* **6:** 1621-1624.

Caruccio, N. (2011). Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. *Methods in Molecular Biology*, **733***:* 241–255.

Chandrasekharaiah, M., Thulasi, A., Vijayarani, K., Kumar, D. P., Santosh, S. S., Palanivel, C., et al. (2012). Expression and biochemical characterization of two novel feruloyl esterases derived from fecal samples of *Rusa unicolor* and *Equus burchelli*. *Gene*, **500**: 134–139.

Chevreux, B., Wetter, T., & Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information computer science and biology. *Proceedings of the German Conference on Bioinformatics*, **99:** 45–56.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O'Toole, P.W (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic acids research,* **38**: e200.

Cole, J. R., Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., et al. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research*, *31*, 442–443.

Curtis, T.P., Sloan, W.T, and Scannell, J.W (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci*. USA, **99:** 10494–10499.

Dadheech, P. K., Glockner, G., Casper, P., Kotut, K., Mazzoni, C. J., Mbedi, S. and Krienitz, L (2013). Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake. *FEMS Microbiol Ecol,* **85 (2):** 389-401.

Daniel R. (2005). The metagenomics of soil. *Nat Rev Microbiol,* **3:** 470–478.

Drake, H.L., Gossner, A.S., Daniel, S.L (2008). Old acetogens, new light. *Ann N Y Acad Sci*., **1125:** 100-128.

DasSarma, S., Arora, P. (2002). *Halophiles*, *Encyclopedia of Life Sciences*; Nature Publishing Group: London, UK.

Delavat, F., Phalip, V., Forster, A., Plewniak, F., Lett, M.-C and Lie`vremont, D (2012). Amylases without known homologues discovered in an acid mine drainage: significance and impact. *SCIENTIFIC REPORTS*, **2:** 354, | DOI: 10.1038/srep00354.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics (Advance online version).

Delgado, O., Quillaguama, J., Bakhtiar S., Mattiasson B., Gessesse A., and Hatti-Kaul R. (2006). *Nesterenkonia aethiopica* sp. nov., an alkaliphilic, moderate halophile isolated from an Ethiopian soda lake. *International Journal of Systematic and Evolutionary Microbiology*, **56:** 1229–1232.

Demain, A. L. (2000). Microbial biotechnology. *Trends in Biotechnology*, **18(1):** 26 – 31.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72:** 5069–5072.

Dias, A. C. F., de Pereira E Silva, M. C., Cotta, S. R., Dini-Andreote, F., Soares, F. L., Salles, J. F., et al. (2012). Abundance and genetic diversity of *nifH* gene sequences in anthropogenically affected Brazilian mangrove sediments. *Applied and Environmental Microbiology*, **78:** 7960–7967.

Dorigo, U., Volatier, L., Humbert, J.F (2005). Molecular approaches to the assessment of biodiversity in aquatic microbial communities. *Water Research*, **39:** 2207–2218.

Duckworth A.W, Grant W.D, Jones, B.E., van Steenburgen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbiol* Ecol **9:**181–191.

Duckworth, A.W, Grant W.D, Jones B.E., Meijer, D., Marquez, M.C., 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.S., Grant, W.D., Jones ,B.E., Meijer, D (2004). Dietzia natronolimnaea sp. nov.

Dumound, M.G., and Murrell, J.C (2005). Stable isoptope probing-linking microbial identity to function. *Nat. Rev.Microbiol.,* **3:** 499-504.

Ekkers, D. M., Cretoiu, M. S., Kielak, A. M., & Elsas, J. D. (2012). The great screen anomaly— a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, **93:** 1005–1020.

Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J.D, Steele, H.L, Reymond, J.L, Jaeger, K.E, Streit, W.R (2006). Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl Environ Microbio;* **72:** 3637-3645.

El-Hajj, Z.W., Tryfona, T., Allcock, D.J., Hasan, F., Lauro, F.M., Sawyer, L., Bartlett, D.H., Ferguson, G.P (2009). Importance of proteins controlling initiation of DNA replication in the growth of the high-pressure-loving bacterium *Photobacterium profundum* SS9. *J. Bacteriol.*, **191:** 6383–6393.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman,H., and Hugenholtz, P (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J,* **4**: 642-647.

Farelly, V., Rainey, F., Stackebrandt, E (1995). Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl Environ Microbiol.*, **61:** 2798-2801.

Feng, Y.D, Cheng-Jie , Hao Pang, Xin-Chun Mo, Chun-Feng Wu, Yuan Yu, Ya-Lin Hu, Jie Wei, Ji-Liang Tang, Feng J.X (2007). Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol;* **75:** 319-328.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins dos Santos VAP., Yakimov, M.M., Timmis, K.N., Golyshin, P.N (2005a). Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol*, **2:** 895–904.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes- Duarte D, Martins Dos Santos V.A., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N (2005b). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbil.,* **7(12):** 1966-2010.

Ferrer, M., Beloqui, A., Timmis,K.N and Golyshin, P.N (2009). Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol., Biotechnol*. **16:**109-123.

Ferrer, M, Giuliano, L., Golyshin, P.N., Yakimov, V (2011). Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part 1: Prokaryotes and environmental settings. Environ Microbiol. doi:10.1111/j.1462-2920.2011.02478.x.

Ferrer M., Guazzaroni, M.E., Richter, M., et al. (2011). Taxonomic and functional metagenomic profiling of the microbial community in the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo, Central Spain). *Microb Ecol*; **62(4):** 824-37.

Ferrer, M., Ruiz, A., Lanza, F., Haange, S. B., Oberbach, A., Till, H., et al. (2013). Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure.Environmental Microbiology.,**15** :211–226.

Foti, M., Sorokin, D.Y., Lomans, B., Mussman, M., Zakharova, E.E., Pimenov, N.V et al. (2007). Diversity, activity and abundance of sulfate-reducing bacteria in saline and hypersaline soda lakes. *Appl Environ Microbiol*, **73:** 2093–2100.

Fry, J. C. (1998) in *Methods in Aquatic Bacteriology*, ed. Austin, B. (Wiley, London), pp. 27–72.

Fu, J., Leiros, H.-K.S., Pascale, D., Johnson, K. A., Blencke, H.-M., & Landfald, B. (2012). Functional and structural studies of a novel cold-adapted esterase from an Arctic intertidal metagenomic library. *Applied Microbiology and Biotechnology*, 1–14.

Galinsky, E.A (1993). Compatible solutes of halophilic eubacteria: Molecular principles, water-solute interaction, stress protection. *Experientia*, **49:** 487–496.

Gareeb, A.P. and Setati, M.E (2009). Assessment of alkaliphilic haloarchaeal diversity in Sua pan evaporator ponds in Botswana. *Afr. J. Biotechnol.*, **8:** 259-267.

Garrity, G.M., Bell, J.A. and Lilburn T (2005). The Revised Road Map to the Manual. In Brenner, Krieg, Staley and Garrity (ed.), Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 2, The Proteobacteria, Part A, Introductory Essays. Springer, New York, pp. 159–220.

Gessesse A, Gashe B.A. (1997a). Production of alkaline protease by an alkalophilic bacteria isolated from an alkaline soda lake. *Biotechnol Lett*, **19:** 479–81.

Gessesse A, Gashe B.A. (1997b). Production of alkaline xylanase by an alkalophilic *Bacillus* sp. isolated from an alkaline soda lake. *J Appl Microbiol*, **83:** 402– 6.

Gessesse A. (1998). Purification and properties of two thermostable xylanases from an alkaliphilic *Bacillus* sp. *Appl Env Microbiol*, **64:** 3533–5.

Gessesse, A., Hatti-Kaul, R., Gashe, B. A. & Mattiasson, B. (2003). Novel alkaline proteases from alkaliphilic bacteria grown on chicken feather. *Enzyme and Microbial Technology;* **32:** 519-524.

Gilbert, J.A and Dupont, C.L (2011). Microbial metagenomics: Beyond the genome. *Annu. Rev. Mar. Sci.*, **3:** 347–371.

Giri, B.J., Bano, N. and Hollibaugh, J.T (2004). Distribution of RuBisCO genotypes along a redox gradient in Mono Lake, California. *Applied and Environmental Microbiology*, **70:** 3443–3448.

Goh, K.M., Kahar, U.M., Chai, Y.Y., Chong, C.S., Chai, K.P., et al. (2013). Recent discoveries and applications of *Anoxybacillus*. *Appl Microbiol Biotechnol*., **97:** 1475–1488. doi: 10.1007/s00253-012-4663-2.

Gonzalez, J. M., Portillo, M. C., Belda-Ferre, P., and Mira, A. (2012). Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS ONE*, **7:** e29973.

Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nature Communications*, **2:** 375.

Grant, S., Sorokin, D.Y., Grant, W.D., Jones, B.E., Heaphy, S (2004). A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. *Extremophiles*, **8:** 421–429.

Grant, W. D., Mwatha, W. E., and Jones, B. E. (1990). Alkaliphiles: ecology, diversity, and applications. *FEMS Microbiol. Rev*., **75:** 255–270.

Grant W.D (1992). Alkaline environments. In: Lederberg J (ed) Encyclopaedia of microbiology, 1st edn. Academic, London, pp 73–80.

Grant, S., Grant, W.D., Jones, B.E., Kato, C., Li, L (1999). Novel archaeal phytotypes from an East African alkaline saltern. *Extremophiles*, **3:**139–145.

Grant, W.D. and Heaphy S. (2010). Metagenomics and recovery of enzyme genes from alkaline saline environments. *Environ Technol*, **31:** 1135–1143.

Grant, W.D and Sorokin, D.Y (2011). Distribution and diversity of Sodal Lake Alkaliphiles. Koki Horikoshi (ed.), Extremophiles Handbook, DOI 10. 1007/978-4-431-53898-1_2.2, Springer.

Grant, W.D. (2003). Alkaline environments and biodiversity. In *Extremophiles*: *Basic Concepts*; Gerdsy, C., Glansdorff, N., Eds.; Encyclopedia of Life Support Systems: Paris, France.

Gupta, R., Beg, Q. K., Lorenz, P (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol.,* **59:** 15–22.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G. et al (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **21**: 494-504.

Haile, G and Gessesse, A (2012). Properties of Alkaline Protease C45 Produced by Alkaliphilic Bacillus Sp. Isolated from Chitu, Ethiopian Soda Lake. *J Biotechnol Biomater*; **2:**4.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman R.M (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. **5:**R245-249.

Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev*, **68(4):** 669-85.

Hawley, E.R and Hess, M (2014). Metagenome sequencing of the prokaryotic microbiota of the hypersaline and meromictic Soap Lake, Washington. *Genome Announc*., 2(1).

Hayakawa, M., Otoguro, M., Takeuchi, T., Yamazaki, T., & Limura, Y. (2000). Application of a method incorporating differential centrifugation for selective isolation of motile actinomycetes in soil and plant litter. *Antonie van Leeuwenhoek*, **78**: 171-185.

Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G (1999). Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol*., **65:**3901–3907.

Henne, A., Schmitz, R., Bomeke, A. M., Gottschalk, G. and Daneil R (2000). Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. *Appl. Environ. Microbiol*., **66:** 3113-3116.

Hoff, K. J., Lingner, T., Meinicke, P., & Tech, M. (2009). Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Research*, **37:** W101–W105.

Hong, S., Bunge, J., Leslin, C., Jeon, S., and Epstein, S.S (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME journal* **3**: 1365-1373

Horikoshi, K (1971). Production of alkaline enzymes by alkalophilic microorganisms. I. Alkaline protease produced by Bacillus No. 221, *Agric. Biol. Chem*., **35:** 1407–1414.

Horikoshi, K (**1999).** Alkaliphiles: Some applications of their products for biotechnology. *Microbiol. Mol. Biol. Rev*, **63:** 735-750.

Horikoshi, K (2006). *Alkaliphiles*: *Genetic Properties and Applications of Enzymes*; Springer: Berlin, Germany.

Hu *et al* (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications.* Article number: 2151.

Huang, Y., Gilna, P., and Li, W. (2009). Annotation of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25:** 1338 – 1340.

Huber, J.A., Mark, Welc D.B., Morrison, H.G, Huse, S.M., Neal, P.R., et al. (2007). Microbial Population Structures in the Deep Marine Biosphere. *Science*, **318:** 97–100.

Humayoun, S.B., Bano, N., Hollibaugh, J.T (2003). Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Appl Environ Microbiol*., **69:** 1030–1042.

Humblot, C. and Guyot, J.P (2009). Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as Pearl Millet slurries. *Applied and Environmental Microbiology*, **75 (13):** 4354e4361

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*, **40:** D306–D312.

Huson, D. H., & Mitra, S. (2012). Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Methods in Molecular Biology*, **856:** 415–429.

Ilmberger, N and Streit, W.R (2010). Screening for cellulase encoding clones in metagenomic libraries. In Metagenomics: Methods and Protocols, Methods in Molecular Biology 668 ed. Streit, W.R. and Daniel, R.E. pp. 177–188. New York: Humana Press, Springer.

Imhoff, J., Sahl, H., Soliman, GS.H., Tru¨per, H.G (1979). The Wadi Natrun: chemical composition and microbial mass developments in alkaline brines of eutrophic desert lakes. *Geomicrobiology,* **1**:219–234.

Ito, S., Shikata, S., Ozaki,K., Kawai, S., Okamoto, K., Inoue, S., Takei, A., Ohta, Y. and Satoh, T (1989). Alkaline cellulase for laundry detergents: Production by Bacillus sp. KSM-635 and enzymatic properties, *Agric. Biol. Chem*. **53:** 1275–1281.

Iwai, S., Chai, B., Jesus, E. d. C., Penton, C. R., Lee, T. K., Cole, J. R., *et al*. (2011). Gene-targeted metagenomics (GT metagenomics) to explore the extensive diversity of genes of interest in microbial communities. In F. J. de Bruijn (Ed.), *Handbook of molecular microbial ecology I: Metagenomics and complementary approaches*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Jain, R., Rivera, M. C., Lake, J. A. **(1999).** Horizontal gene transfer among genomes: the complexity hypothesis. P*roc Natl Acad Sci U S A*, **96:** 3801–3806.

Jannasch, H. (1995). Microbial Interactions with Hydrothermal Fluids. In *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*; American Geophysical Union: Washington, DC, USA; 273–296.

Jiang, Z., Wang, H., Ma, Y., and Wei, D. (2006). Characterization of two novel lipase genes isolated directly from environmental sample. *Applied Microbiology and Biotechnology*, **70:** 327–332.

Jones B.E and Grant W.D (1999). Microbial diversity and ecology of the soda lakes of East Africa. Microbial Biosystems: New Frontiers, Proceedings of the 8th International Symposium on Microbial Ecology Bell CR, Brylinsky M, Johnson-Green P (eds),Atlantic Canada Society for Microbial Ecology, Halifax, Canada,

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

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

Jones, B. V., and Marchesi, J. R. (2007). Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Natural Methods*, 4: 55–61.

Joshi, A.A., Kanekar, P.P., Kelkar, A.S., Shouche, Y.S., Vani, A.A., Borgave, S.B., Sarnaik, S.S (2008). Cultivable bacterial diversity of alkaline Lonar Lake, India. *Microb Ecol* **55:** 163–172.

Kalyuzhnaya, M.G., Khmelenina, V., Eshinimaev, B., Sorokin, DYu., Fuse, H., Lidstrom, M., Trotsenko, Yu A (2008). Reclassification and emended description of halo (alkali) philic and halo (alkali) tolerant methanotrophs of genera Methylomicrobium and Methylobacter. I*nt J Syst Evol Microbiol,* **58:**591–596.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, **36:** D480–D484.

Kebede, E., Getachew, T., Taylor, W.D., Zinabu, G.M. (1994). Eutrophication of L. Hayq in the Ethiopian highlands. *Journal of Plankton Research*, **14:** 1473–1482.

Kelley, D. R., Liu, B., Delcher, A. L., Pop, M., & Salzberg, S. L. (2012). Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Research*, ***40*:** e.

Kennedy, J., O'Leary, N.D., Kiran, G.S., Morrissey, J.P., O'GaraF., Selvin J. and Dobson A.D. (2011). Functional metageno micstrategies for the discovery of novel enzymes and biosurfac-tants with biotechnological applications from marine ecosys-tems. *J Appl. Microbiol*. 111: 787–799.

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." *Curr Microbiol.,* **37:** 94-100.

Khmelenina, V.N., Eshinimaev, B.T., Kalyuzhnaya, M.G., Trotsenko, I. (2000). Potential activity of methane and ammonia oxidation by methanotropic communities from soda lakes of the southern Transbaikal. Mikrobiologiia, **69:** 553–558.

Kim, U.J., Shizuya, H, de Jong, P.J, Birren, B., Simon, M.I (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res*. **20:**1083-1085.

Kim, Y..J., Choi, G.S et al (2006). Screening and characterization of a novel esterase from a metagenomic library. *Protein Expression and Purification*, **45:** 315-323.

Kirkpatrick, J., Oakley, B., Fuchsman, C., Srinivasan, S., Staley, J.T., Murray, J.W (2006). Diversity and distribution of planctomycetes and related bacteria in the suboxic zone of the Black Sea. *Appl. Environ. Microbiol*. **72:** 3079–3083.

Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., et al. (2008). Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*, **36:** 2230–2239.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.*, **305:** 567–580.

Krulwich, T.A., Ito, M., Hicks, D.B., Gilmour, R., Guffanti, A.A (**1998)**. pH homeostasis and ATP synthesis: Studies of two processes that necessitate inward proton translocation in extremely alkaliphilic *Bacillus* species. *Extremophiles*, **2:** 217-222.

Kompantseva, E.I., Sorokin, D.Y., Gorlenko, V.M., Namsaraev, B.B (2005). The phototrophic community found in Lake Khilganta (an alkaline saline lake located in the southeastern Transbaikal Region). Microbiology, **74:**352–361.

Kouker, G. and Jaeger, K.E. (1987) Specific and sensitive plate assay for bacterial lipases. *Applied and Environmental Microbiology,* **53:** 211–213.

Kumar, L., Awasthi, G., Singh, B (2011). Extremophiles: A novel source of industrially important enzymes. *Biotechnology*, **10 :** 1–15.

Kurata, A., Uchimura, K., Kobayashi, T., Horikoshi, K (2010). Collagenolytic subtilisin-like protease from the deep-sea bacterium *Alkalimonas collagenimarina* AC40T *appl. Microbiol. Biotechnol*., **86 (2):** 589-598.

Lakay, F. M., Botha, A., & Prior, B. A. (2007). Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. Journal of Applied Microbiology, 102, 265–273.

Lämmle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., Rupp, S (2007). Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol.,* **127:** 575-592.

Lanze´n, A., Simachew, A., Gessesse, A., Chmolowska, D., Jonassen, I., Øvrea°s, L (2013). Surprising Prokaryotic and Eukaryotic Diversity, Community Structure and Biogeography of Ethiopian Soda Lakes. *PLoS ONE*, **8(8):** e72577. doi:10.1371/journal.pone.0072577.

LeCleir, G.R., Buchan, A., Maurer, J., Moran, M.A, Hollibaugh, J.T (2007). Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. *Environ Microb.,* **9:** 197–205.

Lee, C.C., Kibblewhite-Accinelli, R.E., Wagschal, K., Robertson, G.H., Wong, D.W (2006). Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles;* **10:** 295-300.

Lee, D.G., Jeon, J.H., Jang, M.K., Kim, N.Y., Lee, J.H., Kim, S.J., Kim, G.D., Lee, S.H (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. *Biotechnol Lett.,* **29:** 465–472.

Leis, B1, Angelov, A., Liebl, W (2013). Screening and expression of genes from metagenomes. *Adv. Appl. Microbiol.*, **83:** 1–68.

Liebl, W. (2011). Metagenomics. In J. Reitner & V. Thiel (Eds.), *Enzyclopedia of geobiology* (pp. 553–558). Dordrecht: Springer.

Li, L., Kato, C., and Horikoshi, K. (1999). Microbial diversity in sediments collected from the deepest cold-seep area, the Japan Trench. *Marine Biotechnology* **1:** 391-400.

Li, J.L., Radajewski, S., Eshinimaev, B.T., Trotsenko, Yu.A., McDonald, I.R., Murell, J.C (2004). Molecular diversity of methanotrophs in Transbaikal soda lake sediments and identification of potentially active populations by stable isotope probing. Environ Microbiol 6:1049–1060.

Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24:** 713–714.

LI, X., BHASKAR, R., YANG, H., WANG, D. and MIAO, Y. (2009). Screening and identification of new isolate: Thermostable Escherichia coli with novel thermoalkalotolerant cellulases. *Current Microbiology*, **59 (4):** 393-399.

Liang, Y., Nostrand J. D. V., Deng, Y., He, Z., Wu, L., Zhang, X., Li, G and Zhou J. (2011). Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *The ISME Journal,* **5:** 403–413.

Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., & Pop, M. (2011). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics*, *12*(Suppl. 2), S4.

Lin, J.L, Radajewski, S., Eshinimaev, B.T., Trotsenko, Y.A., McDonald, I.R., Murrell, J.C. (2004). Molecular diversit.y of methanotrophs in Transbaikal soda lake sediments and identification of potentially active populations by stable isotope probing. *Environ Microbiol*. **6:** 1049–1060.

Lin, J.L., Joye, S.B., Scholten, J.C., Schafer, H., McDonald, I.R., Murrell, J.C. (2005). Analysis of methane monooxygenase genes in Mono lake suggests that increased methane oxidation activity may correlate with a change in methanotroph community structure. *Appl Environ Microbiol*., **71:** 6458–6462.

Litchfield, C.D., Gillevet, P.M. (2002). Microbial diversity and complexity in hypersaline environments: A preliminary assessment. *J. Ind. Microbiol. Biotechnol.*, **28:** 48-55.

Litthauer, D., Abbai, N.S., Piater, L.A., van Heereden, E (2010). Pitfalls using tribuytrin agar screening to detect lipolytic activities in metagenomic studies. *Afr. J. Biotechnol*., **9 (27):** 4282-4285.

Liu, W.T., Marsh, T.L., Cheng, H., Forney, L.J (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied Environmental Microbiology,* **63(11):** 4516–4522.

Logares, R., Haverkamp, T.H., Kumar, S., Lanzen, A., Nederbragt, A.J., Quince, C., and Kauserud, H (2012). Environmental microbiology through the lens of high throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches. *Journal of microbiological methods,* **91**: 106-113.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., de Vargas, C., Lima-Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels-Lewis, S., Karsenti, E., Bork, P., Acinas, S. G (2013). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology*, doi: 10.1111/1462-2920.12250.

Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr. Opin. Biotechnol*., 1**3:** 572–577.

Lopez-Garcia, P., Kazmierczak, J., Benzerara, K., Kempe, S., Guyot, F., Moreira, D (2005) Bacterial diversity and carbonate precipitation in the giant microbialites from the highly alkaline Lake Van, Turkey. *Extremophiles*, **9:**263–274.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*., **25:** 955–964.

Ma, Y., Xue, Y, Grant, W.D., Collins, NC, Duckworth, AW, van Steenbergen, R.S, Jones, B.E (2004a). Alkalimonas amylolytica gen. nov., sp. nov., and Alkalimonas delamerensis gen. nov., sp. nov., novel alkalophilic bacteria from soda lakes in China and East Africa. *Extremophile*s **8:**193–200.

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

MacLean, D., Jones, J.D.G., Studholme, D.J (2009). Application of ‗next-generation' sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, **7:** 287–296.

Mamo, G, Hatti-Kaul, R., and Mattiasson, B (2006). A thermostable alkaline active endo-beta-1-4-xylanase from Bacillus halodurans S7: Purification and characterization. *Enzyme and Microbial Technology*, **39 (7):** 1492 – 1498.

Mamo, G. and Gessesse, A. (2000). Immobilization of alkaliphilic Bacillus sp. cells for xylanase production using batch and continuous culture. *Appl Biochem Biotechnol.,* **87:** 95-101.

Marine, R., Polson, S. W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., et al. (2011). Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and Environmental Microbiology*, **77**: 8071–8079.

Martins, R. F., Davids, W., Abu Al-Soud, W., Levander, F., Radstrom, P. and Hatti-Kaul, R. (2001). Starch-hydrolyzing bacteria from Ethiopian soda lakes. *Extremophiles,* **5:** 135-44.

Markowitz, V. M., Chen, I.-M. A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., et al. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research*, **40:** D123–D129.

Mesbah, N.M., Hedrich, D.B., Peacock, A.D., Rohde, M., Wiegel, J (2007). Natranaerobius themophilus gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi an Natrun, Egypt, and proposal of Natranaerobiaceae fam. Nov. and Natranaerobiales ord. nov. Int J Syst Evol *Microbiol*., **57:** 2507–2512.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **., 11(1):** 31-46.

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., et al. (2003) GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res*., 31**:** 2187–2195.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics***, 9:** 386.

Michael, T.; Madigan, M.; Orent, A (1999). Thermophilic and halophilic extremophiles. *Curr. Opin. Microbiol*., **2:** 265-269.

Milford, A. D., Achenbach, L. A., Jung, D. O. and Madigan, M. T. (2000). Rhodobaca bogoriensis gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes. *Arch Microbiol*., **174:** 18–27.

Miller, D.N., Bryant, J.E., Madsen, E.L., Ghiorse, W.C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol*., **65**: 4715–4724.

Miller, J.H. (1972). Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1972) pp. 431–435.

Minig, M., Walker, D. and Ledesma, P. (2009). Bacterial Isolates from Ethiopian Soda Lakes Producers of Alaline-Active beta-Glucanases Resistant to Chelating and Surfactant Compounds. *Research Journal of Microbiology,* **48:** 194-201.

Minoche, A.E., Dohm, J.C., and Himmelbauer, H (2011). Evaluation of genomic highthroughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome biology,* **12**: R112.

Monzoorul Haque, M., Ghosh, T. S., Komanduri, D., & Mande, S. S. (2009). SOrt-ITEMS: sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics*, **25:** 1722–1730.

Moreira, D (1998). Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucl Acids Res*; **26**:3309–10.

Morimoto, S. and Fujii, T. (2009). A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Applied Microbiology and Biotechnology*, **83:** 389–396.

Morris, L. S., Evans J., Marchesi J. R (2012). A robust plate assay for detection of extracellular microbial protease activity in metagenomic screens and pure cultures. *Journal of Microbiological Methods*, **9 (1):** 144-146.

Moussard, H., Smith, T.J., Murrell, J.C. (2011). DNA-stable isotope probing and gene mining. In: Murrell JC, Whiteley A (eds). Stable Isotope Probing and Related Technologies. ASM press: Washington, DC, 3–24.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al (2005) InterPro, progress and status in 2005. Nucleic Acids Res 33: D201–D205.

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

Mwichia, R., Cousin, S., Muigai, A.W., Boga, H.I., Stackebrandt, E (2010). Archaeal diversity in the haloalkaline Lake Elmenteita in Kenya. Curr Microbiol., **60:**47–52.

Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. J Microbiol Biotechnol., **22(4):**462–469. doi: 10.4014/jmb.1108.08037.

Nakai, R.; Abe, T.; Takeyama, H.; Naganuma, T (2011**).** Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. *Mar. Biotechnol.*, **13:** 900–908.

Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research*, **40:** e155.

Narasingarao, P., Podell, S., Ugalde, J.A (2011). De novo meta-genomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbialcommunities. *The ISME Journal*. doi:10.1038/ismej.2011.78.

Neef, A., Witzenberger R., and Kampfer, P (1999). Detection of sphingomonads and *in situ*identification in activated sludge using 16S rRNA-targeted oligonucleotide probes. *J. Ind. Microbiol. Biotechnol.* **23:** 261-267.

Neveu, J., Regeard, C. and DuBow, M.S (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and death Valley deserts. *Appl Microbiol Biotechnol* ., **91:** 634–644.

Niehaus F., Gabor E., Wieland S., Siegert P. Maurer K. H. and Eck J. (2011). Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases; Microbial Biotechnology, **4(6):** 767–776.

Nieto, J.J and Vargas, C (**2002**). Synthesis of osmoprotectants by moderately halophilic bacteria: Genetic and applied aspects. *Recent. Res. Devel. Microbiol*., **6:** 403-418.

Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., & Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *Journal of Microbiology and Biotechnology*, **22:** 462–469.

Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., & Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *Journal of Microbiology and Biotechnology*, **22:** 462–469.

Noguchi, H., Park, J., & Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, **34:** 5623–5630.

Noguchi, H., Taniguchi, T., & Itoh, T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Research*, **15:** 387–396.

Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res*, **34:** 5623–5630.

Nyyssönen, M., Tran, H. M., Karaoz, U., Weihe, C., Hadi, M. Z., Martiny J. B. H., Martiny, A. C and Brodie, E. L (2013). Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. *frontiers in Microbiology*, **4:** 282, doi: 10.3389/fmicb.2013.00282.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., Stahl, D.A (1986). Microbial ecology and evolution – a ribosomal-RNA approach. *Annual Review of Microbiology*, **40:** 337–365.

Oren, A (1999). Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev.*, **63:** 334–348.

Oren, A (2004). Adaptation of halophilic archaea to life at high salt concentrations. In *Salinity: Environment—Plants—Molecules*; Lauchli, A., Luttge, U., Eds.; Springer: Dordrecht, The Netherlands.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, **33:** 5691–5702.

Oxley, A.P.A., Lanfranconi, M.P., Wurdemann, D (2010). Halophilic archaea in the human intestinal mucosa. *Environ-mental Microbiology*, **12:** 2398–2410.

Pace, N. R (1997). A molecular view of microbial diversity and the biosphere. *Science.*, **276:** 734-740.

Pagaling, E., Wang,H., Venables, M., Wallace, A., Grant, W.D., Cowan, D.A, Jones, B.A., Ma, Y., Ventosa, A., Heaphy, S (2009). Microbial biogeography of six salt lakes in Inner Mongolia and one salt lake in Argentina. *Appl Environ Microbiol.*, **75:**5750–5760.

Pailin, T., Kang, D.H., Schmidt, K. and Fung, D.Y.C. (2001) Detection of extracellular bound proteinase in EPS-producing lactic acid bacteria cultures on skim milk agar. *Lett Appl Microbiol.,* **33**: 45–49.

Parameswaran, P., Jalili R., Tao L., Shokralla S., Gharizadeh B., Ronaghi M. and Fire A. Z. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research*, **35(19):** e130: 131-139.

Park, S.J., Kang, C.H., Chae, J.C and Rhee S.K (2008). Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS Microbiol Lett*, **284:** 28–34.

Pastor, J.M., Salvador, M., Argandona, M., Bernal, V., Reina- Bueno M., Csonka LN et al. (2010). Ectoines in cell stress protection: uses and biotechnological production. *Biotechnol Adv* ., **28:** 782–801.

Pawaraya, R., Zambare, V., Barve, S., Paratkar, G (2009). Application of Protease isolated from *Bacillus sp.*158 in enzymatic cleansing of contact lenses. *Biotechnology*, **8:** 267-280.

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2011). Meta-IDBA: a de novo assembler for metagenomic data. *Bioinformatics***, 27:** i94–i101.

Pedersen, N.R., Wimmer, R., Matthiesen, R., Pedersen, L.H. and Gessesse, A. (2003). Synthesis of sucrose laurate using a new alkaline protease. Tetrahedron: *Asymmetry*, **14:** 667–673.

Pérez-Pantoja D., Donoso R., Agulló L., Córdova M., Seeger M., Pieper D. H. and González B. (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. *Environmental Microbiology*, **14 (5):** 1091–1117.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**: 3094–3101.

Pikuta, E.V., Zhilina, T.N., Zavarzin, G.A., Kostrikina, N.A, Osipov, G.A., Rainey, F.A (1998). *Desulfonatronum lacustre* gen. nov., sp. nov. A new alkaliphilic sulfate- reducing bacterium utilizing ethanol. *Microbiology*, **67:**105–113.

Polz, M. F. and Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, **64:** 3724–3730.

Poster, A., Lohmayer R., Voget, C., Knocller, K., Planer, F.B., Sorokin, D., Richnow, H.H., and Finster, K (2013). Disproportionation of elemental sulfur by haloalkaliphilic bacteria from soda lakes. *Extremophiles*, **17:** 1003-1012.

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

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Research*, **40:** D290–D301.

Quast, C. (2006) MicHanThi – design and implementation of a system for the prediction of gene functions in genome annotation projects. Master Thesis. Bremen, Germany: University of Bremen.

Rainey, F. A., Ward-Rainey, N. L., Janssen P. H., Hippe H., Stackebrandt E. (1996). *Clostridium paradoxum* DSM 7308^T contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology,* **142**: 2087–2095.

Rainey, F. A and Oren, A (2006). Methods in Microbiology: *Extremophiles* (Vol. 35). Sandiego, CA: Elsevier.

Rampelotto, P.H. (**2009).** The search for life on other planets: Sulfur-based, silicon-based, ammonia-based life. *J. Cosmol.*, **1:** 818-827.

Rao, M.B., Tanksale, A.M., Ghatge, M.S., and Deshpande,V.V. (1998). Molecular and biotechnological aspects of microbial proteases. Microbiol. Molec. Rev., 62: 597-635.

Rascovan, N., Carbonetto, B., Revale, S., Reinert, M., Alvarez, R., et al. (2013). The PAMPA datasets: a metagenomic survey of microbial communities in Argentinean pampean soils. *Microbiome*, **1:** 21.

Rees, H.C., Grant, S., Jones, B., Grant, W.D., Heaphy, S (2003). Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles,* **7:** 415-421.

Rees, H.C., Grant, W.D., Jones, B.E., Heaphy, S (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles*, **8:**63–71.

Reysenbach, A., L.J. Giver, G.S. Wickham, and N.R. Pace 1992 Differential Amplification or rRNA Genes by Polymerase Chain Reaction. Applied Environ. Microbiol., 58:3417-3418.

Rhee, J. K, Ahn D. G., Kim, Y.G., Oh, J.W (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol,* **71:** 817–825.

Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research*, **38:** e191.

Richter, M., Lombardot, T., Kostadinov, I., Kottmann, R., Duhaime, M.B., Peplies, J., and Glöckner, F.O. (2008) JCoast – a biologist-centric software tool for data mining and comparison of prokaryotic metagenomes. *BMC Bioinformatics*, **9:** 177.

Riesenfeld, C.S., Schloss, P.D., Handelsman, J (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*., **38:** 525-52.

Rodrigues, D.F., Jesus, E.d.C., Ayala-del-Rio, H.L., Pellizari, V.H., Gilichinski, D., Sepulveda-Torres, L., Tiedje, J.M., 2009. Biogeography of two cold adapted genera Psychrobacter and Exiguobacterium. *ISME J*., **3:** 658e665.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Rothschild, L.J and Mancinelli, R.L. (**2001).** Life in extreme environments. *Nature*, **409:** 1092-1101.

Roux, K.H. (1995). Optimization and troubleshooting in PCR. In *PCR Primer: a Laboratory Manual*. Dieffenbach, C.W., and Dveksler, G.S. (eds). Plainview, NY: Cold Spring Harbor Laboratory Press, pp. 53–62.

Rusch, D.B., Halpern, A.L., Heidelberg, K.B., Sutton, G., Williamson, S.J., Yooseph, S. *et al*. (2007). The sorcerer II global ocean sampling expedition: I, the northwest Atlantic through the eastern tropical Pacific. *PLoS Biol* **5**: e77.

Russell, N.J., Evans, R.I., ter Steeg, P.F., Hellemons, J., Verheul, A., Abee, T (1995). Membranes as a target for stress adaptation. *Int. J. Food Microbiol*., **28:** 255–261.

Sambrook and Russell (2001). *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor Laboratory Press. ISBN 978-0-87969-577-4.

Sanchez, S., Demain, A.L (2011). Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Org. Process Res. Dev*., **15:** 224–230.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., et al. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, **265:** 687–695.

Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., et al. (2012). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **40:** D13–D25.

Schauer, M., Jiang, J., and Hahn, M. (2006). Recurrent Seasonal Variations in Abundance and Composition of Filamentous SOL Cluster Bacteria (Saprospiraceae, Bacteroidetes) in Oligomesotrophic Lake Mondsee (Austria*). Applied and Environmental Microbiology*, **72(7):** 4704-4712.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., et al. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*, **39:** D670–D676.

Schmidt, T.M., DeLong, E.F., Pace, N.R (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriology,* **173:** 4371–4378.

Schloss PD, Handelsman J (2003). Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* **14:** 303-310Scholten, J.C.M., S.B. Joye, J.T. Hollibaugh & J.C. Murrell (2005).

Scholten, J.C.M., Joye, S.B., Hollibaugh, J.T., and Murrwll, J.C (2005). Molecular analysis of the sulfate reducing and archaeal community in a meromictic soda lake (Mono Lake, California) by targeting 16S rRNA, *mcrA*, *apsA*, and *dsrAB* genes. *Microbial Ecology*, **50:** 29–39.

Schopf, J. W., Kudryavtsev, A. B., Agresti, D. G., Wdowiak, T. J., and Czaja, A. D (2002). Laser–raman imagery of earth‘s earliest fossils. *Nature*, **416 (6876):** 73–76.

Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., et al. (2007). TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Research*, **35:** D260–D264.

Senesi, N., Loffredo, E (2001). Soil humic substances. In: Hofrichter M, Steinbuchel A, editors. Biopolymers. Lignin, humic substances and coal, *Wiley-VCH,* **1**: 247–99.

Sharma, S., Sharma, K. K. and Kuhad, R.C (2014). An efficient and economical method for extraction of DNA amenable to biotechnological manipulations, from diverse soils and sediments. *Journal of Applied Microbiology*, **116 (3):** 477–746.

Sharma, P.K., Capalash, N. and Kaur, J (2007. An improved method for single step purification of metagenomic DNA. Molecular Biotechnology, **36 (1):** 61-63.

Sheridan, C. (2004). Kenyan dispute illuminates bioprospecting difficulties. *Nat Biotechnol*., **22:** 1337.

Shi, Y., McCarren, J., Delong, E.F (2012). Transcriptional responses of surface water marine microbial assemblages to deep-sea water amendment. *Environ. Microbiol*., **14:** 191–206.

Simonen, M and Palva, I (1993). Protein secretion in *Bacillus* species, *Microbiol. Rev*., **57:** 109–137.

Sinha, R., Joshi, R., Dodia, M.S., Singh, S.P. (2007). Production, purification and characterization of an alkaline protease from an alkalophilic *Bacillus* sp. *Journal of Cell and Tissue Research*, **7:** 1031-1037.

Skerman, V. B. D. (1969). Starch hydrolysis detection method, p. 808. In V. B. D. Skerman (ed.), Abstracts of microbiological methods. Wiley-Interscience, New York.

Schippers, A., Kock, D., Höft, C., Köweker, G., and Siegert, M. (2012). Quantification of microbial communities in subsurface marine sediments of the black sea and off Namibia. *Frontiers in Microbiology*, **3:** 16.

Schloss P.D. and Handelsman J (2003). Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology,* **14:**303–310.

Shokralla, S., Spall, J. L., Gibson, J. F., and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21:** 1794–1805.

Smibert, R. M. and Krieg N. R. (1994). Phenotypic characterization. In *Methods for General and Molecular Bacteriology*, pp. 607–654. Edited by Gerhardt P., Murray R. G. E., Wood W. A., Krieg N. R. . Washington, DC: American Society for Microbiology.

Sogin, M. L., Morrison,H. G., 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." Proc. Natl. Acad. Sci. USA 103:12115–12120.

Sorokin, D.Y., Jones, B.E., Kuenen, J.G. (2000a). A novel obligately methylotrophic, methane-oxidizing Methylomicrobium species from a highly alkaline environment. Extremophiles, **4:** 145–155.

Sorokin, D.Y., Robertson, L.A., Kuenen, J.G. (2000b). Isolation and characterization of obligately chemolithoautotrophic alkaliphilic sulfur-oxidizing bacteria. Ant V Leeuwenhoek, **77:** 251–260.

Sorokin, D.Y., Tourova, T.P., Lysenko, A.M., Kuenen, J.G (2001a) . Microbial thiocyanate utilization under highly alkaline conditions. Appl Environ Microbiol., **67:**528–538.

Sorokin, D.Yu., Tourova, T.P., LysenkoAM MLL, Kuenen J.G (2002). Thioalkalivibrio thiocyanooxidans sp. nov. and Thioalkalivibrio paradoxus sp. nov., novel alkaliphilic, obligately autotrophic, sulfuroxidizing bacteria from the soda lakes able to grow with thiocyanate. *Int J Syst Evol Microbiol.,* **52:**657–664.

Sorokin, D.Y., Antipov, A.N., Kuenen, J.G. (2003). Complete denitrification in coculture of obligately chemolithoautotrophic haloalkaliphilic sulfur-oxidizing bacteria from a hypersaline soda lake. Arch Microbiol., **180:** 127–133.

Sorokin, D.Yu Gorlenko, V.M., Namsaraev, B.B., Namsaraev, Z.B., Lysenko, A.M., Eshinimaev, B.T., Khmelenina, V.N., Trotsenko, Y.A., Kuenen, J.G (2004). Prokaryotic communities of the north-eastern Mongolian soda lakes. *Hydrobiologia,* **522:**235–248.

Sorokin, D.Y. and Kuenen, J.G (2005b). Haloalkaliphilic sulphur-oxidizing bacteria in soda lakes. FEMS Microbiol Rev., **29**:685–702.

Sorokin, D.Y., Tourova, T.P., Henstra, A.M., Stams, A.J.M., Galinski, E.A., Muyzer, G (2008a). Sulfidogenesis under extremely haloalkaline conditions by *Desulfonatronospira thiodismutans* gen. nov., sp. nov., and *Desulfonatronospira delicata* sp. nov.– a novel lineage of *Deltaproteobacteria* from hypersaline soda lakes. *Microbiology,* **154:**1444–1453

Sorokin, D.Y., Tourova, T.P., Mussmann, M., Muyzer, G (2008b). *Dethiobacter alkaliphilus* gen. nov sp nov., and *Desulfurivibrio alkaliphilus* gen. nov sp nov.: two novel representatives of reductive sulfur cycle from soda lakes. *Extremophiles*, **12:**431–439.

Sorokin, D. Y., Detkova, E. N., and Muyzer, G (2010b). Propionate and butyrate dependent bacterial sulfate reduction at extremely haloalkaline conditions and description of *Desulfobotulus alka-liphilus* sp. nov. *Extremophiles,* **14:** 71–77.

Sorokin, D.Y., Kuenen, J.G., and Muyzer G (2011). The microbial sulfur cycle in soda lakes. *Front Microbial Physiol.,* **2:** 44. DOI: 10.3389/fmicb.2011.00044.

Sorokin, D. Y., Tourova, T. P., Panteleeva, A. N., Kaparullina, E. N., Muyzer, G. (2012). Anaerobic utilization of pectinous substrates at extremely haloalkaline conditions by Natranaerovirga pectinivora gen. nov., sp. nov., and Natranaerovirga hydrolytica sp. nov., isolated from hypersaline soda lakes. *Extremophiles*, **16:**307–315. DOI 10.1007/s00792-012-0431-6.

Stach J.E., Maldonado L.A., Masson D.G., Ward A.C., Goodfellow M. and Bull A.T. (2003a). Statistical approaches to estimating bacterial diversity in marine sediments. *Appl.Env. Microbiol.* **69:** 6189–6200.

Stahl, D.A., Lane, D.J., Olsen, G.J., and Pace, N.R. (1985). Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl Environ Microbiol*. **49:**1379-1384.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and DeLong, E.F (1996). Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment front a planktonic marine archaeon. *Journal of Bacteriology*. **178:**591-599.

Staley, J.T and Konopka, A (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*, **39:** 32-346.

Steele, H.L., Jaeger, K.E., Daniel, R. Streit, W.R (2009). Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol.*, **16**: 25–37.

Stetter, K.O. (2002). Hyperthermophilic microorganisms. In Astrobiology. The Quest for the Conditions of Life, G. Horneck, and C. Baumstark-Khan, eds. (Berlin: Springer), pp. 169 - 184.

Stewart, F.J., Ottesen E.A., DeLong E.F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *Isme J.,* **4:** 896-907.

Streit, W.R., Daniel, R., Jaeger, K.E (2004). Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Curr Opin Biotechnol*., **15:** 285–290.

Streit, W.R., Schmitz, R.A (2004). Metagenomics-the key to the uncultured microbes. *Curr Opin Microbiol.,* **7:** 492-8.

Sukarasi, V.P., Wani, A., Shouche, Y.S., Ranade, D.R (2007). Phylogenetic analysis of methanogenic enrichment cultures obtained from Lonar Lake in India: isolation of Methanocalculus sp. and Methanoculleus sp. *Microb Ecol*., **54:** 697–704

Tadesse, S., Milesi, J. & Deschamps, Y. (2003). Geology and mineral potential of Ethiopia: a note on geology and mineral map of Ethiopia. *Journal of African Earth Sciences*, **36:** 273–313.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., Horikoshi, K. (2008). Cell proliferation at 122 °C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Nat. Acad. Sci. USA,* **105:** 10949-10954.

Tanasupawat, S., Thongsanit, J., Okada, S., Komagata, K (2002). "Lactic acid bacteria isolated from soy sauce mash in Thailand". *The Journal of General and Applied Microbiology* (The Microbiology Research Foundation), **48 (4):** 201–209.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., et al. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research,* **29:** 22–28.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics,* **4:** 41.

Teather, R.M. and Wood P.J. (1982). Use of congo red–polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl. Environ. Microbiol.,* **43:** 777–780.

Tebbe, C.C., Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl. Environ. Microbio,* **59:** 2657–2665.

Tekedar, H. C., & Sanlı-Mohamed, G. (2011). Molecular cloning, over expression and characterization of thermoalkalophilic esterases isolated from *Geobacillus* sp. *Extremophiles*, 1**5***:* 203–211.

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics—a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, **2:** 3.

Tien, C.C., Chao, C.C. and Chao, W.L (1999). Methods for DNA extraction from various soils: a comparison. *J Appl. Microbiol*., **86:** 937–943.

Tilahun, G. & Ahlgren, G. (2010). Seasonal variations in phytoplankton biomass and primary production in the Ethiopian Rift Valley lakes Ziway, Awassa and Chamo– The basis for fish production. *Limnologica*.

Tindall, B.J., Ross, H.N.M., Grant, W.D (1984). Natronobacterium gen. nov. and Natronococcusgen. nov., two new genera of haloalkaliphilic archaebacteria. Syst Appl Microbiol., **5:**41–57.

Tindall, B.J., Rosselló-Móra, R., Busse, H.-J., Ludwig W., Kämpfer P. **(**2010). Notes on the characterization of prokaryote strains for taxonomic purposes.Int J Syst Evol Microbiol **60:** 249–266.

Tiwari, S., Singh,P., Tiwari, R. *et al*. (2011). ―Salt-tolerant rhizobacteria-mediated induced tolerance in wheat (Triticum aestivum) and chemical diversity in rhizosphere enhance plant growth," *Biology and Fertility of Soils*,. 47(8): 907–916, 2011.

Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S., & Pop, M. (2011). Next generation sequence assembly with AMOS. *Current Protocols in Bioinformatic*, *11*. (Unit 11.8).

Tringe, S.G. and Hugenholtz P. (2008). A renaissance for the pioneering 16S rRNA gene.*Current Opinion in Microbiology,* **11(5):** 442-446.

Trotsenko, Y.A and Khmelenina, V.N. (2002). Biology of extremophilic and extremotolerant methanotrophs. Arch Microbiol, **177:** 123–131.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428:** 37–43.

Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, **23:**88–93.

Uchiyama, T. and Miyazaki, K. (2010). Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology*, **76:** 7029–7035.

UniProt Consortium. (2010). The Universal protein resource (UniProt) in 2010. *Nucleic Acids Research*, **38:** D142–D148.

Uroz, S., Ioannidis, P., Lengelle, J., Cébron, A., Morin, E., *et al*. (2013) Functional Assays and Metagenomic Analyses Reveals Differences between the Microbial Communities Inhabiting the Soil Horizons of a Norway Spruce Plantation. *PLosONE*, **8(2):** e55929. doi:10.1371/journal.pone.0055929.

Van de Peer, Y., Chapelle, S., De Wachter, R (1996). A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res.,* 24**:** 3381-3391.

Vaz-Moreira, I., Egas, C., Nunes, O. C. and Manaia, C. M. (2013). Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA-DGGE and culture-dependent methods. *FEMS Microbiology Ecology*, **8*3*:** 361–374.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304:** 66–74.

Verma, D., Satynarayana, T (2011). An improved protocol for DNA extraction from alkaline soil and sediment samples for constructing metagenomic libraries. *Appl Biochem Biotechnol* **165:** 454–464. doi: 10.1007/s12010-011-9264-5

Vergin, K. L., Urbach, E., Stein, J. L., DeLong, E. F., Lanoil, B. D. and Giovannoni, S. J. (1998). Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl Environ Microbiol*, **64:** 3075–3078

Vilchez-Vargas, R., Geffers, R., Suárez-Diez, M., Conte, I., Waliczek, A., Kaser, V. S., *et al*. (2012). Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel internally calibrated microarray system. *Environmental Microbiology,* 14 (5): 1016–1039.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E., Streit, W.R (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol,* **69:** 6235-6242.

Voget, S, Steele, H.L., Streit, W.R (2006). Characterization of a metagenome-derived halotolerant cellulase. *J Biotechnol*, **126:** 26-36.

Wani, A.A., Surakasi, V.P., Siddharth, J., Raghavan, R.G., Patole, M.S., Ranade, D., Shouche, Y.S. (2006). Molecular analyses of microbial diversity associated with the Lonar soda lake in India: an impact crater in a basalt area. *Res Microbiol*, **157:** 928–937.

Wagner, M., Amann, R., Lemmer, H., Schleifer, K (1993). Probing activated sludge with oligonucleotides specific for proteobacteria: Inadequacy of culture-dependent methods for describing microbial community structure. *Applied Environmental Microbiology,* **59:** 1520–1525.

Wang, G., Meng, K., Luo, H., Wang, Y., Huang, H., Shi, P., et al. (2012). Phylogenetic diversity and environment-specific distributions of glycosyl hydrolase family 10 xylanases in geographically distant soils. *PLoS ONE*, **7:** e43480.

Wang, Q, G., Garrity, M., Tiedje, J. M., and Cole, J. R (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol*., **73(16):**5261-7.

Wang, Y., Hayatsu, M. and Fujii, T. (2012). Extraction of bacterial RNA from soil: challenges and solutions. *Microbes and the Environment*, **27:** 111–121.

Ward, D.M., Weller, R., Bateson, M.M (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature,* **345:** 63–65.

Ward, B.B., D.P. Martino, M.C. Diaz & S.B. Joye (2000). Analysis of ammonia-oxidizing bacteria from hypersaline Mono Lake, California, on the basis of 16S rRNA sequences. *Applied and Environmental Microbiology*, **66:** 2873–2881.

Waschkowitz, T., Rockstroh, S. and Daniel, R (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol.,* **75:** 2506–2516.

Werner, J.J., Zhou, D., Caporaso, J.G., Knight, R., and Angenent, L.T (2012). Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J,* **6:** 1273-1276.

Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci*. USA, 95: 6578–6583.

Williamson, L. L., Borlee, B. R., Schloss, P. D., Guan, C., Allen, H. K., & Handelsman, J. (2005). Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Applied and Environmental Microbiology*, ***71*:** 6335–6344.

Williamson, S.J., Rusch, D.B, Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass JI *et al*. (2008). The sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3**: e1456.

Woese, C. R (1987). Bacterial evolution. *Microbiol Rev.*, **51:** 221-271.

Wolanin, P.M., Thomason, P.A., Stock, J.B. (2002). Histidine protein kinases: Key signal transducers outside the animal kingdom. *Genome Biol.* **3:** doi:10.1186/gb-2002-3-10-reviews3013.

Wooley, J. C., and Ye, Y. (2009). Metagenomics: facts and artifacts, and computational challenges. *Journal of Computer Science and Technology*, **25:** 71–81.

Wooley, J. C., Godzik, A. and Friedberg, I (2010). A primer on metagenomics. *PLoS Comput Biol*, **6(2):** e1000667.

Wu, J., Gao, W., Johnson, R. H., Zhang, W., and Meldrum, D. R. (2013). Integrated Metagenomic and Metatranscriptomic Analyses of Microbial Communities in the Meso- and Bathypelagic Realm of North Pacific Ocean. *Mar. Drugs*, **11:** 3777-3801.

Xia, S.Q., Duan, L.A., Song, Y.H., Li, J.X., Piceno, Y.M., et al. (2010). Bacterial community structure in geographically distributed biological wastewater treatment reactors. *Environmental Science and Technology,* **44:** 7391–7396.

Young, C.C., Burghoff, R.L., Keim, L.G.,Minak-Bernero, V., Lute, J.R., Hinton, S.M (1993). Polyvinylpyrrolidone- agarose gel electrophoresis purification of polymerase chain reaction-amplifiable DNA from soils. *Appl. Environ. Microbiol*., **59**: 1972–1974.

Yooseph, S., G. Sutton, D.B. Rusch, A.L. Halpern, S.J. Williamson, K. Remington, J.A. Eisen, K.B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C.S. Miller, H. Li, S.T. Mashiyama, M.P. Joachimiak, C. van Belle, J.M. Chandonia, D.A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B.J. Raphael, V. Bafna, R. Friedman, S.E. Brenner, A. Godzik, D. Eisenberg, J.E. Dixon, S.S. Taylor, R.L. Strausberg, M. Frazier, and J.C. Venter (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol*. **5:**e16.

Youssef, N., Steidley, B. L., and Elshahed, M. S. (2012). Novel high-rank phylogenetic lineages within a sulfur spring (Zodletone Spring, Oklahoma), revealed using a combined pyrosequencing-sanger approach. *Applied and Environmental Microbiology*, **78:** 2677–2688.

Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering*, **89(6):** 670–9. doi:10.1002/bit.20347.

Yun, J., Kang, S., Park, S., Yoon, H., Kim, M.J., Heu, S., Ryu, S (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl Environ Microbiol.,* **70:** 7229-7235.

Zavarzin, G. A., Zhilina, T. N. & Kevbrin, V. V. (1999). The alkaliphilic microbial community and its functional diversity. *Mikrobiologiya,* **68:** 579-599.

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

Zhang, T., Fang, H.H.P (2006). Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. *Applied Microbiology and Biotechnology,* **70:** 281–289.

Zhao, J., Yang, N., and Zeng, R (2008). Phylogenet ic analysis of type 1 polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediments. *Extremophlies,* **12:** 97-105.

Zhilina, T.N., Zavarzin, G.A., Rainey, F.A., Kevbrin, V.V., Kostrikina, N.A., Lysenko, A.M (1996). *Spirochaeta alkalica sp. nov*., *Spirochaeta africana sp. nov*., and *Sprirochaeta asiatica sp. nov*., alkaliphilic anaerobes from the continental soda lakes in Central Asia and East African rift. *Int J Syst Bacteriol*, **46:** 305-312.

Zhilina, T.N., Zavarzina, D.G., Kuever, J., Lysenko, A.M., Zavarzin, G.A (2005b). *Desulfonatronum cooperativum sp*. nov., a novel hydrogenotrophic, alkaliphilic, sulfate- reducing bacterium, from a syntrophic culture growing on acetate. *Int J Syst Evol Microbiol*, **55:**1001–1006.

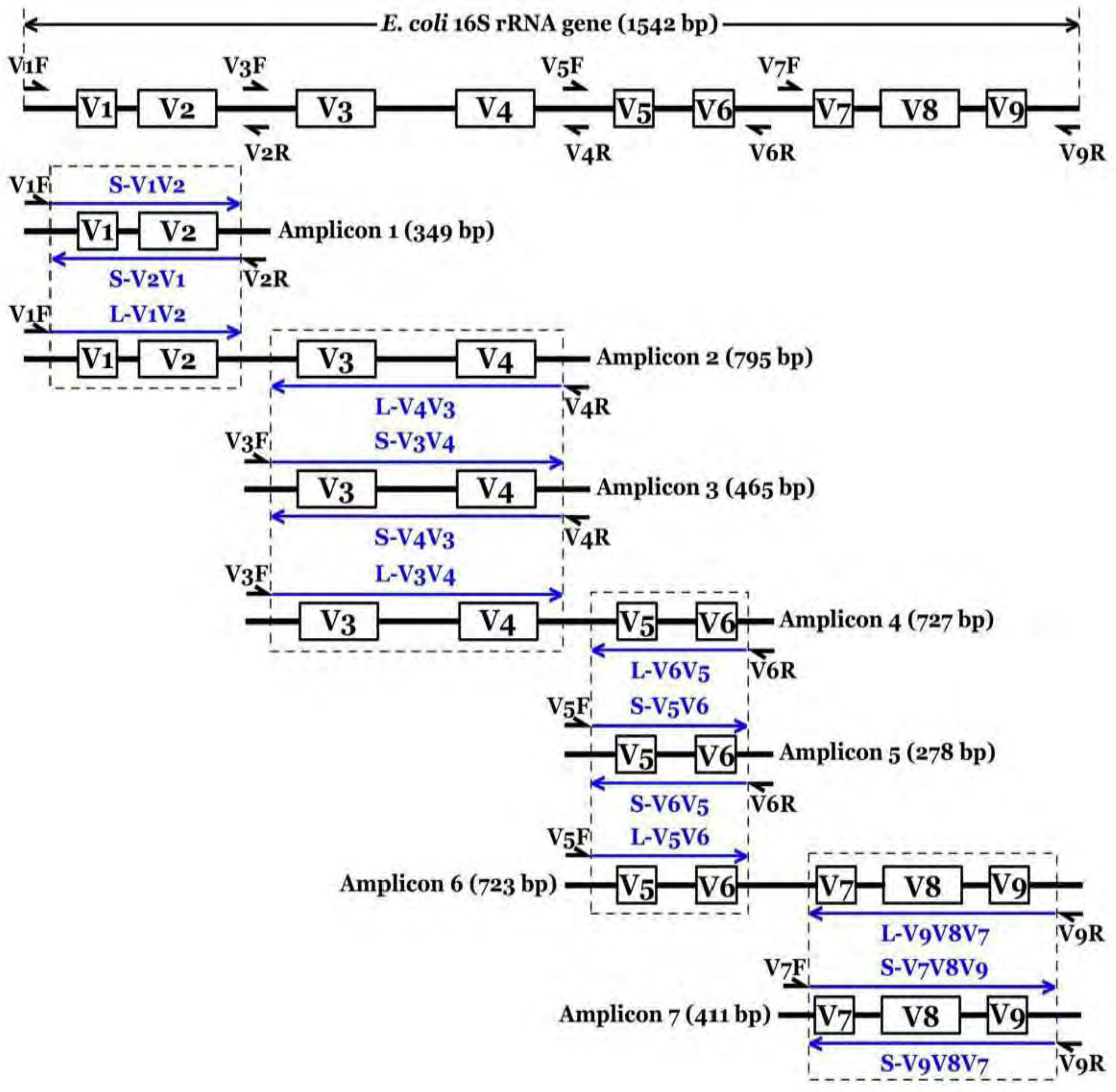
Zhou, J., Bruns, M.A., Tiedje, J.M (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol,* **62:**316–22.

Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, **38:** e132.

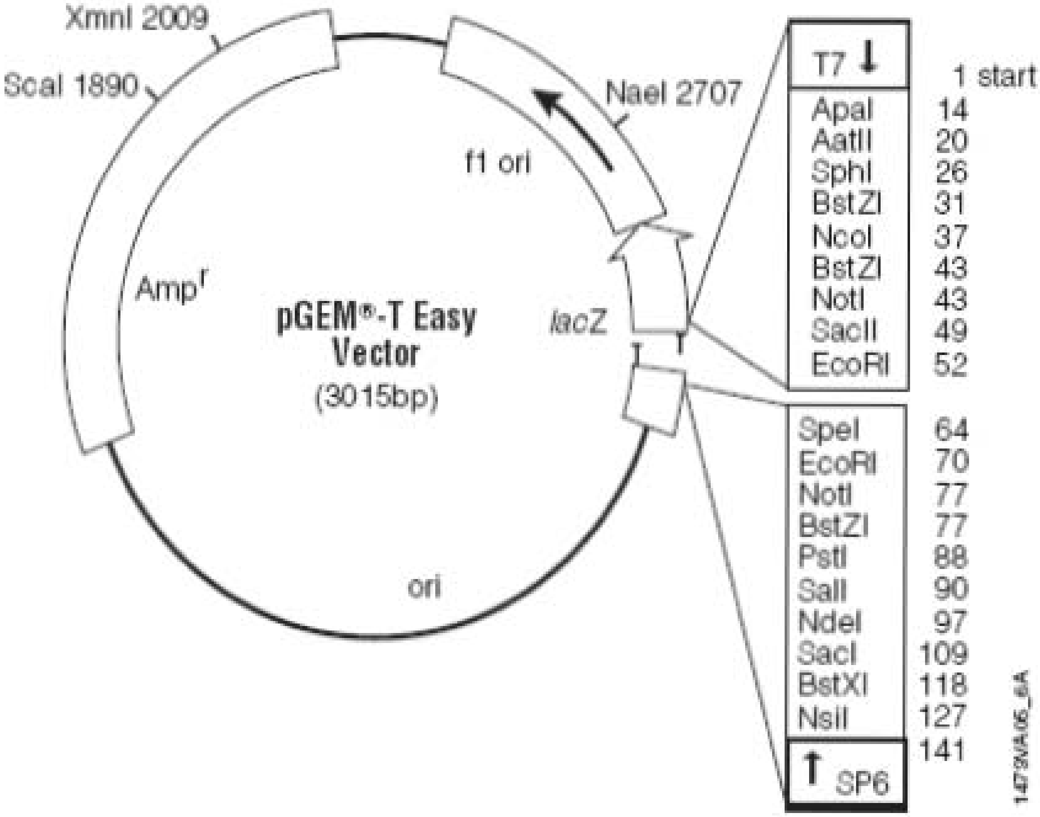
Zinger, L., Gobet, A.L., and Pommier, T (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, **21:** 1878–1896.

APPENDIX

Apendix I: The *E. coli* 16 rRNA genes with 9 variable regions



Apendix II: **The pGEM-T cloning vector and its features**



Apendix III: Table 4.3, 4.4, 4.5, and 4.6

Table 4.3: Phylogenetic affiliation of 16S rRNA sequences of actinomycete clones obtained from Shalla Lake

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group	Function/ Habitat source
LSH_A01 (HM106304)	*Rhodococcus erythropolis* (AB546303)	100%	*Actinobacteria*	Soil, Japan
LSH_A03 (HM106305)	*Uncultured actinomycete* (JX242838)	98%	*Actinobacteria*	soil, Chaina
LSH_A05 (HM106306)	*Uncultured actinobacterium* (JF916617)	99%	*Actinobacteria*	Soil, Chaina
LSH_A06 (HM106307)	Uncultured *Ilumatobacter sp.* (KC817061)	98%	*Actinobacteria*	Ballast water, USA
LSH_A8 (HM106308)	*Uncultured actinobacterium* (JF916617)	95%	*Actinobacteria*	soil, Chaina
LSH_A9 (HM106309)	Uncultured *Ferrimicrobium sp* (JX505325)	98%	*Actinobacteria*	Wetland soil, Chaina
LSH_A10 (HM1063010)	*Uncultured actinobacterium* (JF916589)	97%	*Actinobacteria*	Soil, Chaina
LSH_A12 (HM1063011)	*Uncultured actinobacterium* (DQ811922)	97%	*Actinobacteria*	Mangrove soil,
LSH_A13 (HM1063012)	*Uncultured bacterium* (GQ860294)	94%	*Actinobacteria*	Ohio, River sediments
LSH_A19 (HM1063013)	*Uncultured actinobacterium* (KC514305)	98%	*Actinobacteria*	Coral
LSH_A20 (HM1063014)	*Uncultured bacterium* (KC112930)	99%	*Actinobacteria*	Yanqi saltern
LSH_A21 (HM1063015)	*Uncultured actinobacterium* (JN037894)	96%	*Actinobacteria*	Saline-alkaline soil
LSH_A46 (HM1063016)	*Uncultured bacterium* (FJ948246)	97%	*Actinobacteria*	Lake Chaka

Table 4.4: Phylogenetic affiliation of 16S rRNA sequences of actinomycete clones obtained from Chitu Lake

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group	Function/ Habitat source
LCH_A01 (HM106292)	*Streptomyces sp*. TTH-DM-4 (KF729630)	100%	*Actinobacteria*	Soil, Chaina
LCH_A02 (HM106293)	*Streptomyces sp*. MJM164 (KF767870)	99%	*Actinobacteria*	Soil, Korea
LCH_A03 (HM106294)	*Streptomyces sp*. TTH-DM-4 (KF729630)	99%	*Actinobacteria*	Soil, Chaina
LCH_A04 (HM106295)	*Nocardioides aquaticus* (X94145)	99%	*Actinobacteria*	Hypersaline, Ekho lake, Vestfold hills, Antarctic
LCH_A08 (HM106296)	*Kitasatospora sp*. (FN550139)	98%	*Actinobacteria*	Marion Island terrestrial habitats (Sub-Antarctica)
LCH_A10 (HM106297)	*Uncultured actinobacterium* (EF447044)	96%	*Actinobacteria*	Volcanic environment, Spain
LCH_A19 (HM106298)	*Nocardioides sp*. (X90830)	99%	*Actinobacteria*	Unknown
LCH_A24 (HM106299)	*Streptomyces sp*. (KF729626)	99%	*Actinobacteria*	Soil, Chaina
LCH_A29 (HM106300)	*Uncultured actinobacterium* (JN825528)	98%	*Actinobacteria*	Alkaline lake Alchichica (Mexico)
LCH_A36 (HM106301)	*Uncultured actinomycete* (JX242838)	96%	*Actinobacteria*	Salty Beach, Tarim, Chaina
LCH_A44 (HM106302)	*Uncultured Pseudonocardia sp*. (AY571815)	98%	*Actinobacteria*	Hydrocarbon-contaminated soil, Antarctica
LCH_A55 (HM106303)	*Uncultured actinomycete* (JX507167)	98%	*Actinobacteria*	Seawater, CA

Table 4.5: Phylogenetic affiliation of 16S rRNA sequences of bacterial clones obtained from Lake Chitu

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from NCBI Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group from RDP	Function/ Habitat source
LCH_B05 (HM582691)	Uncultured low G+C Gram-positive bacterium (DQ432446)	98%	*Firmicutes/ Clostridia*	Lake Fazda water in Wadi An Natrun, Egypt
LCH_B06 (HM582692)	Uncultured candidate division OP11 bacterium (AB294914)	93%	*OD1*	Hot spring, Japan
LCH_B07 (HM582693)	*Geosporobacter subterraneus* (DQ643978)	98%	*Firmicutes/ Clostridia*	Deep subsurface aquifer, France
LCH_B10 (HM582694)	*Bacillus sp*. (FJ764770)	98%	*Firmicutes/Bacilli*	Haloalkaline Lake Elmenteita, Kenya
LCH_B12 (HM582695)	Gamma proteobacterium (FJ764787)	99%	*Gammaproteobacteria*	Haloalkaline Lake Elmenteita, Kenya
LCH_B13 (HM582696)	Uncultured bacterium (EU570894)	93%	*Firmicutes*	Nitinat Lake at a depth of 13 m, Canada
LCH_B14 (HM582697)	Uncultured *Sphingobacteria bacterium* (EU283540)	97%	*Bacteroidetes /Flavobacteria*	Anderson Lake, USA
LCH_B15 (HM582698)	Uncultured *Bacteroidetes bacterium* (DQ432348)	92%	*Bacteroidetes /Flavobacteria*	Lake Hamra Wadi An Natrun, Egypt
LCH_B16 (HM582699)	Uncultured *Cytophagales bacterium* (EU328102)		*Firmicutes/ Sphingobacteria Chitinophagaceae*	Moderate saline soil, Chaina
LCH_B101 (HM582700)	Uncultured *Firmicutes bacterium* (EU283544)	98%	*Firmicutes/ Bacilli*	Anderson Lake, USA
LCH_B102 (HM582701)	*Thioalkalivibrio paradoxus* (NR_025014)	96%	*Gammaproteobacteria Thioalkalivibrio*	Soda Lake, Kenya
LCH_B103 (HM582702)	Uncultured *Clostridiaceae* (FJ764410)	99%	*Firmicutes/ Clostridia*	Lake Elmenteita sedimen, Kenya

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from NCBI Database (Accession No.)	Similarit y (%)	Phylogenetic Affiliation/group from RDP	Function/ Habitat source
LCH_104 (HM582703)	Uncultured bacterium (GU083693)	95%	*Bacteroidetes*	Inner Mongolia, Xiarinur soda lake, Chaina
LCH_105 (HM582704)	*Alkaliphilus transvaalensis* (NR_024748)	90%	*Firmicutes /Clostridia*	Deep South African gold mine
LCH_107 (HM582705)	Uncultured bacterium (EU645094)	93%	*Firmicutes/ Bacilli*	Soap Lake, Washington
LCH_108 HM582706	Uncultured *Verrucomicrobiales* (DQ351768)	92%	*Verrucomicrobia/ Opitutae*	Heavy metal contaminated marine sediments, Belgium
LCH_110 (HM582707)	Uncultured *Bacillus sp.* (AF454301)	93%	*Firmicutes*	Alkaline, hypersaline Mono Lake, California
LCH_B111 (HM582708)	*Azoarcus toluclasticus* (AF123076)	97%	*Proteobacteria/ Betaproteobacteria Rhodocyclaceae*	Denitrifying bacteria as Aromatic degrader, USA
LCH_B112 HM582709	Uncultured bacterium (AY862794)	93%	*Firmicutes*	Reductive sulfur cycle from soda lakes, Rassia
LCH_B114 (HM582710)	*Dethiobacter alkaliphilus* (EF422412)	99%	*Firmicutes/ Clostridia/Dethiobacter*	Inner Mongolia, Xiarinur Soda Lake, Chaina
LCH_B115 (HM582711)	Uncultured *Firmicutes* (GQ848202)	97%	*Firmicutes/ Clostridia/ Tindallia*	Inner Mongolia, Xiarinur Soda Lake, Chaina
LCH_B123 (HM582712)	Uncultured bacterium (EU645255)	96%	*Firmicutes/ Clostridia*	Soap Lake, Washington
LCH_B127 (HM582713)	Uncultured *Bacteroidetes* (AF507869)	98%	*Bacteroidetes/ Flavobacteria*	Mono Lake at a depth of 35 m, California
LCH_B128 (HM582714)	*Azoarcus sp.* (GQ389714)	98%	*Betaproteobacteria /Rhodocyclaceae/Azoarcus*	Hot spring, Yang Ming Mountain, Taiwan
LCH_B129	Uncultured bacterium (EF573867)	93%	*Verrucomicrobia/ Opitutae*	Coco's Island, Costa Rica

Sample site/ Clone code (Accession No.)	**Nearest phylogenetic neighbor from NCBI Database (Accession No.)**	**Similarity (%)**	**Phylogenetic Affiliation/group from RDP**	**Function/ Habitat source**
(HM582715)				
LCH_B130 (HM582716)	Uncultured bacterium (EU645012)	95%	*Tenericutes/ Mollicutes/ Acholeplasma*	Soap Lake, Washington
LCH_B131 (HM582717)	Uncultured bacterium (HM127230)	98%	*Proteobacteria*	Chaerhan Lake, Tibet
LCH_B132 (HM582718)	Uncultured bacterium (EU358743)	95%	*Bacteroidetes*	Mesophilic biogas digester, Chaina
LCH_B134 (HM582719)	Uncultured *Sphingobacteria* (EU283540)	92%	*Bacteroidetes/ Flavobacteria*	Anderson Lake, USA
LCH_B137 (HM582720)	Uncultured bacterium (GQ979965)	99%	*Betaproteobacteria /Rhodocyclaceae/Azoarcus*	Microbial biofilm from soluble nitrate reducing bioreactor, Chaina
LCH_B139 (HM582721)	*Roseinatronobacter monicus* (DQ659236)	97%	*Alphaproteobacteria/ Rhodocyclaceae/Rhodobaca*	Hypersaline Soda Mono Lake, California
LCH_B141 (HM582722)	Uncultured *Bacteroidetes* (DQ432435)	97%	*Bacteroidetes/ Sphingobacteria*	Lake Fazda water, Wadi An Natrun, Egypt
LCH_B142 (HM582723)	*Roseinatronobacter monicus* strain (DQ659236)	97%	*Alphaproteobacteria/ Rhodocyclaceae/Rhodobaca*	hypersaline Soda Mono Lake, California
LCH_B144 (HM582724)	*Geosporobacter subterraneus* (DQ643978)	96%	*Firmicutes/ Clostridia/*	Deep subsurface aquifer, France
LCH_B146 (HM582725)	Uncultured bacterium (AB297418)	99%	*Alphaproteobacteria/ Rhodocyclaceae/Rhodobaca*	Denitrification system of saline wastewater, Japan
LCH_B147 (HM582726)	Uncultured *Bacillus sp.* (AF454301)	91%	*Firmicutes*	Alkaline, hypersaline Mono Lake, California
LCH_B149	Uncultured bacterium (HM128201)	97%	*Gammaproteobacteria*	Xiaochaidan Lake, Tibet

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from NCBI Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group from RDP	Function/ Habitat source
(HM582727)				
LCH_B150 (HM582728)	Uncultured *alpha proteobacterium* (DQ432301)	98%	*Alphaproteobacteria/ Rhodocyclaceae/ Roseinatronobacter*	Lake Hamra sediment, Wadi An Natrun, Egypt
LCH_B151 (HM582729)	*Alkalimonas collagenimarina* (NR_041515)	98%	*Gammaproteobacteria/ Alkalimonas*	Alkaliphile isolated from deep-sea sediment
LCH_B152 (HM582730)	Uncultured *Desulfuromusa sp.* (EU283459)	98%	*Deltaproteobacteria/ Desulfuromusa*	Anderson Lake, USA
LCH_B153 (HM582731)	Uncultured bacterium (GU133253)	96%	*Actinobacteria*	toluene degraders in tar-oil-contaminated aquifer sediment, Germany
LCH_B154 (HM582732)	Uncultured *Bacteroidetes* (DQ432447)	92%	*Bacteroidetes/ Flavobacteria*	Mono Lake at a depth of 35 m,
LCH_B155 (HM582733)	Uncultured *Bacteroidetes* (AF507873)	92%	*Bacteroidetes/Sphingobacteria Chitinophagaceae*	Mono Lake at a depth of 23 m, California
LCH_B156 (HM582734)	Uncultured gamma proteobacterium (GQ848209)	99%	*Gammaproteobacteria/Chromatiales/ Thioalkalivibrio*	Inner Mongolia, Xiarinur Soda Lake, Chaina
LCH_B157 (HM582735)	Uncultured gamma *proteobacterium* (DQ432152)	97%	*Gammaproteobacteria/Chromatiales/ Thioalkalivibrio*	Hypersaline Lake Wadi An Natrun, Egypt
LCH_B169 (HM582736)	Uncultured Gram-positive bacterium (DQ302446)	99%	*Firmicutes/Clostridia*	Lonar Soda Lake, India
LCH_B170 (HM582737)	*Nitrincola sp.* (FJ764761)	98%	*Gammaproteobacteria/ Oceanospirillaceae/ Nitrincola*	Haloalkaline Lake Elmenteita, Kenya
LCH_B173 (HM582738)	Gamma proteobacterium (FJ764787)	99%	*Gammaproteobacteria*	Haloalkaline Lake Elmenteita, Kenya

Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from NCBI Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group from RDP	Function/ Habitat source
LCH_B175 (HM582739)	*Thioalkalivibrio sp.* (EU709875)	97%	*Gammaproteobacteria /Thioalkalivibrio*	Soda lake sediment
LCH_B176 (HM582740)	Uncultured bacterium (EU644999)	93%	*Bacteroidetes*	Soap Lake, Washington
LCH_B177 (HM582741)	*Clostridium aceticum* (AB910752)	96%	*Firmicutes/ Clostridia*	Culture Collection, Japan

Table 4. 6: Phylogenetic affiliation of 16S rRNA sequences of bacterial clones obtained from Shalla Lake

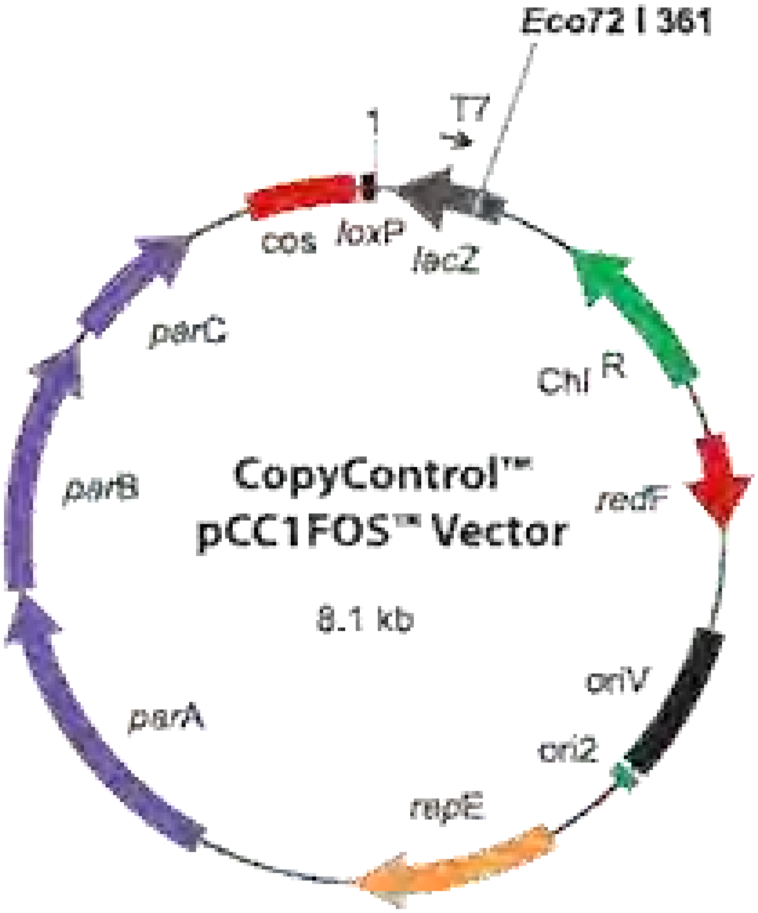
Sample site/ Clone code (Accession No.)	Nearest phylogenetic neighbor from Database (Accession No.)	Similarity (%)	Phylogenetic Affiliation/group from RDP	Function/ Habitat source
LSH_B01 (HM582742)	Uncultured bacterium (FJ712595)	100%	*Gammaproteobacteria/ Escherichia/Shigella*	Kazan mud volcano, East Mediterranean Sea
LSH_B03 (HM582743)	*Pseudidiomarina sp.* (GQ202579)	97%	*Gammaproteobacteria/ Idiomarinaceae*	Lonar soda lake, India
LSH_B07 (HM582744)	Uncultured bacterium (HM150244)	100%	*Gammaproteobacteria/ Escherichia/Shigella*	Sediment sample, thermal spring, Russia
LSH_B08 (HM582745)	*Cellulomonas sp.* (JQ901952)	99%	*Actinobacteria/ Isoptericola*	Soda solonchak soil near soda lake, Kenya
LSH_B10 (HM582746)	*Alkalimonas delamerensis* (NR_044879)	97%	*Gammaproteobacteria/ Alkalimonas*	Alkaline saline Lake, Lake Elmenteita
LSH_B12 (HM582747)	Uncultured bacterium (HM128261)	94%	*Deinococcus-Thermus*	Xiaochaidan Lake, Tibet
LSH_B13	*Bacillus halmapalus*	99%	*Firmicutes/Bacillus*	Culture Collection

(HM582748)	strain (NR_026144)			
LSH_B14 (HM582749)	*Alkalibacterium* sp. (FJ764767)	99%	*Firmicutes/ Alkalibacterium*	Isolates from Lake Elmenteita, Kenya
LSH_B17 (HM582750)	Uncultured *Cellulomonas sp.* (JQ860981)	96%	*Actinobacteria/ Demequina*	Semi-desert soil, India
LSH_B18 (HM582751)	Unidentified Hailaer soda lake bacterium (AF275711)	99%	*Firmicutes/Bacillus*	Soda Lake, Inner Mongolia, Chaina
LSH_B19 (HM582752)	*Alkalimonas collagenimarina* strain (NR_041515)	98%	*Gammaproteobacteria/ Alkalimonas*	Deep-sea sediment
LSH_B20 (HM582753)	Uncultured bacterium (HQ697807)	98%	*Firmicutes/ Anaerobacillus*	Hydrocarbon contaminated saline-alkali soil, Chaina
LSH_B21 (HM582754)	Uncultured bacterium clone SINH789 (HM128261)	87%	*Bacteroidetes*	Xiaochaidan Lake, Tibet
LSH_B22 (HM582755)	*Rhodocyclaceae* bacterium (DQ343836)	98%	*Betaproteobacteria/ Rhodocyclaceae*	Alkaline, saline, Washington
LSH_B24 (HM582756)	Uncultured bacterium (JQ738933)	98%	*Gammaproteobacteria/ Oceanospirillales*	Lonar sediment surface rocks
LSH_B25 (HM582757)	Uncultured bacterium (JQ427177)	98%	*Alphaproteobacteria/ Rhizobiales*	Alkaline saline soil
LSH_B26 (HM582758)	*Rhodobaca bogoriensis* strain (NR_025089)	99%	*Alphaproteobacteria/ Rhodobaca*	Alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes
LSH_B27 (HM582759)	*Propionibacterium acnes* (AB538431)	100%	*Actinobacteria/ Propionibacterium*	Human body
LSH_B28 (HM582760)	Uncultured bacterium (FJ152677)	99%	*Actinobacteria/ Ilumatobacter*	Alkaline saline soils of the former Lake Texcoco
LSH_B29 (HM582761)	Uncultured bacterium (EU644961)	96%	*Actinobacteria/ Iamiaceae*	Soap Lake, Washing ton
LSH_B31	Uncultured bacterium	99%	*Alphaproteobacteria/*	Alkaline saline soil

HM582762	(JQ426440)		*Rhizobiales*	
LSH_B32 (HM582763)	*Demequina aestuarii* (AB639015)	98%	*Actinobacteria/ Demequina*	Rhizosphere of a mangrove
LSH_B33 (HM582764)	Uncultured bacterium (GQ263848)	91%	*Actinobacteria*	Simulated low level waste site
LSH_B34 (HM582765)	Gamma proteobacterium (FJ764788)	99%	*Gammaproteobacteria/ Idiomarinaceae*	Moderately alkaline saline shallow lake, Lake Elmenteita, Kenya
LSH_B35 (HM582766)	*Isoptericola halotoleran* (AB489222)	99%	*Actinobacteria/ Isoptericola*	Culture collection
LSH_B38 (HM582767)	Wandonia haliotis strain (FJ424814)	97%	*Bacteroidetes/ Flavobacteria/ Wandonia*	Culture collection
LSH_B39 (HM582768)	Uncultured bacterium (FJ462107)	93%	*Acidobacteria/Gp3*	Mesophilic anaerobic reactor fed with effluent from the chemical industry
LSH_B41 (HM582769)	Uncultured Bacterium (HM243958)	95%	*Planctomycetes/ Blastopirellula*	Lower sediment from Honghu Lake, Chaina
LSH_B42 (HM582770)	Uncultured gamma proteobacterium (HM072376)	98%	*Gammaproteobacteria/ Idiomarinaceae*	Pangong Lake, India
LSH_B43 (HM582771)	Uncultured bacterium (KC852964)	97%	*Gammaproteobacteria/ Alishewanella*	Tukh Lake sediment, Mongolia
LSH_B45 (HM582772)	Uncultured bacterium (GU455196)	97%	*Planctomycetes/ Rhodopirellula*	anaerobic fermentation reactor with waste activated sludge at mesophilic and pH 9.0
LSH_B46 (HM582773)	Uncultured bacterium (JN178135)	98%	*Firmicutes/ Clostridia*	Extreme saline-alkaline soil of the former lake Texcoco
LSH_B49	*Methylophaga lonarensis* strain	98%	*Gammaproteobacteria/ Methylophaga*	Lonar Soda Lake, India

(HM582774)	(JF330773)			
LSH_B50 (HM582775)	Uncultured *Rhodospirillaceae* bacterium (AM936262)	97%	*Alphaproteobacteria/ Rhodospirillales*	Pilot-scale bioremediation process of a hydrocarbon-contaminated soil
LSH_B51 (HM582776)	*Escherichia albertii* strain (NR_025569)	100%	*Gammaproteobacteria/ Escherichia/Shigella*	Culture collection
LSH_B52 (HM582777)	*Bacillus sp*. (AB437410)	99%	*Firmicutes/ Anaerobacillus*	Alkaliphilic nitrate-reducing bacteria from soil
LSH_B53 (HM582778)	Uncultured *Desulfuromonadales* bacterium (JF727697)	97%	*Deltaproteobacteria/ Desulfuromonadaceae/ Pelobacter*	Petroleum-contaminated saline-alkali soils, Chaina
LSH_B54 (HM582779)	*Mongolicoccus roser* strain (HQ699467)	97%	*Bacteroidetes/ Sphingobacteriales/ Cytophagaceae Litoribacter*	Mongolia Plateau, Hunsandake Desert, Lake Dugerno, Chaina
LSH_B55 (HM582780)	*Isoptericola sp.* (KC608148)	99%	*Actinobacteria/ Promicromonosporaceae/ Isoptericola*	Rhizosphere soil
LSH_B56 (HM582781)	*Aliidiomarina sp.* (JX981923)	96%	*Gammaproteobacteria/ Idiomarinaceae*	Pool for *Spirulina platensis* cultivation
LSH_B57 (HM582782)	Uncultured *planctomycete* (DQ329815)	95%	*Planctomycetes/ Rhodopirellula*	Hypersaline microbial mat: Guerrero Negro, Mexico
LSH_B58 (HM582783)	*Alishewanella jeotgali* strain (EU817498)	97%	*Gammaproteobacteria/ Alishewanella*	Traditional fermented food in Korea
LSH_B59 (HM582784)	*Alkalibacterium sp.* (FJ764767)	99%	*Firmicutes/ Bacilli/ Alkalibacterium*	Moderately alkaline saline shallow lake, Lake Elmenteita, Kenya
LSH_B60 (HM582785)	*Symbiobacterium thermophilum* strain (NR_075044)	85%	*Firmicutes/ Clostridia*	Culture collection

Apendix IV: Map and features of the Fosmid vector (CopyControl PCC1FOS vector)



Appendix V. Fig. 5.4 Representatives of Prokaryotic community distribution in LCH, LSH and ABJ sediment samples

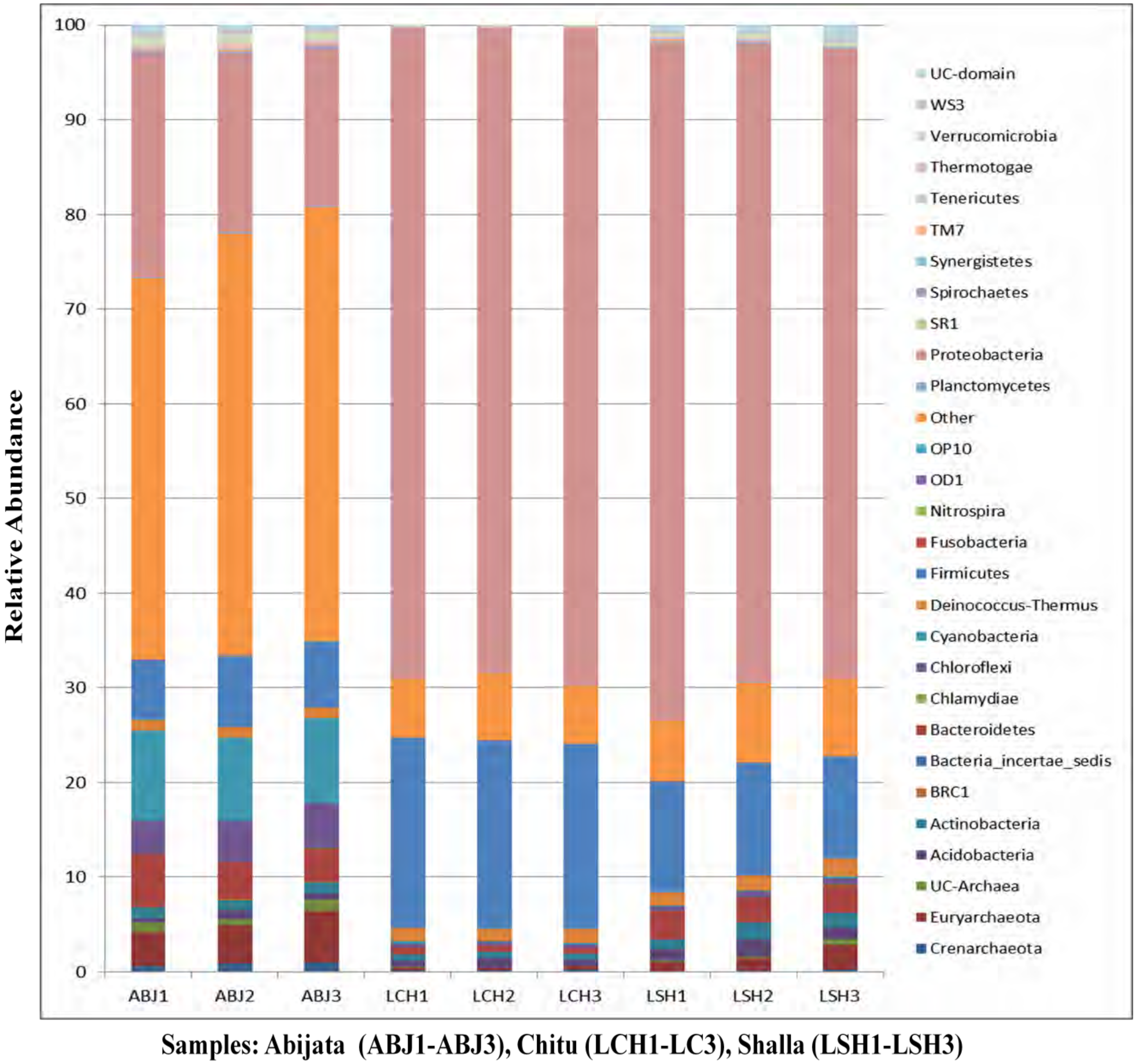


Fig. 5.4: Prokaryotic community distribution in LCH, LSH and ABJ sediment samples. The figure shows the relative abundances of different phyla. The relative abundance is presented in terms of percentage in total microbial sequences per sample.

Apendix VI. COG Categories and Codes

Energy production and conversion	**C**
Cell cycle control and mitosis	**D**
Amino Acid metabolis and transport	**E**
Nucleotide metabolism and transport	**F**
Carbohydrate metabolism and transport	**G**
Coenzyme metabolis	**H**
Lipid metabolism	**I**
Tranlsation	**J**
Transcription	**K**
Replication and repair	**L**
Cell wall/membrane/envelop biogenesis	**M**
Cell motility	**N**
Post-translational modification, protein turnover, chaperone functions	**O**
Inorganic ion transport and metabolism	**P**
Secondary Structure	**Q**
Signal Transduction	**T**
Intracellular trafficing and secretion	**U**
General Functional Prediction only	**R**
Function Unknown	**S**
Defense mechanisms	**V**

Appendix VII. Table 7.4: Illumina Sequence Data Showing Major KEGG, COGs, and Metabolic Pathways

ORF (Functional Gene)	Taxonomic Affliation			KEGG	Functional Category (COGs)	Metabolic Pathway
Glycerophosphoryl diester phosphodiesterase	*Alphaproteobacteria*			K01126	Lipid Metabolism	Glycerophospholipid metabolism [PATH:ko00564]
Long-chain-fatty-acid-CoA ligase	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae (Rhodopseudomonas)*	K00666	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
Multi-sensor hybrid histidine kinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K07678	Enzyme Families	Protein kinases [BR:ko01001]
Two-component system, cell cycle response regulator	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02488	Signal Transduction	Two-component system [PATH:ko02020]
Adenosylmethionine--8-amino-7-oxononanoate transaminase	*Gammaproteobacteri*	*Enterobacteriales*	*Enterobacteriaceae*	K00833	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Type III secretion protein	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K03227	Membrane Transport	Bacterial secretion system [PATH:ko03070]
Helicase, partial	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K03724	Replication and Repair	DNA repair and recombination proteins [BR:ko03400]
Amino-7-oxononanoate synthase Short=AONS AltName: Full=7-keto-8-amino-pelargonic acid synthase Short=7-KAP synthase Short=KAPA synthase AltName: Full=8-amino-7-ketopelargonate synthase	*Gammaproteobacteri*	*Enterobacteriales*	*Enterobacteriaceae*	K00652	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Transporter, major facilitator family protein	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	K08156	Membrane Transport	Transporters [BR:ko02000]
Deoxyxylulose-5-phosphate synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Deflftia)*	K01662	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
DNA polymerase III subunit epsilon	*Betaproteobacteria*	*Burkholderiales*		K02342	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Lipoprotein, YaeC family				K02073	Membrane Transport	Transporters [BR:ko02000]
protein containing Acyl-CoA oxidase/dehydrogenase, type 1 domain				K00249	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Acid-coenzyme A ligase				K08295	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
3-hydroxyacyl-CoA dehydrogenase oxidoreductase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae*	K00074	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]

Long-chain-fatty-acid--CoA ligase				**K00666**	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
Extracellular ligand-binding receptor	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01999**	Membrane Transport	Transporters [BR:ko02000]
Conserved hypothetical protein				**K03520**	Energy Metabolism	Methane metabolism [PATH:ko00680]
Salicylyl-CoA 5-hydroxylase	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae (Delftia)*	**K09461**	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Excinuclease ABC subunit B, partial				**K03702**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
Peptidase M20D, amidohydrolase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01451**	Amino Acid Metabolism	Phenylalanine metabolism [PATH:ko00360]
Mandelate racemase/muconate lactonizing protein	*Betaproteobacteria*	*Burkholderiales*		**K01631**	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
Phosphate regulon sensor histidine kinase protein PhoR	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	**K07636**	Enzyme Families	Protein kinases [BR:ko01001]
(p)ppGpp synthetase I SpoT/RelA	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Signal recognition particle protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03106**	Membrane Transport	Bacterial secretion system [PATH:ko03070]
Carbon monoxide dehydrogenase large subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03520**	Energy Metabolism	Methane metabolism [PATH:ko00680]
1-deoxy-D-xylulose 5-phosphate synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01662**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
Transcription-repair coupling factor	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03723**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00790**	Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism [PATH:ko00520]
Cytosine deaminase				**K01485**	Nucleotide Metabolism	Pyrimidine metabolism [PATH:ko00240]
Malic enzyme-	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00028**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Acetamidase/formamidase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K01455**	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]

O-succinylbenzoate--CoA ligase				K01904	Amino Acid Metabolism	Phenylalanine metabolism [PATH:ko00360]
Sugar transport protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03762	Membrane Transport	Transporters [BR:ko02000]
LysR family Transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K01672	Energy Metabolism	Nitrogen metabolism [PATH:ko00910]
Salicylyl-CoA 5-hydroxylase	*Betaproteobacteria*	*Burkholderiales*		K09461	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Heat shock protein 90				K04079	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
protein containing Glutamine amidotransferase type 1 domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01951	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K00991	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
protein containing RNA polymerase sigma factor 54, DNA-binding domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K03092	Transcription	RNA polymerase [PATH:ko03020]
Ttranscription-repair coupling factor	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03723	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
D-galactonate transporter, partial	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae (Pseudomonas)*	K08194	Membrane Transport	Transporters [BR:ko02000]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01595	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
GntR family transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*		K00375	Transcription	Transcription factors [BR:ko03000]
GntR family transcriptional regulator,	*Betaproteobacteria*	*Burkholderiales*		K00375	Transcription	Transcription factors [BR:ko03000]
ABC transporter ATP-binding/permease protein,				K06147	Membrane Transport	Transporters [BR:ko02000]
ABC transporter, ATP-binding protein/permease protein				K06147	Membrane Transport	Transporters [BR:ko02000]
IMP dehydrogenase/GMP reductase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00088	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
heavy metal sensor signal transduction histidine kinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Comamonas)*	K07644	Enzyme Families	Protein kinases [BR:ko01001]
4-hydroxybenzoate 3-monooxygenase,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00481	Xenobiotics Biodegradation and Metabolism	2,4-Dichlorobenzoate degradation [PATH:ko00623]
MarR family transcriptional regulator	*Alphaproteobacteria*			K06075	Transcription	Transcription factors [BR:ko03000]

Acyl-CoA dehydrogenase domain-containing protein	*Alphaproteobacteria*			K00249	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
2-dehydro-3-deoxygalactonokinase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K00883	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01814	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Acid-coenzyme A ligase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K08295	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Glutathione S-transferase domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00799	Metabolism of Other Amino Acids	Glutathione metabolism [PATH:ko00480]
Carbamoyl-phosphate synthase I chain ATP-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01968	Amino Acid Metabolism	Valine, leucine and isoleucine degradation [PATH:ko00280]
Acyl-coa dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K00249	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Ribonucleotide-diphosphate reductase subunit beta				K00526	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
ATP-dependent Clp protease, proteolytic subunit ClpP	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01358	Enzyme Families	Peptidases [BR:ko01002]
Protease ATP-binding subunit ClpX	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03544	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
Endopeptidase La	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01338	Enzyme Families	Peptidases [BR:ko01002]
Acetyl-CoA carboxylase, biotin carboxylase subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01961	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Major facilitator family protein				K08196	Membrane Transport	Transporters [BR:ko02000]
Methylmalonate -semialdehyde dehydrogenase, partial	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00140	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
D-alanyl-D-alanine carboxypeptidase / D-alanyl-D-alanine-endopeptidase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K07259	Glycan Biosynthesis and Metabolism	Peptidoglycan biosynthesis [PATH:ko00550]
AMP-dependent synthetase/ligase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K08295	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Branched-chain amino acid ABC transporter ATP-binding protein	*Bacilli*	*Lactobacillales*	*Streptococcaceae (Streptococcus)*	K01996	Membrane Transport	Transporters [BR:ko02000]
ABC transporter, ATP-binding protein,	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K02003	Membrane Transport	Transporters [BR:ko02000]

Winged helix family two component heavy metal response transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K07665**	Signal Transduction	Two-component system [PATH:ko02020]
Thiamine-phosphate pyrophosphorylase				**K00788**	Metabolism of Cofactors and Vitamins	Thiamine metabolism [PATH:ko00730]
Tartrate/fumarate subfamily Fe-S type hydro-lyase subunit alpha	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01676**	Carbohydrate Metabolism	Citrate cycle (TCA cycle) [PATH:ko00020]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
N-formylglutamate amidohydrolase				**K01458**	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Formiminoglutamate deiminase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K05603**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Excinuclease ABC subunit B	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	**K03702**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
Porphobilinogen deaminase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01749**	Metabolism of Cofactors and Vitamins	Porphyrin and chlorophyll metabolism [PATH:ko00860]
Phosphoribosylaminoimidazolesuccinocarboxamide synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01923**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Peptidase C56, PfpI	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K05520**	Enzyme Families	01002 Peptidases [BR:ko01002]
Copper resistance protein	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae (Pseudomonas)*	**K00423**	Carbohydrate Metabolism	Ascorbate and aldarate metabolism [PATH:ko00053]
Acetyl-CoA biotin carboxyl carrier	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02160**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Enoyl-CoA hydratase/isomerase				**K01692**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
NAD-dependent aldehyde dehydrogenase				**K00128**	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Molybdate ABC transporter, ATPase subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02017**	Membrane Transport	Transporters [BR:ko02000]
EAL domain protein	*Gammaproteobacteria*			**K02761**	Membrane Transport	Transporters [BR:ko02000]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]

Protein containing Carbamoyl phosphate synthase, large subunit, N-terminal domain				K01961	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
D-3-phosphoglycerate dehydrogenase,				K00058	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
D-3-phosphoglycerate dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K00058	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
Sulfatase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03760	Glycan Biosynthesis and Metabolism	Lipopolysaccharide biosynthesis proteins [BR:ko01005]
3-oxoadipate CoA-transferase subunit B	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01032	Xenobiotics Biodegradation and Metabolism	Benzoate degradation via hydroxylation [PATH:ko00362]
3-oxoacid CoA-transferase subunit A	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01031	Xenobiotics Biodegradation and Metabolism	Benzoate degradation via hydroxylation [PATH:ko00362]
Peptidase M20	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01423	Amino Acid Metabolism	Lysine degradation [PATH:ko00310]
Histidine ammonia-lyase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01745	Energy Metabolism	Nitrogen metabolism [PATH:ko00910]
Integral membrane sensor signal transduction histidine kinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02668	Enzyme Families	Protein kinases [BR:ko01001]
Type IV fimbriae expression regulatory protein PilR	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02667	Signal Transduction	Two-component system [PATH:ko02020]
Potassium-transporting ATPase subunit B,	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01547	Signal Transduction	Two-component system [PATH:ko02020]
Phosphoserine phosphatase SerB	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01079	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
type III secretion apparatus H+-transporting two-sector ATPase				K03224	Membrane Transport	Bacterial secretion system [PATH:ko03070]
Ku domain-containing protein	*Alphaproteobacteria*			K01972	Replication and Repair	DNA replication [PATH:ko03030]
4-hydroxyphenylacetate 3-hydroxylase				K00483	Amino Acid Metabolism	Tyrosine metabolism [PATH:ko00350]
Transcription-repair coupling factor,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03723	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
Ribosomal protein L20	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02887	Translation	Ribosome [PATH:ko03010]

Enoyl-CoA hydratase/isomerase	*Betaproteobacteria*	*Burkholderiales*		**K01692**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Aminobenzoyl-glutamate utilization protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01436**	Amino Acid Metabolism	Arginine and proline metabolism [PATH:ko00330]
CoA synthetase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01906**	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Pseudaminic acid synthase				**K05304**	Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism [PATH:ko00520]
Nucleotide sugar epimerase/dehydratase WbpM				**K00100**	Carbohydrate Metabolism	Fructose and mannose metabolism [PATH:ko00051]
Spore coat polysaccharide biosynthesis protein SpsC				**K07806**	Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism [PATH:ko00520]
Imidazoleglycerol phosphate synthase, cyclase subunit				**K02500**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Imidazole glycerol phosphate synthase subunit HisH				**K02501**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
ATP dependent DNA ligase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K01971**	Replication and Repair	Base excision repair [PATH:ko03410]
Histidinol-phosphate aminotransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K00817**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Thioredoxin	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K05838**	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
D-galactonate transporter	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae (Pseudomonas)*	**K08194**	Membrane Transport	Transporters [BR:ko02000]
Phosphoenolpyruvate carboxylase				**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Pyridoxal phosphate (active vitamin B6) biosynthesis PdxA	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K00097**	Metabolism of Cofactors and Vitamins	Vitamin B6 metabolism [PATH:ko00750]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Protein containing Peptidase M23 domain	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae (Bordetella)*	**K08259**	Enzyme Families	Peptidases [BR:ko01002]
secreted protein containing HAMP linker domain domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K07638**	Enzyme Families	Protein kinases [BR:ko01001]

Glutamate/aspartate ABC superfamily ATP binding cassette transporter, ABC protein				K02028	Membrane Transport	Transporters [BR:ko02000]
Acetyl-CoA carboxylase, biotin carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01968	Amino Acid Metabolism	Valine, leucine and isoleucine degradation [PATH:ko00280]
D-isomer specific 2-hydroxyacid dehydrogenase nad-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00058	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
Homoserine O-acetyltransferase-	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00641	Energy Metabolism	Sulfur metabolism [PATH:ko00920]
IMP dehydrogenase/GMP reductase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00088	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Protein containing Signal transduction response regulator, C-terminal domain	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K07665	Signal Transduction	Two-component system [PATH:ko02020]
Heavy metal sensor kinase family protein 1	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K02484	Enzyme Families	Protein kinases [BR:ko01001]
Protein containing Aldehyde dehydrogenase domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00140	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Diacylglycerol kinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00901	Lipid Metabolism	Glycerolipid metabolism [PATH:ko00561]
Imidazolonepropionase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01468	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Coenzyme A transferase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K01031	Xenobiotics Biodegradation and Metabolism	Benzoate degradation via hydroxylation [PATH:ko00362]
IclR family Transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K02624	Transcription	Transcription factors [BR:ko03000]
Bbinding-protein-dependent transport systems inner membrane component				K02050	Membrane Transport	Transporters [BR:ko02000]
AMP-dependent synthetase/ligase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K08295	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Inner-membrane translocator				K01998	Membrane Transport	Transporters [BR:ko02000]
3-oxoadipate enol-lactone hydrolase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01055	Xenobiotics Biodegradation and Metabolism	Benzoate degradation via hydroxylation [PATH:ko00362]

Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Methyl-accepting chemotaxis protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03406**	Signal Transduction	Two-component system [PATH:ko02020]
YdeP domain protein				**K00122**	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
protein containing GMP synthase, C-terminal domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Protein containing Allophanate hydrolase subunit 2 domain	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	**K01941**	Amino Acid Metabolism	Arginine and proline metabolism [PATH:ko00330]
Protein containing Aldehyde dehydrogenase domain				**K00155**	Metabolism of Terpenoids and Polyketides	Limonene and pinene degradation [PATH:ko00903]
Protein containing Bacterial regulatory protein GntR, HTH domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K05836**	Transcription	Transcription factors [BR:ko03000]
Transcriptional regulator, histidine utilization repressor, GntR family	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K05836**	Transcription	Transcription factors [BR:ko03000]
CoA-binding domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K01906**	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Phosphoribosylaminoimidazole carboxylase ATPase subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01589**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Translation elongation factor P	*Alphaproteobacteria*			**K02356**	Translation	Translation factors [BR:ko03012]
phosphoribosylaminoimidazole carboxylase catalytic subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Comamonas)*	**K01588**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Type I secretion system ATPase, PrtD, partial	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae (Serratia)*	**K06147**	Membrane Transport	Transporters [BR:ko02000]
Rieske (2Fe-2S) domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00499**	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
D-3-phosphoglycerate dehydrogenase				**K00058**	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
Enoyl-CoA hydratase/isomerase				**K01692**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]

Protein containing Transcriptional regulator IclR, C-terminal domain				**K02624**	Transcription	Transcription factors [BR:ko03000]
Histidine ammonia-lyase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01745**	Energy Metabolism	00910 Nitrogen metabolism [PATH:ko00910]
	Gammaproteobacteria	*Pseudomonadales*	*Pseudomonadaceae (Pseudomonas)*	**K02029**	Membrane Transport	Transporters [BR:ko02000]
Transcription-repair coupling factor	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03723**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
ATP-dependent DNA helicase RecQ	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03654**	Folding, Sorting and Degradation	RNA degradation [PATH:ko03018]
Hydantoinase/carbamoylase family amidase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01431**	Nucleotide Metabolism	00240 Pyrimidine metabolism [PATH:ko00240]
Hhydantoinase/carbamoylase family amidase				**K01431**	Nucleotide Metabolism	Pyrimidine metabolism [PATH:ko00240]
Salicylyl-CoA 5-hydroxylase	*Betaproteobacteria*	*Burkholderiales*		**K09461**	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
Chemotaxis protein methyltransferase CheR	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K00575**	Signal Transduction	Two-component system [PATH:ko02020]
Carbonate dehydratase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae*	**K01673**	Energy Metabolism	Nitrogen metabolism [PATH:ko00910]
1-deoxy-D-xylulose-5-phosphate synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01662**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
Kinase protein (partial sequence), partial	*Betaproteobacteria*	*Burkholderiales*		**K03338**	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Glutamate racemase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K01776**	Metabolism of Other Amino Acids	D-Glutamine and D-glutamate metabolism [PATH:ko00471]
Catalase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03781**	Energy Metabolism	Methane metabolism [PATH:ko00680]
Phosphoribosylaminoimidazole carboxylase catalytic subunit				**K01588**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Phosphoribosylaminoimidazole carboxylase ATPase subunit				**K01589**	Nucleotide Metabolism	00230 Purine metabolism [PATH:ko00230]
Protein containing FAD-linked oxidase, C-terminal domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00102**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]

Histidine kinase				K07638	Enzyme Families	Protein kinases [BR:ko01001]
Histidinol dehydrogenase, prokaryotic-type	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00013	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
4-hydroxyphenylacetate 3-hydroxylase	*Alphaproteobacteria*			K00483	Amino Acid Metabolism	Tyrosine metabolism [PATH:ko00350]
Histidinol dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00013	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
ABC transporter, permease/ATP-binding protein	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K06147	Membrane Transport	Transporters [BR:ko02000]
Mammalian cell entry related domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02067	Membrane Transport	Transporters [BR:ko02000]
Protein containing Catalase-related immune responsive domain	*Betaproteobacteria*	*Burkholderiales*	*Acidovorax*	K03781	Energy Metabolism	Methane metabolism [PATH:ko00680]
Acetyl-coenzyme A synthetase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01895	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Acetyl-CoA synthetase,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01906	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Heat shock protein 90	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K04079	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
Multidrug ABC transporter permease	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K09686	Membrane Transport	Transporters [BR:ko02000]
Homoserine O-acetyltransferase-	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00641	Energy Metabolism	Sulfur metabolism [PATH:ko00920]
Short-chain dehydrogenase/reductase SDR	*Betaproteobacteria*	*Burkholderiales*		K00100	Carbohydrate Metabolism	Fructose and mannose metabolism [PATH:ko00051]
Protein containing Threonyl/alanyl tRNA synthetase, SAD domains	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01868	Translation	Aminoacyl-tRNA biosynthesis [PATH:ko00970]
Phosphoenolpyruvate carboxylase				K01595	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Threonyl-tRNA synthetase	*Betaproteobacteria*	*Burkholderiales*		*K01868*	Translation	Aminoacyl-tRNA biosynthesis [PATH:ko00970]

Pyridoxamine 5'-phosphate oxidase AltName: Full=PNP/PMP oxidase Short=PNPOx AltName: Full=Pyridoxal 5'-phosphate	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae (Brevundimonas)*	**K00275**	Metabolism of Cofactors and Vitamins	Vitamin B6 metabolism [PATH:ko00750]
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03270**	Glycan Biosynthesis and Metabolism	Lipopolysaccharide biosynthesis [PATH:ko00540]
Threonyl-tRNA synthetase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01868**	Translation	Aminoacyl-tRNA biosynthesis [PATH:ko00970]
Conserved hypothetical protein, membrane				**K01992**	Membrane Transport	Transporters [BR:ko02000]
Polyprenyl synthetase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K00795**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
Conserved hypothetical protein, membrane				**K01992**	Membrane Transport	Transporters [BR:ko02000]
ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component				**K02049**	Membrane Transport	Transporters [BR:ko02000]
Binding-protein-dependent transport system inner membrane protein				**K02050**	Membrane Transport	Transporters [BR:ko02000]
Myo-inositol catabolism IolB domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03337**	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Ribokinase-like domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K03338**	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Aminotransferase class-III	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00822**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
periplasmic-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02016**	Membrane Transport	Transporters [BR:ko02000]
MmgE/PrpD family protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01720**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Short-chain alcohol dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae (Burcholderia)*	**K00059**	Lipid Metabolism	Fatty acid biosynthesis [PATH:ko00061]
Protein containing DNA/RNA helicase, C-terminal domains	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	**K03702**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00991**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]

ABC transporter, ATP-binding protein	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K01990	Membrane Transport	Transporters [BR:ko02000]
Fe(3+)-transporting ATPase				K01996	Membrane Transport	Transporters [BR:ko02000]
Conserved hypothetical protein				K01431	Nucleotide Metabolism	Pyrimidine metabolism [PATH:ko00240]
LysR family transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae (Burcholderia)*	K01672	Energy Metabolism	Nitrogen metabolism [PATH:ko00910]
3-octaprenyl-4-hydroxybenzoate carboxy-lyase	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K03186	Metabolism of Cofactors and Vitamins	Ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]
AraC family transcriptional regulator				K01455	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Phosphate acetyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00028	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Galactonate dehydratase				K01684	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
ATP phosphoribosyltransferase catalytic subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00765	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
AMP-dependent synthetase and ligase	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae (Brevundimonas)*	K00666	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
Carbon monoxide dehydrogenase large chain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03520	Energy Metabolism	Methane metabolism [PATH:ko00680]
Dihydroxy-acid dehydratase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01687	Amino Acid Metabolism	Valine, leucine and isoleucine biosynthesis [PATH:ko00290]
ABC transporter-like protein				K02032	Membrane Transport	Transporters [BR:ko02000]
Short-chain dehydrogenase/reductase SDR				K00034	Carbohydrate Metabolism	Pentose phosphate pathway [PATH:ko00030]
membrane protein containing DUF214, permase predicted				K02004	Membrane Transport	Transporters [BR:ko02000]
Class III aminotransferase				K00823	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
N-acetyltransferase GCN5				K00657	Amino Acid Metabolism	Arginine and proline metabolism [PATH:ko00330]
ABC-type multidrug transport system, ATPase component	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K01990	Membrane Transport	Transporters [BR:ko02000]

3-hydroxybutyrate dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00019**	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]
ATP-dependent DNA helicase RecQ	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03654**	Folding, Sorting and Degradation	RNA degradation [PATH:ko03018]
Carbamoyl-phosphate synthase L chain, ATP-binding protein, partial	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01968**	Amino Acid Metabolism	Valine, leucine and isoleucine degradation [PATH:ko00280]
3-hydroxybutyryl-CoA dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K00074**	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]
LexA repressor	*Betaproteobacteria*	*Gallionellales*	*Gallionellaceae*	**K01356**	Enzyme Families	Peptidases [BR:ko01002]
Conserved hypothetical protein				**K01631**	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
GMP synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
GMP synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Protein containing Aldehyde dehydrogenase domain	*Betaproteobacteria*	*Burkholderiales*		**K00140**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Dihydroxy-acid dehydratase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01687**	Amino Acid Metabolism	Valine, leucine and isoleucine biosynthesis [PATH:ko00290]
Hydroxymethylglutaryl-CoA lyase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01640**	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]
Glycerophosphoryl diester phosphodiesterase, periplasmic				**K01126**	Lipid Metabolism	Glycerophospholipid metabolism [PATH:ko00564]
Enoyl-CoA hydratase/isomerase				**K01692**	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Protein containing 3-hydroxyacyl-CoA dehydrogenase, NAD binding domain	*Betaproteobacteria*			**K00074**	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]
N-formylglutamate amidohydrolase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01458**	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
ATP phosphoribosyltransferase catalytic subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00765**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Type III secretion protein SpaR/YscT				**K03228**	Membrane Transport	Bacterial secretion system [PATH:ko03070]
Protein containing Sulfatase domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03760**	Glycan Biosynthesis and Metabolism	Lipopolysaccharide biosynthesis proteins [BR:ko01005]

Integral membrane sensor signal transduction histidine kinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K07638**	Enzyme Families	Protein kinases [BR:ko01001]
Succinyl-diaminopimelate desuccinylase-like protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01423**	Amino Acid Metabolism	Lysine degradation [PATH:ko00310]
Thioesterase superfamily protein	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K01075**	Metabolism of Cofactors and Vitamins	Ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]
IclR family transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02624**	Transcription	Transcription factors [BR:ko03000]
GMP synthase, large subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Xaa-Pro dipeptidase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01262**	Enzyme Families	Peptidases [BR:ko01002]
Uroporphyrin-III C-methyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02496**	Metabolism of Cofactors and Vitamins	Porphyrin and chlorophyll metabolism [PATH:ko00860]
Protein containing AMP-dependent synthetase and ligase domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00666**	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
KU domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01972**	Replication and Repair	DNA replication [PATH:ko03030]
Extracellular ligand-binding receptor				**K01999**	Membrane Transport	Transporters [BR:ko02000]
LysR family transcriptional regulator				**K03566**	Transcription	Transcription factors [BR:ko03000]
branched-chain amino acid ABC-type transport system, permease component	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01997**	Membrane Transport	Transporters [BR:ko02000]
ABC-type branched-chain amino acid transport system, ATPase component	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01996**	Membrane Transport	Transporters [BR:ko02000]
ABC transporter-like protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01995**	Membrane Transport	Transporters [BR:ko02000]
Fe-sulfur protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00499**	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (bordetella)*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Chaperone protein htpG	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K04079**	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]

Stress response DNA-binding protein (Dps)				K04047	Replication and Repair	Chromosome [BR:ko03036]
Betaine aldehyde dehydrogenase				K00128	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Choline dehydrogenase				K00108	Amino Acid Metabolism	Glycine, serine and threonine metabolism [PATH:ko00260]
Peptidase M20D, amidohydrolase				K01451	Amino Acid Metabolism	Phenylalanine metabolism [PATH:ko00360]
Deoxyxylulose-5-phosphate synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01662	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01595	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Iimidazolonepropionase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K01468	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Dihydrolipoamide dehydrogenase,	*Alphaproteobacteria*	*Rhodospirillales*	*Acetobacteraceae (Acetobacter)*	K00382	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Receptor family ligand-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K01999	Membrane Transport	Transporters [BR:ko02000]
Protein containing Signal transduction response regulator, C-terminal domains	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae (Pseudomonas)*	K07665	Signal Transduction	Two-component system [PATH:ko02020]
Phosphoribosylaminoimidazole carboxylase, catalytic subunit				K01588	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
RNA polymerase sigma factor RpoE	*Alphaproteobacteria*			K03088	Transcription	RNA polymerase [PATH:ko03020]
Biotin synthase	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	K01012	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Ribonucleotide-diphosphate reductase subunit alpha,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00525	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Hydrolase or acyltransferase of alpha/beta superfamily	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01259	Amino Acid Metabolism	Arginine and proline metabolism [PATH:ko00330]
Enoyl-CoA hydratase/isomerase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01692	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Thymidylate synthase				K00560	Nucleotide Metabolism	Pyrimidine metabolism [PATH:ko00240]

Phosphoenolpyruvate carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K01595	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Conserved hypothetical protein				K01923	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Branched-chain amino acid ABC transporter ATP-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K01996	Membrane Transport	Transporters [BR:ko02000]
Acetyl-CoA carboxylase, biotin carboxylase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01961	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Lipid ABC transporter ATPase/inner membrane protein	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K06147	Membrane Transport	Transporters [BR:ko02000]
Transporter, major facilitator family protein	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K08156	Membrane Transport	Transporters [BR:ko02000]
Outer membrane protein lom	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	K07804	Signal Transduction	Two-component system [PATH:ko02020]
protein containing Malic enzyme, NAD-binding domains	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00028	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Imidazole glycerol phosphate synthase, glutamine amidotransferase subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K02501	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Protein containing Aldehyde dehydrogenase domain				K00155	Metabolism of Terpenoids and Polyketides	Limonene and pinene degradation [PATH:ko00903]
Porphobilinogen deaminase, partial	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01749	Metabolism of Cofactors and Vitamins	Porphyrin and chlorophyll metabolism [PATH:ko00860]
Conserved hypothetical protein				K01992	Membrane Transport	Transporters [BR:ko02000]
Inner-membrane translocator				K01998	Membrane Transport	Transporters [BR:ko02000]
FMN-dependent alpha-hydroxy acid dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00104	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Protein containing Biotin carboxylase, C-terminal domains	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01961	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
IolC protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K03338	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Phosphoribosylaminoimidazole carboxylase, ATPase subunit	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K01589	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]

Protein containing Helicase and RNase D C-terminal, HRDC domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03654**	Folding, Sorting and Degradation	RNA degradation [PATH:ko03018]
Winged helix family two component transcriptional regulator	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	**K07667**	Signal Transduction	Two-component system [PATH:ko02020]
Two-component sensor KdpD	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	**K07646**	Enzyme Families	Protein kinases [BR:ko01001]
K+-transporting ATPase, C subunit	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	**K01548**	Signal Transduction	Two-component system [PATH:ko02020]
Potassium-transporting ATPase subunit B	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	**K01547**	Signal Transduction	Two-component system [PATH:ko02020]
ModE family transcriptional regulator	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02019**	Transcription	Transcription factors [BR:ko03000]
Molybdenum ABC transporter periplasmic molybdate-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02020**	Membrane Transport	Transporters [BR:ko02000]
2-dehydro-3-deoxygalactonokinase				**K00883**	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
Exonuclease VII small subunit	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K03602**	Replication and Repair	Mismatch repair [PATH:ko03430]
Catalase,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Acidovorax)*	**K03781**	Energy Metabolism	Methane metabolism [PATH:ko00680]
Oxidoreductase alpha (molybdopterin) subunit,	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae*	**K00122**	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Thioredoxin				**K05838**	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02049**	Membrane Transport	Transporters [BR:ko02000]
Protein containing ABC-1 domain	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae*	**K03688**	Metabolism of Cofactors and Vitamins	Ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]
Chaperone protein DnaK	*Betaproteobacteria*	*Burkholderiales*		**K04043**	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
DNA-binding transcriptional regulator TyrR,	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae (Serratia)*	**K03721**	Transcription	Transcription factors [BR:ko03000]
GTP pyrophosphokinase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00951**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]

PTS transporter subunit IIA-like nitrogen-regulatory protein PtsN	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02806**	Membrane Transport	Transporters [BR:ko02000]
Acetyl-coenzyme a synthetase	*Betaproteobacteria*	*Burkholderiales*		**K01895**	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Protein containing AMP-dependent synthetase and ligase domain	*Betaproteobacteria*	*Burkholderiales*		**K01895**	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Cytochrome O ubiquinol oxidase protein, subunit I	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02298**	Energy Metabolism	Oxidative phosphorylation [PATH:ko00190]
Dihydrolipoamide dehydrogenase	*Alphaproteobacteria*	*Rhizobiales*		**K00382**	Carbohydrate Metabolism	Glycolysis / Gluconeogenesis [PATH:ko00010]
Long-chain-fatty-acid--CoA ligase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00666**	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
Cysteine ABC transporter permease/ATP-binding protein	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	**K06148**	Membrane Transport	Transporters [BR:ko02000]
Histidinol-phosphate aminotransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00817**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Myo-inositol catabolism IolB domain-containing protein	*Betaproteobacteria*	*Burkholderiales*		**K03337**	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Phosphoenolpyruvate carboxylase				**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Inosine-5'-monophosphate dehydrogenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00088**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
4-hydroxyphenylacetate 3-hydroxylase	*Alphaproteobacteria*			**K00483**	Amino Acid Metabolism	Tyrosine metabolism [PATH:ko00350]
Ribonucleotide-diphosphate reductase subunit beta	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00526**	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Adenosylmethionine--8-amino-7-oxononanoate transaminase	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	**K00833**	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
2Fe-2S iron-sulfur cluster-binding domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03518**	Energy Metabolism	Methane metabolism [PATH:ko00680]
Type III secretion protein	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K03222**	Membrane Transport	Bacterial secretion system [PATH:ko03070]

Branched-chain amino acid ABC superfamily ATP binding cassette transporter, ABC protein				**K01995**	Membrane Transport	Transporters [BR:ko02000]
Phosphoenolpyruvate carboxylase,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01595**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Transcription-repair coupling factor,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03723**	Replication and Repair	Nucleotide excision repair [PATH:ko03420]
Urocanate hydratase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01712**	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Type III secretion exporter				**K03229**	Membrane Transport	Bacterial secretion system [PATH:ko03070]
Binding-protein-dependent transport system inner membrane protein	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	**K02034**	Membrane Transport	Transporters [BR:ko02000]
Sensor histidine kinase KdpD	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	**K07646**	Enzyme Families	Protein kinases [BR:ko01001]
Myo-inositol catabolism IolB domain-containing protein	*Betaproteobacteria*	*Burkholderiales*		**K03337**	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
1-deoxy-D-xylulose-5-phosphate synthase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K01662**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]
PpiC-type peptidyl-prolyl cis-trans isomerase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03771**	Folding, Sorting and Degradation	Chaperones and folding catalysts [BR:ko03110]
Binding-protein-dependent transport systems inner membrane component	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K02050**	Membrane Transport	Transporters [BR:ko02000]
Protein containing Phosphate acetyl/butaryl transferase domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00028**	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Long-chain-fatty-acid--CoA ligase				**K01897**	Lipid Metabolism	Fatty acid metabolism [PATH:ko00071]
ATP-dependent DNA helicase RecQ,	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K03654**	Folding, Sorting and Degradation	RNA degradation [PATH:ko03018]
Translation initiation factor IF-3	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	**K02520**	Translation	Translation factors [BR:ko03012]
Polyprenyl synthetase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	**K00795**	Metabolism of Terpenoids and Polyketides	Terpenoid backbone biosynthesis [PATH:ko00900]

Chorismate synthase	*Alphaproteobacteria*			K01736	Amino Acid Metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis [PATH:ko00400]
Protein containing AMP-dependent synthetase and ligase domain	*Betaproteobacteria*	*Burkholderiales*		K08295	Xenobiotics Biodegradation and Metabolism	Carbazole degradation [PATH:ko00629]
NLPA lipoprotein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02073	Membrane Transport	Transporters [BR:ko02000]
(p)ppGpp synthetase I SpoT/RelA	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00951	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
protein containing ATP-binding region, ATPase-like domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K07638	Enzyme Families	Protein kinases [BR:ko01001]
UbiE/COQ5 methyltransferase				K03183	Metabolism of Cofactors and Vitamins	Ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]
Branched-chain amino acid ABC transporter periplasmic amino acid-binding protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01999	Membrane Transport	Transporters [BR:ko02000]
CoA-binding domain-containing protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K01906	Metabolism of Cofactors and Vitamins	Biotin metabolism [PATH:ko00780]
Long-chain fatty-acid-CoA ligase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00666	Lipid Metabolism	Lipid biosynthesis proteins [BR:ko01004]
D-alanyl-D-alanine carboxypeptidase / D-alanyl-D-alanine-endopeptidase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K07259	Glycan Biosynthesis and Metabolism	Peptidoglycan biosynthesis [PATH:ko00550]
Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Acidovorax)*	K01814	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Histidinol-phosphate aminotransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K00817	Amino Acid Metabolism	Histidine metabolism [PATH:ko00340]
Leucine/isoleucine/valine transporter ATP-binding subunit				K01995	Membrane Transport	Transporters [BR:ko02000]
Flagellar transcriptional activator FlhC	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02402	Signal Transduction	Two-component system [PATH:ko02020]
Isocitrate lyase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01637	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Oxidoreductase alpha (molybdopterin) subunit	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae (Burkholderia)*	K00122	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]

Carbohydrate kinase, PfkB, partial	*Betaproteobacteria*	*Burkholderiales*		K03338	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Peptidase M20	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01423	Amino Acid Metabolism	Lysine degradation [PATH:ko00310]
Myo-inositol-1(or 4)-monophosphatase	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K01092	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
protein containing Malic enzyme, N-terminal domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00028	Carbohydrate Metabolism	Pyruvate metabolism [PATH:ko00620]
Mandelate racemase/muconate lactonizing enzyme, C-terminal domain protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01631	Carbohydrate Metabolism	Galactose metabolism [PATH:ko00052]
NLPA lipoprotein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02073	Membrane Transport	Transporters [BR:ko02000]
(p)ppGpp synthetase I SpoT/RelA	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00951	Nucleotide Metabolism	Purine metabolism [PATH:ko00230]
Oxidoreductase alpha (molybdopterin) subunit,	*Betaproteobacteria*	*Burkholderiales*		K00122	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Ku domain-containing protein				K01972	Replication and Repair	DNA replication [PATH:ko03030]
Major facilitator superfamily protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K08196	Membrane Transport	Transporters [BR:ko02000]
Phosphate regulon sensor histidine kinase protein PhoR	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Achromobacter)*	K07636	Enzyme Families	Protein kinases [BR:ko01001]
NLPA lipoprotein				K02481	Signal Transduction	Two-component system [BR:ko02022]
Taurine dioxygenase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K03119	Metabolism of Other Amino Acids	Taurine and hypotaurine metabolism [PATH:ko00430]
ABC transporter-like protein	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K02049	Membrane Transport	Transporters [BR:ko02000]
IolC protein	*Betaproteobacteria*	*Burkholderiales*		K03338	Carbohydrate Metabolism	Inositol phosphate metabolism [PATH:ko00562]
Sensor protein RstB	*Gammaproteobacter*	*Enterobacteriales*	*Enterobacteriaceae*	K07639	Enzyme Families	Protein kinases [BR:ko01001]
Glutamate-1-semialdehyde aminotransferase				K01845	Metabolism of Cofactors and Vitamins	Porphyrin and chlorophyll metabolism [PATH:ko00860]
Oxidoreductase, aldo/keto reductase family protein				K00100	Carbohydrate Metabolism	Fructose and mannose metabolism

						PATH:ko00051]
Polysaccharide biosynthesis protein CapD				K01726	Carbohydrate Metabolism	Butanoate metabolism [PATH:ko00650]
Glucose-1-phosphate thymidylyltransferase				K00973	Metabolism of Terpenoids and Polyketides	Polyketide sugar unit biosynthesis [PATH:ko00523]
Glycosyl transferase family protein				K00754	Carbohydrate Metabolism	Fructose and mannose metabolism [PATH:ko00051]
Extracellular solute-binding protein, family 5	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae (Delftia)*	K02035	Membrane Transport	Transporters [BR:ko02000]
Peptidase M22 glycoprotease	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01423	Amino Acid Metabolism	Lysine degradation [PATH:ko00310]
GntR family transcriptional regulator				K00375	Transcription	Transcription factors [BR:ko03000]
Extracellular ligand-binding receptor	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K01999	Membrane Transport	Transporters [BR:ko02000]
ABC-2 type transporter	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K09686	Membrane Transport	Transporters [BR:ko02000]
Oxidoreductase alpha (molybdopterin) subunit				K00122	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
Short chain fatty acid transporter	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae (Bordetella)*	K02106	Signal Transduction	Two-component system [PATH:ko02020]
protein containing Aldehyde dehydrogenase domain	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00140	Carbohydrate Metabolism	Propanoate metabolism [PATH:ko00640]
Oxidoreductase alpha (molybdopterin) subunit,	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae (Burkholderia)*	K00122	Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K00790	Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism [PATH:ko00520]
Membrane protein containing ABC transporter, transmembrane region domain	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	K06147	Membrane Transport	Transporters [BR:ko02000]
Carbon monoxide dehydrogenase form II large subunit, partial	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	K03520	Energy Metabolism	Methane metabolism [PATH:ko00680]

Declaration

The undersigned declare that this thesis is my original work, that it has not been submitted for any degree or examination in any other University, and that all the resources I have used or quoted have been indicated and acknowledged by complete references.

Mesfin Tafesse

Date

July 30/2014

This Thesis has been submitted for Examination with my approval as a University Advisor:

Kassahun Tesfaye (PhD)

Date

July 30/2014