ACTINOBACTERIAL DIVERSITY OF THE ETHIOPIAN RIFT VALLEY LAKES

By

Gerda Du Plessis

Submitted in partial fulfillment of the requirements for the degree of
Magister Scientiae (M.Sc.) in the Department of Biotechnology,
University of the Western Cape

Supervisor: Prof. D.A. Cowan
Co-Supervisor: Dr. I.M. Tuffin
November 2011

DECLARATION

I declare that „The Actinobacterial diversity of the Ethiopian Rift Valley Lakes“ is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Gerda Du Plessis

ABSTRACT

The class *Actinobacteria* consists of a heterogeneous group of filamentous, Gram-positive bacteria that colonise most terrestrial and aquatic environments. The industrial and biotechnological importance of the secondary metabolites produced by members of this class has propelled it into the forefront of metagenomic studies. The Ethiopian Rift Valley lakes are characterized by several physical extremes, making it a polyextremophilic environment and a possible untapped source of novel actinobacterial species. The aims of the current study were to identify and compare the eubacterial diversity between three geographically divided soda lakes within the ERV focusing on the actinobacterial subpopulation. This was done by means of a culture-dependent (classical culturing) and culture-independent (DGGE and ARDRA) approach. The results indicate that the eubacterial 16S rRNA gene libraries were similar in composition with a predominance of α-Proteobacteria and Firmicutes in all three lakes. Conversely, the actinobacterial 16S rRNA gene libraries were significantly different and could be used to distinguish between sites. The actinobacterial OTUs detected belonged to both the *Rubrobacterales* and *Actinomycetales* orders with members of the genus *Arthrobacter* being found in all three lakes. Geochemical properties were significantly different between the lakes, although more than one property attributed to the variance between community compositions. The diversity detected in the culture-based study differed significantly and all isolates belonged to the genus *Streptomyces*. Two novel strains were characterized by means of phylogenetic (16S rRNA gene sequence), physiological, morphological and biochemical analyses. Both novel isolates were capable of growing under "extreme" conditions- pH 12, 10% NaCl and 45°C. Partial enzyme characterization revealed that both strains produced xylanase enzymes that were active at pH 6.5 and 8.5 with an increase in activity up to 45°C. The results obtained revealed a previously undetected diversity of actinobacteria in the Ethiopian Rift Valley with a potentially novel subpopulation adapted to haloalkaline conditions. The low 16S rRNA sequence similarity of a substantial proportion of the libraries suggests that culture-based isolation may play a vital role in deciphering the community fingerprint.

PUBLICATIONS ORIGINATING FROM THIS THESIS

Poster/Presentations

1. Soil Metagenomics Symposium 2010. Berlin: Germany. Poster
2. Biotechnology Departmental open day 2011. Cape Town. Poster
3. South African Society for Microbiology conference 2011. Cape Town. Poster

ACKNOWLEDGEMENTS

I would like to extend my gratitude to:

- My supervisors for their effort and time.
- Dr. Bronwyn Kirby, for her unfailing guidance and motivation when I needed it the most.
- Dr. Lise Øvreås and everyone at UIB for the opportunity to present my work.
- Everyone at IMBM for their support and motivation, especially Dr. Heidi Goodman for the endless hours of work.
- My family, especially Johan because without his patience and undying love I would have faltered long ago.
- The National Research Foundation and the Norwegian research counsel for funding allocated to the project.

DEDICATION

This thesis is dedicated to

Johan C. Theron and

Roelof Du Plessis

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ARDRA	Amplified Ribosomal DNA Restriction Analysis
ATCC	American Type Culture Collection
bp	Base pair
BLAST	Basic Local Alignment Search Tool
°C	Degrees Celsius
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide tri-phosphates
DGGE	Denaturing Gradient Gel Electrophoresis
EDTA	Ethylendiaminetetraacetatic acid
ERV	Ethiopian Rift Valley
EtBr	Ethidium Bromide
G+C	Guanine and Cytosine
gDNA	Genomic DNA
x g	Centrifugal force
g	Gram
HS1	Lake Shala Hot spring 1
HS2	Lake Shala Hot spring 2
HS3	Lake Shala Hot spring 3
IPTG	Isopropyl- D-thiogalactoside
LA	Lake Arenguade
LAb	Lake Abitjata
MDS	Multi Dimensional Scaling
MER	Mid Ethiopian Rift
OTU	Operational taxonomic unit
PC	Principle Component
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PVPP	Polyvinylpolypyrrolidone
RBB	4-O-Methyl-D-glucurono-D-xylan−remazol
rpm	Revolutions per minute
SDS	Sodium Dodecyl Sulfate
sp.	Species (singular)
spp.	Species (Plural)
subsp.	Subspecies
TAE	Tris-acetate-EDTA
TE	Tris-HCl-EDTA
U	Units
w/v	Weight per volume
X-Gal	5-Bromo-4-Chloro-3-indolyl- -D-galactopyranoside

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction to haloalkaline environments

Soda lakes are the dominant naturally occurring alkaline environments on the planet and severe abiotic pressures are imposed on organisms that inhabit the lakes (Table 1). The unique composition of high pH and salt concentrations results in the formation of a specialised niche for opportunistic halo-alkaliphiles. However, until recently, very few true halo-alkaliphiles had been discovered (Antony-Babu & Goodfellow 2008; Bowers *et al*., 2009). As has been suggested by Duckworth *et al*. (1996) the complicated nutrient cycles associated with alkaline environments supports the notion that these environments were once common and are not monophyletic.

Soda lakes can be grouped according to several properties of which pH and salt content is the most commonly used. The salt content of saline lakes is estimated to be 5-10 times higher than that of seawater (Sabet 2009) and falls within the range of 150-300g.L^{-1}. Soda lakes, on the other hand, are significantly less saline and the salt content can range between 0.5g-30g.L^{-1}. Saline lakes can be divided into two groups according to the origin of the water mass. Thalassohaline environments are composed of concentrated seawater which consists of, predominantly, NaCl while athalassohaline environments originate from non-seawater sources and contain a mix of ions (Litchfield & Gillevet, 2002). The latter is dominated by K^+ and soda, whereas the former can result from any salt mine drainage, playas and natural coastal pools. It is because of the high soda concentrations that the athalassohaline soda lakes are characteristically alkaline in nature. The origin of various alkali earth and alkali metal salts, which inhibit bacterial growth, is inherently unique to each site and will be discussed in section 1.5.

Soda lakes and deserts are confined to subtropical latitudes and continental interiors, as well as rain-shadow zones (Jones & Grant, 1999). Most of the limnological and microbial research has been conducted on three geographically dispersed lakes; Lake Mono (USA), Lake Wadi Al Natrun (Egypt) and Lake Magadi (Kenya) (Baumgarte, 2003; Litchfield *et al.,* 2002; Mwirichia *et al.* 2010). In addition, scattered studies on lakes in the former USSR and Inner Mongolia have provided further insights into the bacterial diversity of soda lakes (Sorokin *et al.,* 2006). The unique composition of soda lakes makes them ideal for the industrial production of borax, borates, halite, lithium, soda-ash (sodium carbonate), sodium sulphate, trona (sodium sesquicarbonate); uranium (Williams, 1981), and zeolites (for the silica they contain). Soda lakes are also used in aquaculture for the production of algae used by the pharmaceutical industry (Kebede, 1997). Although soda lakes are considered to be one of the most productive

ecosystems on earth they remain an untapped source of novel organisms including polyextremophiles which may conceal a hidden consortium of industrially applicable secondary metabolites. In order to understand the ecology and diversity within these lakes, the structure of the environment and common inhabitants of similar systems will be discussed.

Table 1. Soda lake and desert distribution throughout the world.

North America	
Canada	Manito
United States	Alkali Valley, Albert Lake, Lake Lenre, Soap Lake, Big Soda Lake, Owens Lake, Mono Lake, Searles Lake, Deep Springs, Rhodes Marsh, Harney Lake, Summer Lake, Surprise Valley, Pyramid Lake Walker Lake
Central America	
Mexico	Texcoco
South America	
Venezuela	Langunilla Valley
Chile	Antofagasta
Europe	
Hungary	Lake Fehér
Yugoslavia	Pecena Slatina
Russia	Kulunda Steppe, Tanastar Lakes, Karakul, Arexes plain, Chita, Barnaul, Slavgerod Central Asia: Lakes of Tuva
Asia	
Turkey	Van
India	Lake Looner, Lake Sambhar
China	Qinhgai Hu, Sui-Yian, Heilungkiang, Kirin, Jehol, Chahar, Shansi, Shensi, Kansu
Africa	
Libya	Lake Fezzun
Egypt	Wadi Al Natrun
Ethiopia	Lake Arenguade, Lake Kilotes, Lake Abitjata, Lake Shala, Lake Chitu, Lake Hertale, Lake Metahara
Sudan	Dariba Lakes
Kenya	Lake Bogoria, Lake Nakuru, Lake Elmentieta, Lake Magadi, Lake Simbi, Lake Sonachi
Tanzania	Lake Natron, Lake Embagi, Lake Magad, Lake Manyara, Lake Balangida, Basotu Crate lakes, Lake Kusare, Lake Tulusia, El Kekhooito, Momela Lakes, Lake Lekandiro, Lake Reshitani, Lake Lgarya, Lake Ndutu, Lake Ruckwa North
Uganda	Lake Katwe, Lake Mahega, Lake Kikorongo, Lake Nyamunuka, Lake Munyanayange, Lake Murumuli, Lake Nunyampaka
Chad	Lake Bodu, Lake Rombou, Lake Djikare, Lake Momboio, Lake Yoan
Australia	Lake Corangamite, Red Rock Lake, Lake Werowrap, Lake Chidnup

1.2 Soda Lake environment

1.2.1 Genesis of Soda lakes

Soda lakes per definition can be categorised according to their unique biochemical equilibrium and dominant ions (i.e. Mg^{2+}, Ca^{2+}, HCO_3^-/CO_3^{2-}, Na^-,Cl^- etc.). In soda lakes, the ions within the geographically enclosed endothermic system are derived from inflow water and surrounding structures. A delicate balance is reached with seasonal fluctuations depending on the rainfall

patterns. For example, there is a relative consensus among scientists that the concentration of magnesium and calcium ions within a system is directly proportional to its pH (Gizaw, 1996). When HCO_3^-/CO_3^{2-} and NaCl levels exceed that of Mg^{2+} and Ca^{2+} the latter precipitates as insoluble carbonates resulting in an alkaline brine (see Fig.1). The calcium and magnesium ions act as buffering agents, whereby alkaline carbonates are removed from the system in the form of gypsum, calcite, sepiolite, dolomite and magnesite, resulting in a neutral environment (Grant, 2004). In addition, a surplus of Mg^{2+} ions can skew the balance towards an acidic environment, such as in the Dead Sea (Oren, 2002). In all cases the alkalis outweigh the alkaline metals as the weak acids do the strong acids. The proportion of potassium, sodium, silica and several lesser elements depends on the specific site and sampling conditions (Jones & Grant, 1999).

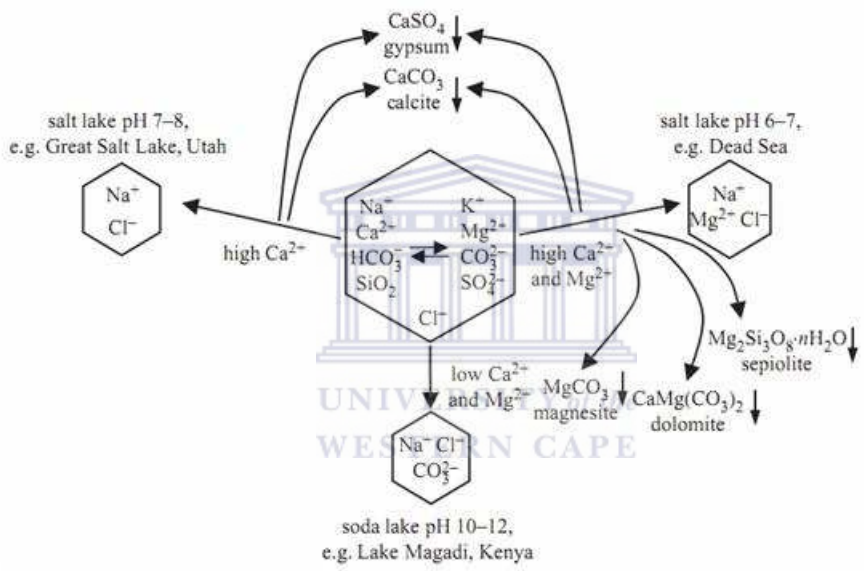


Figure 1. The three predominant geochemical soda lake groups as defined by alkali metal salt and carbonate concentrations. High magnesium and calcium levels cause high precipitation e.g. Dead Sea, whilst low concentrations mimic systems like Lake Magadi. Lakes dominated solely by calcium tend towards NaCl concentrates like the Great Salt Lake. (Grant, 2004)

Salt content is mostly reported as the percentage NaCl rather than the collective concentration of potassium, calcium, sodium and magnesium chloride. This highlights the predominance of NaCl as a naturally occurring ionic inhibitor. However, calcium chloride has been shown to inhibit growth to the same extent, whilst potassium and magnesium chlorides are slightly less effective (Bautista-Gallego *et al.,* 2008). Lake Magadi, the closest lake geochemically to the Ethiopian Rift Valley lakes, lacks magnesium, calcium and potassium ions, as do many saline lakes which might be put forth as the reason for excluding such ions from the definition. However, the potassium levels of Ethiopian Rift Valley lakes are in the order of magnitude 100 times higher than the respective East African lakes. Regardless of the compositional differences,

all soda lakes exist due to the concentration of minerals within a basin. The degree of saturation depends on the location and rock manifestations.

1.2.2 The Ethiopian Rift Valley (ERV)

The East African Rift Valley is composed of three basins that form a triple junction extending from the Afar region of Ethiopia to the eastern branch of Kenya (Alemayehu *et al.,* 2006). These basins are essentially thermal bulges in the earth's surface which stretch and fracture at the edges leaving open faults or rifts. The rift widens over time which is often preceded by volcanic eruptions creating "flood basalts" a sheet of lava covering a large surface area (Le Turdu *et al.,* 1999). These lava sheets, as well as the ignimbrites/tuff (pyroclastic igneous rocks consisting of feldspar and quartz), form part of the current rock formations outlining the ridges of the ERV lakes including Lake Shala and Lake Abijata. While the ridges provide a natural dam, the lava sheets create a source of elements extracted by the interaction with lake water.

Along the Ethiopian Rift, several lakes form a cluster of unusual composite lakes. It is thought that prior to the late Pleistocene period these lakes were part of one large ancestral lake at 1850m elevation. Lake Shala is considered to be the deepest (257m) and largest (crater 12x 15 km) lake. In addition, it is located in the epicenter of the Rift and is encircled by several 'islands' or smaller lakes. These satellite lakes include Lake Abijata and Lake Chitu, which consist of the same ignimbrites as the Shala caldera, although they function as a separate closed system with a complete nutrient cycle. These 'closed' systems refer to the water cycle in each lake (see Fig.2). Each lake is fed by the annual rains and the surrounding rivers, however due to the nature of the basalt sheets there exists almost no groundwater drainage and minimal efflux from streams. The high temperatures and low humidity of the Ethiopian Rift causes water to evaporate, concentrating the minerals and salts present in the lake (see Fig.2).

The flood basalts act as the main source of minerals and characteristically lack sedimentary (Ca^{2+} and Mg^{2+}) deposits. Due to the high productivity of these environments the carbonate levels are high enough to ensure the dissolution of minerals (Na^{+}, K^{+}-silicates) from the surrounding rock; whereas the tectonic activity manifested in the hot springs surrounding these lakes, provide the heat to catalyse these reactions (see Equation 1).

Equation 1 $$CO_2 + H_2O + \text{Na,K-silicates} \leftrightarrow HCO_3 + \text{Na,K} + \text{H-silicates}$$

The clay mineral composites, and to a lesser extent pyroxene, andesite and K-feldspar, produce large amounts of HCO_3^-, whilst the igneous rocks provide a rich source of fluoride (Gizaw, 1996). It is therefore not unusual that Lake Shala is characterized by high CO_3^{2-}, SiO_3^- and

fluoride levels (Alemayehu *et al*., 2006). The effect of this elemental combination on microbial diversity is yet to be investigated. While silica can be utilised by diatoms, studies have shown that large quantities of silica in hydrothermal fluids can result in silification of some prokaryotes (Sayeh, 2010). Silica is thought to absorb biogenically to cell walls, causing the formation of amorphous sheets enclosing bacteria (Sayeh, 2010). In most cases this causes cell death, but recent studies suggest that certain prokaryotes may survive in such a state for a prolonged period of time. The effect of high F concentrations on microbial life has yet to be determined (Zinabu *et al.,* 2002).

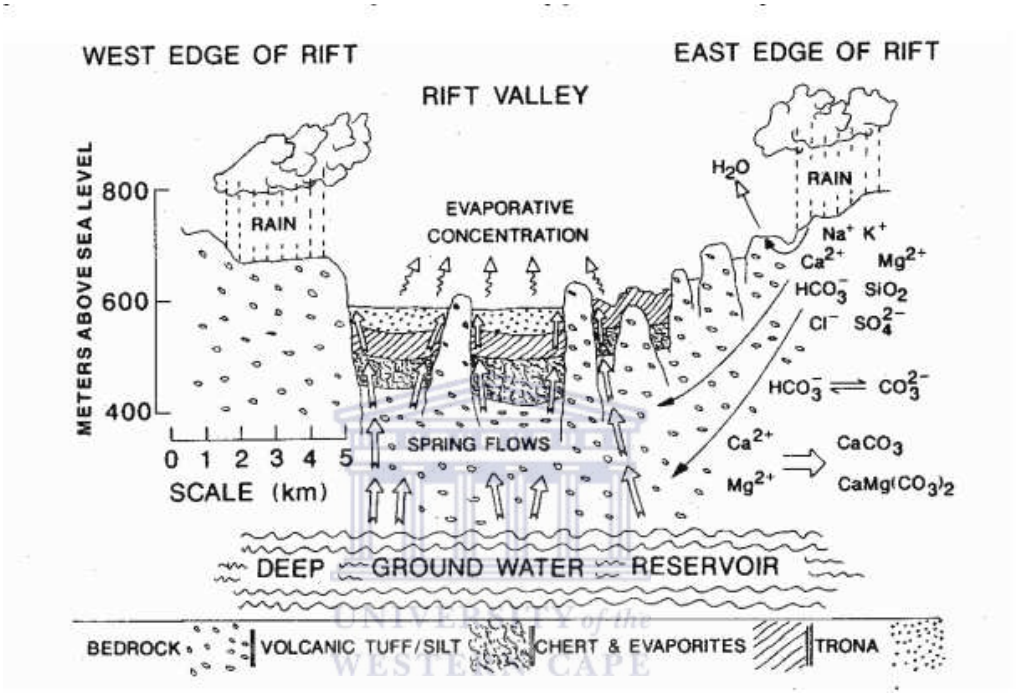


Figure 2. A schematic representation of the proposed mechanisms by which a haloalkaline lake is formed. The model system originally used was Lake Magadi, Kenya. (Grant, 1990)

1.3 Microbial ecology of soda lakes/alkaline biodiversity

1.3.1 Eubacterial placement within a wider ecological niche

Biocoenosis as termed by Karl Möbius (1877) refers to the interaction of organisms within a specific habitat. In extreme environments 'traditional' relationships are slightly modified due to the abiotic stressors inflicted on the inhabitants. Such internal dynamics form one of the constraints that shape the actual species pool colonising a location. The energy within an enclosed system is cycled through primary producers to heterotrophs and chemoorganotrophs, which in turn hydrolyse a range of sugars and polymers for anaerobic fermentors (Fig. 3). The fatty acids produced by anaerobic fermentors are consumed by acetogenic bacteria completing the process. In complex environments such as alkaline soil, the microbial heterogeneity, as well as spatial and temporal changes with regard to generation time, creates what we call "ecological

specialists" (Kassen & Rainey, 2004). These organisms compromise their ability to exist in one niche for a competitive advantage in another. Because of the variety of abiotic stressors (i.e. high temperature; moderate salt and high pH) found in the Rift Valley lakes and the interactions between water- and soil based microsystems, it is imperative to understand the basic components surrounding such an interactive system, as well as the location of various subpopulations.

The high primary productivity (10g carbon $m^{-1}day^{-1}$) of haloalkaline systems supports both oxic and anoxic chemoorganotrophs (see Fig.3). The by-products produced via sugar hydrolysis (i.e. fatty acids) are subsequently consumed by acetogenic bacteria. Conversely, organic material can be utilised by methanogens, which in turn produce methane for methanotrophs. Some anoxygenic phototrophic bacteria release sulphur and sulphate which serve as the electron acceptors for aerobic sulphur oxidizers. Anaerobic fermentors isolated from soda lakes, such as *Tindallia magadii,* (Kevbrin *et al.,* 1998) produce ammonia as a by-product which is in turn metabolised by nitrifiers and methanotrophs, generating nitrite which is co-metabolised with carbon by chemoorganotrophs.

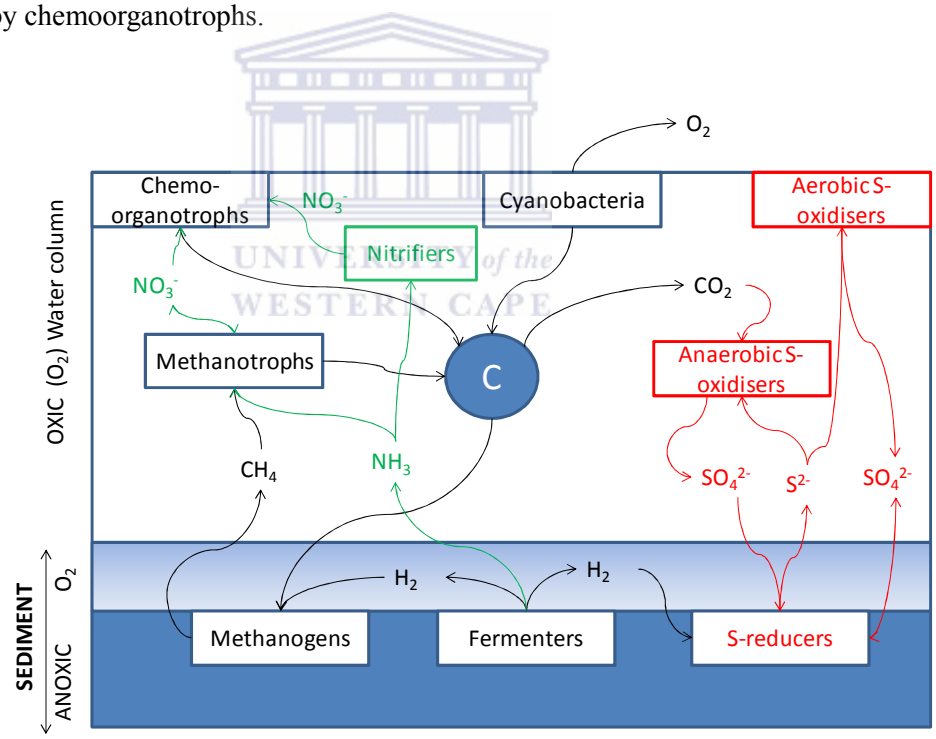


Figure 3. Schematic representation of the nutrient cycles within a soda lake. Black: C-cycle, Red: S-cycle, Green: N-cycle. Modified from Jones *et al*. (1998)

Cultured microorganisms spanning almost all trophic levels (Table 2.) have been isolated from haloalkaline environments world wide (Duckworth *et al*. 1996; Jones *et al*. 1998; Jones & Grant 1999). The following subchapters cover the most prominent eubacteria and their role in the nutrient cycle. This includes published alkaliphilic actinobacteria and their presumed role in the nutrient cycle.

1.3.2. Phototrophic primary producers

The predominant phototrophic primary producers in soda lakes vary depending on salt content, however, some species occur in the majority of soda lakes (Grant, 2006; Bauld, 1981). Alkaliphilic cyanobacteria such as *Arthrospira platensis* (formerly known as *Spirulina platensis*) are thought to drive these systems by providing a fixed carbon source for both alkaliphilic aerobes and anaerobic chemoorganotrophs. These species flourish mostly due to the high ambient temperature, high light intensities and CO_2 supply present in the lakes. They also form part of the main food source for the vast flamingo flocks (*Phoeniconaias minor*) that populate the lakes (Grant, 2006), which in turn provides nitrogen via excrement. Conversely, nitrogen fixation is performed by the heterocystous cyanobacterium *Cyanospira rippkae* and possibly a few chemo-organotrophs.

Several anoxygenic phototrophic bacteria are considered to contribute to primary production rates. Those that have been isolated from soda lakes include species belonging to the genera *Ecotothiorhodospira*, *Halorhodospira, Rhodobacter, Roseinatronobacter* and *Alkalispirillum* (Jones & Grant, 1999). They are also considered to be the major source of nitrogenous compatible solutes (glycine, ectoine and betaine) in these environments (Jones & Grant, 1999). One species has been identified in several lakes along the Rift Valley. Milford *et al.* (2000) was the first to isolate the purple non-sulphur, anoxygenic phototroph by the name of *Rhodobaca bogoriensis* which was originally from Lake Bogoria, Kenya (Milford *et al.,* 2000). Several uncultured clones matching closely with the type strain has also been found in Lake Magadi (Rees *et al*., 2004).

1.3.3. Carbon cycle

a) Aerobes

The majority of the aerobic chemoorgantotrophic isolates have been identified as *γ-Proteobacteria* of which members of the *Halomonodaceae* clade are widely represented (Horikoshi *et al.,* 2010; Jones *et al.,* 1998). Besides their ability to hydrolyse polymeric carbohydrates and proteins (with a preference for lipoproteins), they can reduce nitrate to nitrite which serves as an energy source for nitrite oxidising bacteria (Fig.4). Equally as important are the Gram-positive species belonging to both the low and high G+C lineages, with the low G+C bacterial lineages dominated by *Bacillus* species. Several alkaliphilic species have been isolated from Lake Bogoria and Lake Nakuru (Kenya) which are closely related to *Bacillus alcalophilus*, *Bacillus agaradhaerens* and *Bacillus clarkia*. These bacilli are responsible for the breakdown of a large variety of polymeric carbohydrates in soda lakes. High G+C isolates include several

members of the phylum *Actinobacteria,* including a novel *Streptomyces* species that can grow at pH 10.0 (Jones *et al.* 1998), a novel *Dietzia* species (Duckworth *et al.,* 1998), as well as the type strain of the novel genus *Bogoriella* (Groth *et al*., 1997). *Nesterenkonia aethiopica* was isolated from Lake Abitjata and has an optimal growth range between 30-37°C, pH 9 and 3% NaCl (Delgado *et al.,* 2006). Strains distantly related to *Arthrobacter* and *Terrabacter* type strains (Jones *et al.,* 1998) have also been isolated from soda lakes. As chemolithoorganotrophs these actinobacterial species play an important role in carbon cycling and secondary metabolite production. Their contribution to the adaptive mechanism employed in soda lakes is discussed in section 1.5.

Archaea form part of the aerobic chemoorganoheterotrophs in soda lakes and mainly assimilate amino acids, peptides, organic acids and carbohydrates (Cavicchioli, 2007). Archaeal strains isolated from both the Rift Valley soda lakes and the Dead Sea fall within the order *Halobacteriales* (Cavicchioli, 2007) although several culture-independent studies have detected strains related to the suborders Crenarcheota and Euryarcheota (Wani *et al.,* 2006).

b) Anaerobic alkaliphiles

Anaerobic chemoorganotrophs can form large populations in soda lakes and colony counts in the order of 10^6 cfu/ml have been reported (Seckbach, 2000). Among the obligate anaerobes, most characterized isolates are associated with *Clostridium* group XI (Jones *et al.,* 1998). These isolates ferment a range of sugars producing iso-valeric acid and small amounts of acetic acid. Other anaerobic fermentors include *Spirochaeta* species, which have been detected in both Lake Magadi and Lake Khatyn (Zhilina *et al.,* 1996), and *Methanosalus zhilinae,* which was originally isolated from Lake Magadi (Zhilina & Zavarzin, 1994). The acetate produced by these organisms is broken down by homoacetogenic bacteria such as *Natroniella acetogena*, which was originally isolated from a soda lake in Asia (Zhilina *et al.,* 1996.). Theoretically, the residual lactate could subsequently be used by *Lactobacilli* or propionibacteriums, although both groups have only been detected in soda lakes by culture-independent techniques (Duckworth *et al.,* 1996).

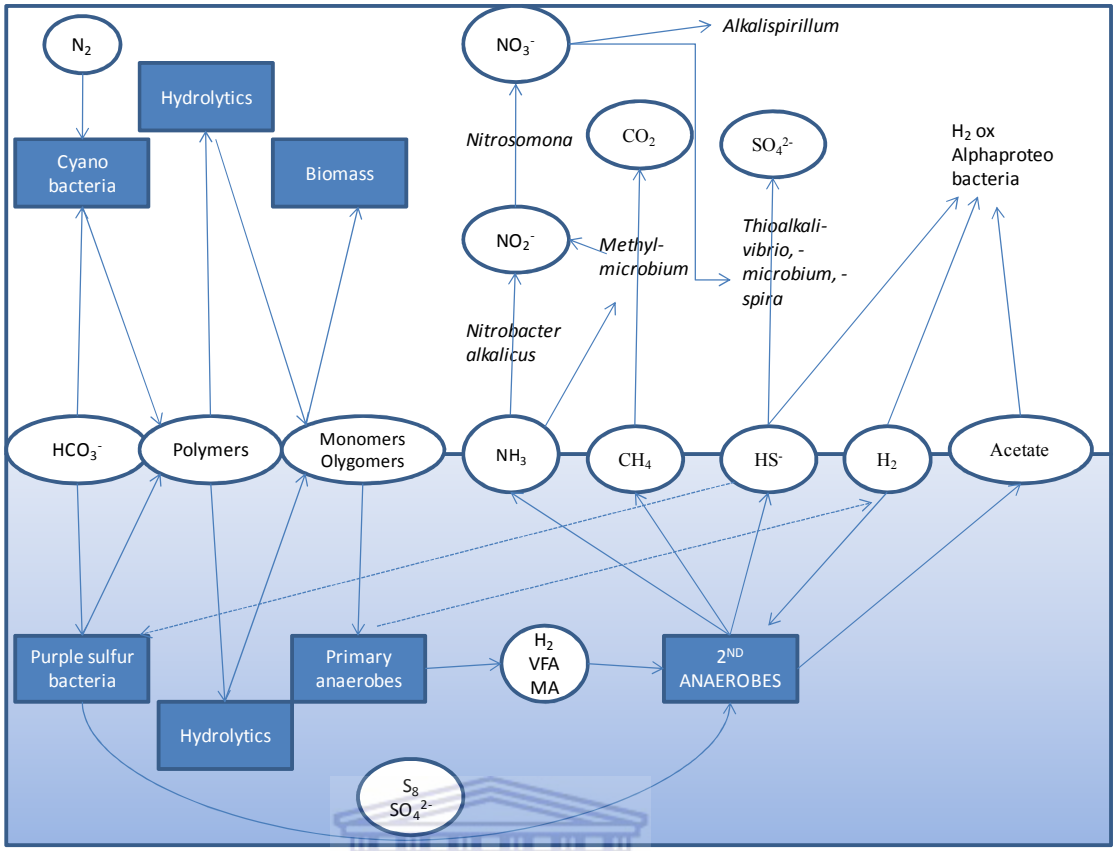


Figure 4. Schematic representation of the relationship between aerobic denitrifying chemolithotrophs in soda lakes. VFA: Volatile Fatty acids/alcohols, MA: Methylamines (Sorokin & Kuenen, 2005a)

The majority of methanotrophs isolated from soda lakes belong to the phylum *γ-Proteobacteria* and exhibit a low salt tolerance (all below 1.5M). These include *Methylomicrobium pelagium* and *Methylomicrobium* sp. AMO1 isolated from Lake Tuva (central Asia) and a Kenyan soda lake (Sorokin *et al.,* 2000). A recently discovered novel species belonging to the same genera has been isolated from Lake Chitu (Ethiopia) and is undergoing further characterization (Mørkeseth, pers comm. 2011).

Table 2. Prokaryotes known to grow under alkaline conditions

EUBACTERIA	Strains and species
Cyanobacteria	*Arthrospira platensis, Cyanospira rippkae, Synechocystis* sp., *Synechococcus* sp., *Phormidium* sp.
Gram-positive (High G+C)	
ACTINOBACTERIA	
Corynebacteria	*Bogoriella caseilyticus, Dietzia natronolimnaea*
Micrococci/*Nesterenkonia*	Lake Bogoria isolate 69B4, *Arthrobacter* sp.[b]
Streptomyces/Nocardiopsis	Lake Nakuru isolate 11AG8
Intraspotangia	*Terrabacter* sp.[b]
Gram-positive (low G+C)	
Bacillus/ Clostridium	*B. agaradhaerens, B. clarkii, B. cohnii, B. alcalophilus, B. halmapalus, B. halodurans, B. horikoshii, B. gibbonsii, B. haloalkaliphilus, B. hortii, B. vedderi, AmphiBacillus fermentum, AmphiBacillus tropicus, Alkalibacterium olivoapovliticus*[a]*, AnoxyBacillus pushchinensis*[a]*, C. paradoxum*[a]*, C. thermoalcaliphilum*[a]*, Anaerobranca gottschalkii, Tindallia Magadiensis, Natronincola histidinovorans*
Haloaerobes	*Natroniella acetigena, Halonatronum saccharophilum, Thermosyntropha lypolyticum*[a]*, Anaerobranca horikoshii*
Proteobacteria	
Sulfur oxidizers	*Thioalkalimicrobium sibericum, Thioalkalimicrobium aerophilum, Thioalkalivibrio versutus, Thioalcalivibrio nitratus, Thioalkalivibrio denitrificans, Thioalcalimicrobium cyclum, Thioalkalivibrio jannaschii*
Nitrifiers	*Nitrobacter alkalicus*
Sulfate-reducing bacteria	*Desulfonatronovibrio hydrogenovorans, Desulfonatronum lacustre*
Anoxygenic phototrophs	*Ectothiorhodospira mobilis, Halorhodospira halophila, Thiorhodospira siberica, Roseinatronobacter thiooxidans, Rhodobacter bogoriensis, Alkalispirillum mobile, Ectothiorhodosipa vacuolata, Ectothiorhodospira halochloris, Ectothiorhodospira haloalkaliphilus*
Other	
Halomonads	*Halomonas magadii, Halomonas campisalis, Halomonas desiderata*[a]*, Halomonas pantellenriensis*[a]
Methylotrophs	*Methylobacter alcaliphilus, Methylomicrobium sp. AM01*
Entrics/Vibrios/Aeromonads	*Lake Nakuru isolate 20N1*
Spirochaetes	*Spirochaeta alkalica, Spirochaeta asiatica, Spirochaeta africana*
Thermotogales	*Thermopallium natronophilum*
ARCHAEA	
Halobacteria	*Halorubrum vacuolatum, Natrialba magadii, Natronobacterium gregoryi, Natronomonas pharaonis, Natronococcus occultus, Natronococcus amylolyticus, Natronorubrum bangense, Natronorubrum tibetense*
Thermophiles	*Thermococcus alcaliphilus*[a]*, Thermococcus fumicolans*[a]
Methanogens	*Methanolobus oregonensis, Methanosalsus zhilinae, Methanobacterium subterraneum*[a]*, Methanobacterium alcaliphilum*

[a] *Indicates alkaliphiles that have not yet been isolated from soda lakes*
[b] *Indicates a genus that has only been detected by culture-independent methods. Uncultured clones.*

1.3.4. Sulphur cycle

The sulphur cycle in soda lakes are mainly driven by purple sulphur bacteria and hydrogenotrophs that utilise polysulfides as electron donors (Sorokin *et al.,* 2005b). The products of the primary anaerobes are used as energy sources, and methane as well as S_2O_3 species are released into the water column (see Fig.4). In the case of Lake Magadi, a sulphate reducing bacteria, *Desulfonatronovibrio hydrogenovorans* was isolated at a NaCl concentration of 3% (pH 9.5) (Zhilina *et al*., 1997). The only other known isolate is *Desulphonatronum lacustre* (Pikuta *et al*., 2003) which was isolated from a Siberian soda lake. Recent studies on planctomycete metabolic activities suggest they might play a role as anaerobic sulphate reducers (Spring *et al*., 2007). While *Planctomyces* species are yet to be isolated from soda lake environments, related sequences have been identified through molecular fingerprinting techniques. The reduced sulphur compound can be employed by both aerobic and anaerobic sulphur oxidizing bacteria (SOB) that form part of the bacterial 'filter' in the sediment and water zones. An estimated hundred SOB strains have been isolated from soda lakes belonging to the new genera of γ-*Proteobacteria*: *Thioalkalimicrobium, Thioalkalispira* and *Thioalkalivibrio* (Sorokin & Kuenen, 2005a). Sulphur-oxidizers can reach viable cell count of 10^3-10^8 cells/cm^3 in soda lakes (Horikoshi *et al*., 2010). The majority of Rift Valley SOB isolated to date belong to either the genus *Thioalkalimicrobium* or *Thioalkalivibrio* (Baumgarte, 2003; Jones & Grant, 1999; Rees *et al*., 2004). Their pH optima range between 9.5-10.2 with salt limits in the range of 1.5M Na^+ and 4.3M Na^+. A smaller number of *Halomonas deleya* have also been detected in some soda lake samples from Siberia and Kenya (Sorokin *et al.,* 2001).

A separate, but industrially relevant group of γ-*Proteobacteria* are able to utilise thiocyanate as an electron donor converting it to cyanate and finally to ammonium, CO_2 and sulphate. The two strains isolated from soda lake brines found to possess this ability are *Thioalkalivibrio thiocyanoxidans* and *Thioalkalivibrio paradoxus* (Horikoshi *et al*., 2010).

1.3.5. Nitrogen cycle

Since nitrite is less toxic at high pH, it can exist at higher concentrations, which is necessary for sustainable growth due to the low energy yield of the reaction (Sorokin & Kuenen, 2005a). This finding accounts for the presence of *Nitrobacter alkalicus* and related species in soda lakes (Sorokin *et al*., 1998). Hydrolytic microorganisms produce ammonium under aerobic conditions which becomes volatilized and is released into the air. One unique organism isolated from Lake Magadi, *Tindallia magadiensis*, is an anaerobe capable of ammonifying amino acids, although the energy yield of this reaction is insufficient to serve as the sole energy source (Grant, 2006). Copiotrophic fermenters (i.e. methane oxidising bacteria) oxidise ammonium to produce NH_3 and H_2, which in turn is used by methanotrophs, nitrifiers and hydrogenotrophs (Seckbach, 2000). The main ammonium oxidising bacterium detected in both Inner Mongolia and in Lake

Wadi Al-Natrun was identified as *Nitrosomonas europea* (Horikoshi *et al*., 2010). Both the nitrifiers and methanotrophs produce nitrite as a by-product which the chemoorganotrophs exploit. In most cases nitrite substitutes as the preferred energy source and as a syntrophic relayed process several bacteria could be involved in nitrification. Two species identified as *Nitrobacter alkalicus* and *Nitrobacter winogradskyi* have been isolated from soda lake environments (Sorokin *et al*., 1998) suggesting an active nitrite oxidizing community.

Thioalkalivibrio strains form part of both the sulphur and nitrogen cycles depending on the availability and preference of individual species. One extensively studied interaction is the denitrification of nitrate to nitrite via two γ-*Proteobacteria*, *Thioalkalivibrio nitratireducens* and *Thioalkalivibrio denitrificans* (Sorokin & Kuenen, 2005b). The former is known for its degradation of nitrate to nitrite whereas the latter converts nitrite to NO_2. Despite the fact that some interaction takes place, obligate syntrophy (metabolically completely dependent on another species) is a complex relationship to study through classical culture-based approaches and in such cases an interdisciplinary approach is warranted.

1.4. Actinobacteria

The class *Actinobacteria* (phylum 'Actinobacteria') is considered to be the major Gram-positive, high G+C taxa consisting of five subclasses and nine orders (Goodfellow & Fiedler, 2010). The relationship between the different actinobacterial lineages have been the subject of a long debate and several theories exist regarding the evolutionary history of the phylum (Ventura *et al*., 2007; Stackebrandt & Schumann, 2006; Chater & Chandra, 2006). Prauser was the first to propose a monophyletic evolution of actinobacteria from pleomorphic rod-shaped coryneforms to nocardioforms which subsequently evolved into the spore-forming actinobacteria (Prauser *et al*., 1970). A second hypothesis is the polyphyletic origin where fungi, actinobacteria and mycobacteria share a common ancestor which later morphed into Bifido- and Lactic acid bacteria (Embley & Stackebrandt, 1994). Regardless of their origin, some of the traits inherently associated with the phylum differ widely within the different actinobacterial lineages. Actinobacteria can be divided into four subclasses: *Acidimicrobidae, Coriobacteridae, Rubrobacteridae* and *Actinobacteridae*. The first two contains only one order although the last two contains three each. The most published of which is the *Actinomycetales* order (Ventura *et al*., 2007; Stackebrandt & Schumann, 2006).

Traditionally, traits such as antibiotic production and cell wall characteristic (menaquinones, mycolic acids and peptidoglycans) were used to differentiate between the different suborders. However, any phenotypic/cellular trait on its own is not sufficient to classify a strain. With the emergence of molecular microbiology the analysis of the 16S rRNA gene has superseded

traditional phenotypic characterization. The current species threshold used by taxonomists is that 99% 16S rRNA sequence similarity denotes different species.

Actinobateria are aerobic/microaerobic, filamentous bacteria that form lichenoid, leathery or butyrous colonies on agar. The assortment of strains in its 18 suborders varies in terms of morphology and chemotaxonomy. The main classifying feature of the phylum is the high G+C content of their genomic DNA, which ranges from 53% in some *Corynebacteria* to 78% in some *Streptomyces* species (Korn-Wendisch & Kurtzner, 1992; Ventura *et al.,* 2007). Actinobacteria, specifically streptomycetes, have been of industrial importance for decades due to their secondary metabolite production (Streit & Schmitz, 2004). Streptomycetes colonise most terrestrial and aquatic environments. This is done with the help of extensive branching mycelia which are seen as the vegetative organ, whilst the hyphae extend from the reproductive part of the colony forming chains of hydrophobic spores. The production of anthrospores confers spore-forming actinobacteria with the ability to withstand drought, as well as nutrient deprivation, whilst providing a means through which wind dispersal is possible. Conversely, in sporangia-forming actinobacteria such as the genera *Actinoplanes*, hydrophilic spores are specially adapted for water dispersal (Whalen & Sampedro, 2010). Multilocular sporangia are mainly formed by nitrogen fixing *Frankia* species and *Geodermatophilus* (Saikia, 2008). Additional genus specific characteristics include spore pigmentation, the production of diffusible pigment and secondary metabolite profiles (Maheshwari *et al*., 2010). Nocardioform actinobacteria are non-spore-forming, aerobes and include the genera *Dietzia, Gordonia, Mycobacterium, Nocardia* and *Rhodococcus*. This group's metabolic capabilities include the degradation of several hydrocarbons - aliphatics, aromatics and xenobiotic pollutants including organochlorines (Whalen & Sampedro, 2010).

Thermophilic, halophilic and alkaliphilic actinobacteria exist among several genera (Cai *et al*., 2009; Kurapova *et al*., 2008; Norovsuren *et al*., 2007). Table3. summarises some of the actinobacteria which fall into the three categories. Colonisation of hot springs, such as those found surrounding Lake Shala, requires the strain to survive at temperatures of up to 90°C. Several thermophilic actinobacteria have been isolated including *Thermobifida fusca*, *Thermobifida alba* (Zhang *et al*., 1998)*, Acidothermus cellulolyticus, Streptomyces clavuligerus, Microbispora, Thermoactinomyces thalpophilus* and more recently *Thermoleophilum album*. Actinobacteria have been isolated from Hot springs in India (Malkawi *et al.,* 2010), Taiwan, Jordan (Wani *et al.,* 2006) and Turkey (Song *et al.,* 2009). However, the combination of alkaline water and high temperature or haloalkaline environments narrows the list of cultured actinobacteria considerably.

Table 3. Extremophilic actinobacteria and their origins

Species	Origin	Tolerance
Alkaliphilic		*pH tolerance*
Nitriliruptor alkaliphilus	**Soda lake sediments**	**8.4-10.6**
Nocardiopsis valliformis	**Alkaline lake soil in China**	**8.0-14.0**
Streptomyces deccanensis	**Gulbarga, Karnataka Province, India**	**8.0-10.5**
Kucuria aegyptia	**Saline, alkaline Egypt**	**9.0-10.0**
Streptomyces sodiiphilus	**Chaka salt lake, China**	**9.0-10.1**
Nocardiopsis alkaliphila	**Eastern desert of Egypt**	**9.5-10.0**
Nocardiopsis metallicus	**Alkaline slag dump, Germany**	**8.5-10.5**
Dietzia natronolimnaios	**East African soda lake**	**9.0-13.0**
Streptomyces sannanensis	**Gujarat Province, India**	**7.0-10.0**
Bogoriella caseilytica	**East African soda lake**	**9.0-10.0**
Streptomyces tanashiensis	**Loktak Lake, India**	**6.5-9.5**
Streptomyces rochei	**Bay of Bengal**	**6.0-10.5**
Halophilic species		*Salt tolerance*
Nocardiopsis litoralis	**Sea anemone, South China**	**1-15%**
Streptomonospora amylolytica	**Salt lake, South-west China**	**5-15%**
Streptomonospora flavalba	**Salt lake, South-west China**	**5-15%**
Saccharopolyspora qijiaojingensis	**Salt lake, North-west China**	**6-22%**
Amycolatopsis marina	**Ocean sediment, South China sea**	**8-15%**
Thermobifida halotolerans	**Salt mine, south-west china**	**0-15%**
Nocardiopsis aegyptia	**Marine sediments, Egypt**	**5-30%**
Nocardiopsis halotolerans	**Kuwait salt marsh soil**	**0-15%**
Saccharopolyspora halophila	**Kuwait salt marsh soil**	**10-25%**
Nocardiopsis kunsanensis	**Saltern in Kunsan,Korea**	**10-20%**
Streptomyces clavuligerous	**Saline soil of Mithapur, India**	**0-15%**
Thermophilic Species		*Temp tolerance (°C)*
Acidothermus cellulolyticus	**Hot spring, Yellow stone national park**	**37-70**
Amycolatopsiseurytherma	**Arid soil, India**	**25-55**
Thermoactinomyces sacchari	**Hay, straw and bagasse**	**35-65**
Thermoactinomyces vulgaris	**Moldy bagasse**	**35-65**
Streptomyces thermocarboxydovorans	**Soil**	**20-55**
Streptomyces thermovulgaris	**Cow manure**	**25-55**
Streptomyces thermodiastaticus	**Soil**	**20-55**
Streptomyces thermonitrificans	**Soil, Bombay India**	**25-55**
Streptomyces thermoviolaceus subsp. *thermoviolaceus*	**Soil**	**20-55**

Compilation of (Lacey, 1971; B. Kim *et al*., 2002; Sahin *et al*., 1999; Kim *et al*., 1999)

1.5. Theoretical adaptive mechanisms: Cellular evolution

Soda lakes form an environment that can be described in terms of its low water activity value. Other than fresh water habitats, all aquatic environments contain large quantities of water molecules which are bound and often inaccessible to microorganisms (Grant, 2004). According to Grant (2004) water activity can be defined as the effective water content expressed as its mole fraction. In a haloalkaline environment, microorganisms need to avoid water loss via osmosis whilst maintaining turgor pressure to allow adequate growth. This can be accomplished via two strategies: either by a 'Salt-in' strategy; which encompasses the generation of a counterbalance of inorganic ions (usually KCl) to achieve osmotic stability; or a 'Salt-out' strategy which implies the production or accumulation of low molecular weight organic compounds that have osmotic potential (also known as osmolytes). The latter strategy is the dominant strategy in soda lakes and is utilised by several eukaryotic plant and algal species. The reason being that as a weaker electrolyte Na_2CO_3 (soda) reduces the osmotic burden of secondary metabolite production giving these organisms an advantage over those living in NaCl brines (Sorokin, 2005a). The 'salt-out' theory forms the excepted hypothesis for halotolerant osmoadaptation. This argument is strengthened by the apparent overproduction of osmolytes by model organisms (i.e. *Salinivibrio costicola*; *Halomonas elongata* and *Halomonas israelensis*). Each strategy is accompanied by four osmoadaptive responses: cytoplasmic pH acidification; cell membrane permeability control; stability of proteins and altered cell wall rigidity.

The 'salt-out' strategy combines the import and synthesis of osmolytes with the export of salts out of the cell. This mediates the problem of maintaining an iso- or hyperosmotic cytoplasm within its internal cellular environment. Osmolytes are polar, normally uncharged zwitterions which under cellular conditions influence proton flow and thus the pH within the cytoplasm, without inhibiting or interfering with enzyme activity (Grant *et al.,* 2004). Proteins form what is known as a hydration shell which, due to surface tension, propels osmolytes from the dense water around proteins (Galinski, 1995). This causes the osmolytes to react with the bulk water phase, thereby minimizing water entropy by affecting the cohesion between molecules and the reinforcement of hydrophobic bonds (Nieto *et al*., 1998; Galinski, 1995). Examples of such compounds are: polyols such as mannitol (which is synthesized *de novo*), betaines such as trimethylammonium compounds (or in the case of terrestrial environments glycine betaine which is accumulated from the environment), aminoacids (specifically proline, glutamate and glutamine) and ectoines (which are the most abundant osmolytes in alkaliphilic organisms). Most halophilic/halotolerant heterotrophic bacteria cannot synthesize glycine betaine and its precursor choline *de novo,* and it is therefore transported by means of membrane proteins into the cell. Thus far a single actinobacterial species, *Actinopolyspora halophila,* has been shown to

synthesize it.

In the case of alkaliphilic bacteria several responses are essential for the maintenance of cellular function under high osmotic stress. These include the alteration to extracellular, membrane-bound and intracellular enzymes, as well as cell envelopes and stress response proteins. An example of a common adaptation is the decrease in hydrophobic amino acids visible in the pyruvate kinase produced by *Salinivibrio costicola* (Nieto *et al*., 1998). The model organism *Micrococcus halophilus,* a known producer of extracellular enzymes, generates halophilic amylases and nucleases which reach optimal activity at 2M NaCl and 1-4M NaCl, respectively (Bowers *et al*., 2009). Both enzymes however are reliant on the ability of the cell envelope to adjust to a fluctuating environment such as lake sediments. Essentially an increase in salinity would result in an increase in charged phospholipids (phosphatidylglycerol or cardiolipin in the case of halotolerant bacteria) and a decrease in uncharged phosphatidylethanolamine resulting in a shift in charge density (Nieto *et al*., 1998). Phosphatidylglycerol tends to suppress the formation of a hexagonal II phase and maintains a lipid bilayer thereby counteracting the increase in salinity. A problem does arise when the organism is exposed to both high temperature and salinity; since the latter increases the production of non-desirable phosphatidylethanolamine (Bowers *et al*., 2009).

The Archaea (salt-in strategy) and Eubacteria (salt-out strategy) occupy essentially two different niches and thus have little overlap. A study on a Spanish solar saltern revealed that the NaCl concentration range where both taxa overlapped was between 25-32%. This is likely to be the main reason why most research on soda lakes is focused on bacteria and not on archae (Rodriguez-Valera *et al*., 1981), the exception being hot springs that reach high enough temperatures and saline lakes such as Lake Magadi (Kenya). They concluded that bacteria tend to dominate in low temperature, -salt and -nutrient environments whereas archaea thrive in high temperature and salt habitats (Nieto *et al*., 1998).

1.6 Ribosomal RNA genes and molecular microbial ecology

1.6.1 The role of taxonomy in phylogenetic studies

Modern taxonomic studies comprise of two important concepts: identification and description of an organism (α taxonomy) as well as the classification of such an organism (β taxonomy) on the basis of phylogeny or phenetic characteristics (Roselló-Mora & Kampfer, 2004). Historically phenetic classification played an important role in both the categorisation of functionally different groups as well as the identification of industrially relevant organisms. The subjectivity of classifying strains by means of phenetic characteristics encouraged the development of a robust phylogenetic system that could be less variable. Both approaches have directed the flow

of information from microbial ecology through to the construction of a framework for the analysis of microbial diversity. Extreme environments, such as found in the Ethiopian Rift Valley, provide a unique glimpse into the understanding of evolutionary 'hot spots' and its contribution to our understanding of taxonomy and phylogenetic cladistics. It is therefore important to discuss the assumptions on which it is based and the factors which influence its propagation.

The species concept and speciation

Taxonomical classification is subject to three influential concepts namely: the species concept, the rate of speciation (species evolution) and the extrinsic environment (which induces change in the form of evolution). All three concepts affect both the number of species detected (richness) and the lineages present in a particular environment, by means of defining what can be classified as an unique operational taxonomic unit (OTU) or phylotype. The term OTU is used in order to define a unit from which community diversity can be analysed. Several reviews have addressed the change in definition of each concept therefore only the currently excepted theories will be discussed briefly.

The species concept encompasses four possible theories (Roselló-Mora & Kampfer, 2004). The first is termed the biological species concept (BSC) which was first proposed by Mayr in 1942, and it essentially states that a species can be defined as "groups of interbreeding natural populations that are reproductively isolated from other such groups". The second is the phenetic species concept which has been incorporated into all fundamental microbiology. This classification scheme is however limited to the taxonomic characterization of culturable organisms which is estimated to account for only 1% of the microorganisms inhabiting soil environments (Torsvik *et al*., 1990). The third is the cohesion species concept that defines a species by "the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms (among them are gene flow, natural selection…)". The fourth theory is that of the phylogenetic species concept and encompasses two formulations: (i) the monophyletic species concept which states that a species can be regarded as "the least inclusive monophyletic group definable by at least one autapomorphy" and (ii) the diagnostic species concept that defines a species as "the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent" (Claridge *et al*. 1997). The phylogenetic species concept is the one currently used to define species.

Several definitions of a bacterial species have been proposed (Rosselló-Mora & Amann, 2001), however the current definition stipulates that a new species is classified by a less than 70%

similarity in DNA-DNA reassociation experiments or have a ΔT_m of 5°C or less accompanied by a unique phenotypic profile (Wayne *et al*. 1987). This 70% threshold has been shown to correspond to 97.5% 16S rRNA gene similarity (Stackebrandt & Goebel, 1994). Speciation, as defined by the separation of one species into two unique/distinct genomic strains, is intertwined with the definition of a species and the classification of novel isolates. During the course of speciation both intrinsic (i.e. translation apparatus) and extrinsic factors (i.e. environment) contribute to the survival of a mutated organism and this classification system is based on a comparison of genetic modifications.

The macromolecules (DNA, RNA and protein) used to identify a genetic relationship between two organisms are subject to change over time. This is the major cause of debate regarding species classification within the phylogenetic species concept. When focussing on a single species, there are several intrinsic factors that influence medium and long-term evolutionary events. One example is the haploid nature of prokaryotes, which makes them ideally suited for the expression of new genes if the necessary translation and transcription mechanisms are in place. This, combined with the ubiquitous nature of horizontal gene transfer (HGT), may account for the divergence of the major lineages. It is known that HGT can occur between organisms through conjugation, transduction and transformation, and HGT has been shown to have occurred in a variety of proteins (Woese, 2004).

It is not yet known how much variability between related proteins is due to natural mutation (Rappe & Giovannoni, 2003) and how much is due to artefacts such as chimeric genes (Wang & Zhang, 1997) and paralogous rRNA operons (Yap *et al*., 1999); but evidence of HGT among actinobacteria is rare and mostly limited to functional proteins other than the translation apparatus. It is however worth noting that many proteins believed to be unique to actinobacteria are in fact present in the proteobaterium, *Magnetospirillum magnetotacticum*, suggesting that the phenomenon does occur (Ventura *et al.,* 2007).

Extrinsic factors are less traceable and niche specific. Such factors include: dispersal (water and wind), biological (i.e. the ability to produce antibiotics or survive desiccation), chemical and physical factors. It has been proposed that environmental pressure might be responsible for a variety of primary and secondary structural differences in macromolecules (Zuckerkandl & Pauling, 1965). This is also the motivation behind the search for "extremozymes" derived from organisms that are adapted to extreme environments (Hough & Danson, 1999). Although these events might cause small structural changes, generally they do not cause a loss of function and are therefore seen as "selectively neutral" (Woese, 1987).

On the grounds of taxonomic classification, it is important to note the contribution of intrinsic/extrinsic factors such as HGT to the formulation of a species concept and phylogenetic relationships. Although the current species concept is far from ideal, it is still evolving and therefore our understanding of microbial relationships is as well. A relatively new approach has been added to the long list of classification systems and due to its use in actinobacterial taxonomy, it will be discussed further.

1.6.2. A polyphasic approach towards actinobacterial taxonomy

The term 'polyphasic taxonomy' was initially coined by Colwell (1970) to describe the delineation of taxa. The constant addition of new molecular techniques and therefore evolution in taxonomy has given rise to the current classification system. Classical taxonomy is no longer sufficient and new approaches must incorporate genotypic, phenetic and phylogenetic methods to the identification and classification of organisms.

The development of a classification system for actinobacteria has followed a slightly different path to that of other phyla. The delineation of taxa underwent several alterations due to the vast array of morphological and biochemically diverse lineages, as well as the industrially driven exploitation of novel isolates. The initial discovery of antibiotic producing streptomycetes in the 1940s resulted in the overspeciation of the genus, which increased from a mere 40 to 3000 described species within twenty years (Anderson & Wellington, 2001). The International *Streptomyces* Project (ISP) was founded to provide a means of introducing standard criteria to classify *Streptomyces* species thereby reducing the number of synonymous species/strains. The initial standard criteria involved spore colour, spore chain morphology, surface ornamentation, the colour of substrate mycelium and soluble pigments, the production of melanin pigment, as well as utilisation of various compounds (Shirling & Gottlieb, 1966). Although this resulted in the re-description of 450 species, the ISP only provided a standard means by which species could be described and does not provide a universal identification scheme.

Originally, the problem of overspeciation was addressed by including chemotaxonomy to characterization protocols. Most of the accompanying methods and genotypic approaches that have been added since are summarised in Table 4. All genotypic and phenotypic methods are limited by the taxonomic level to which they can be applied. In the case of streptomycetes, species can be distinguished based on spore morphology and pigment production (phenotypic characteristics). This is done via the analysis of the convoluted projections found on the surface of the conidial wall together with the arrangement and shape of the spore branching structure (Anderson & Wellington, 2001).

Additional phenotypic approaches include cell wall composition (Lechevalier & Lechevalier, 1970), FAME analysis, whole cell analysis (Sanglier *et al*., 1992), protein profiling (Good-fellow & O'Donnell, 1993), phage typing (Wellington & Williams, 1981), ELISA (Kirby & Rybicki, 1986) and rapid biochemical assay for the utilisation of 4-methyl-umbelliferone-linked substrates (Good-fellow *et al*., 1987b). Cell wall components play an important role in characterizing Gram-positive bacteria. Both cell wall bound peptidoglycan and teichoic acid are used as analytical tools for genus delineation.

Cellular fatty acids have long been used for strain characterization due to its dominant presence in lipopolysaccharides, polar- and sphingolipids. Individual species are defined by the variability in chain length, double bonds and substituent groups. With regards to *Streptomyces* species, cellular fatty acids are mainly straight chain; iso and anteiso-branched fatty acids with 14-18 C side chains. Some streptomycetes contain hydroxylated methyl esters, which helps to distinguish strains at the species level. This technique does not allow the delineation of all *Streptomyces* species and should be limited to genus level profiling within this specific phylum (Anderson & Wellington, 2001). Whole cell analysis, as with FAME, is sensitive to growth media and incubation alterations which makes it a less favoured typing technique, although it has been used to distinguish between species (Sanglier *et al*. 1992).

Protein profiling via 1D and 2D gel electrophoresis has proven to be an effective method for distinguishing between strains and species. Its high correlation with DNA-DNA hybridization also favours its use due to the lower cost. Its use however is not required for the characterization of a novel taxon. Serological classification through ELISA is extremely accurate and can distinguish between species. The only limitation is that low nutrient environments will cause an alteration in the expression of surface antigens and therefore limit the detection level. Phage profiling can distinguish at the species and genus levels due to the occurrence of both polyvalent and species-specific phages. In industrial settings the use of phages for characterization is mainly dismissed due to hazard of possible downstream contamination in processing plants.

BIOLOG plates and API strips provide a fast and relatively standardised protocol for determining phenotypic traits. A small number of tests have shown to exhibit false negatives and due to the large percentage of streptomycetes that produce dark extracellular pigments, it is encouraged to wash cells prior to inoculation to limit false positive results. Ideally, classical methods are still employed when characterizing a novel strain. While this may limit the range, of tests that can be conducted, the results of these tests are likely to be more accurate and reproducible.

Genotypic approaches have become more prevalent due to easy accessibility and rapid automation. These include G+C base content, DNA-DNA hybridization (Labeda, 1992), rRNA homology, low frequency restriction fragment analysis (Beyazova & Lechevalier, 1993) and other PCR-based DNA typing methods.

Generally DNA-DNA hybridization is regarded as the reference method for distinguishing between strains that are closely related or share a higher than 99% 16S rRNA gene homology (Stackebrandt & Goebel, 1994). The method is based upon the reassociation of two individual candidate's genomes whereby an experimentally induced mispairing is measured via the hybrid DNA's thermal stability. A 1-2.2% decrease in thermal stability translates to a 1% mispairing. The concept of a species being defined by a 70% DNA-DNA relatedness or less than 5°C Tm limit is therefore an indication rather than an absolute. A subclause often overlooked is that phenotypic and chemotaxonomic features should agree with this definition. The practical obstacle is of course that discrepancies exist between the different methods used to determine the related level.

rRNA homology has gained considerable weight as a phylogenetic tool and is discussed in great detail in the following sections, however a short mention should be made of its importance to actinobacterial, and particularly to streptomycete, taxonomy. Within the genus *Streptomyces,* variable regions that are both genus-specific (α and β regions) and species-specific (γ regions) have been identified in the 16S rRNA gene (see Fig.5). Although several additional genes may be employed, it is often to highlight other correlations between strains, for example tryptophan synthase has been used to track horizontal gene transfer of antibiotic resistance (Chater & Chandra, 2006). Other PCR-based DNA typing methods are amplicon/gene dependent and include restriction fragment length polymorphisms (RFLP), T-RFLP, DGGE and ARDRA.

The last two methods include low frequency restriction fragment analysis and pulse field gel electrophoresis, which makes use of high molecular weight DNA. While both techniques have been used for species comparison, neither method is frequently employed due to time and equipment constraints.

Table 4. The advantages and limitations of various characterization techniques.

Method	Distinguishing features	Advantages/Limitations	Discernable Level
G+C content	3% range for well-defined taxa; 10% range for Genus; 24-76%	Use on a range in bacteria	Genus-Species
DNA/DNA hybridization	Thermal stability decreases 1-2.2% for every 1% mispairing	Short oligonucleotides and experimentally induced mismatching cannot be extrapolated to an entire genome	Genus-Species
rRNA homology	Basepair similarity of rRNA molecule	Does not always agree with DNA/DNA hybridization results	Family-Genus-Species
Low frequency restriction fragment analysis (LFRFA/PFGE)	Based on enzyme cut patterns of whole genome	Need high molecular weight genomic DNA; Large DNA fragments; Complex patterns are hard to analyse	Species-Strain
Plasmid analysis	Based on Plasmid extraction	Not all strains contain or keep plasmids	Species
PCR-based DNA typing methods	Amplification of areas dependent on type of method and primers used	Applicable, Simple and rapid	Species-Strain
Cell wall composition	Cell wall -bound Peptidoglycan composition	Time consuming	Family-Genus-Species
Cellular fatty acids	Cellular fatty acid methyl ester content determined via G+C-MS	Cheap, rapid and can be automated	Family-Genus-Species
Isoprenoid quinones	Isoprenoid length, saturation and hydrogenation of side chains	Applicable to most strains	Family-Genus-Species
Whole cell protein analysis	Gel pattern via SDS-PAGE	Extremely reliable with good correlation with DNA/DNA hybridization	Species-Strain
Polyamines	All polyamines are qualitative and quantitatively different	Universal	Family-Genus-Species
Phage & bacteriocin typing	Pathogen specificity	Not applicable to all bacteria	Species-Strain
Zymograms	Enzymatic activity	Often media dependent	Species-Strain
Phenotype (BIOLOG, API etc.)	Growth on different energy sources	Parameter dependent	Family-Genus-Species-Strain

1.6.3. Informational macromolecules: 16S rRNA as a priori

In terms of the ever evolving system of taxonomic classification, the most important aspect is in fact the information molecule that forms the foundation of this study. It is important to define and motivate why a specific molecule is used due to differences between speciation patterns when considering functional/housekeeping genes instead of informational molecules such as 16S rRNAs.

Informational macromolecules, also known as semantides, refer to molecules that are present in all organisms. rRNA and cytochrome C are examples of informational macromolecules. Zuckerkandl and Pauling were the first to suggest that a universal molecule could serve as a

chronometer (Zuckerkandl & Pauling, 1965). They proposed that ferrodoxins and cytochromes could be used to track the evolutionary development of the major lineages. It was only later that Woese and co-workers suggested the use of the ribosomal genes due to their central role in the translation apparatus (Woese, 1987). The three universal rRNA molecules are the 23S rRNA (3300bp), the 16S rRNA (1450bp) and the 5S rRNA (120bp) which were named according to their Svedberg sedimentation rates. Woese claimed that the 16S rRNA molecule showed the least susceptibility to horizontal gene transfer and was therefore considered as the most stable. Even though the 23S molecule could give a better resolution, as the 16S is smaller it can be more readily sequenced. Although the 5S was preferred prior to the development of Sanger sequencing it has too few variable regions and invariant domains making it less accurate.

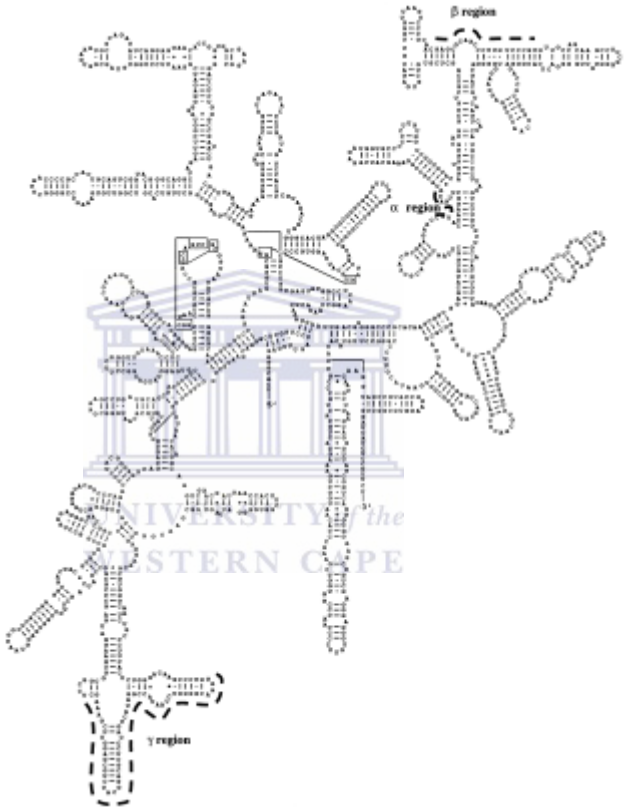


Figure 5. The 16S rRNA protein structure of *Streptomyces coelicolor* indicating the three regions (α, β, γ) commonly employed in phylogenetic studies. (Anderson *et al*., 2001)

The primary structure of 16S rRNA contains three chronometrical subclasses: the hypervariable regions, the moderately conserved regions and the invariate regions. The great variation in frequency of compositional changes between the three subclasses indicates that both deeper lying roots and recent branching lineages can be tracked with relative certainty. Problems arise when considering the peripheral branches where the resolving power is insufficient for correct delineation. This is infact where the species concept comes into play. A species as defined by DNA reassociation and 16S rRNA gene similarity shows certain discrepancies. Examples where

this discrepancy has been observed include strains which share a high level of 16S rRNA similarity but have low levels of genomic similarity (observed in *Amycolatopsis methanolica* and *Amycolatopsis thermoflava),* strains with identical 16S rRNA genes which have different phenotypic and genotypic profiles (observed in *Staphylococcus piscifermentaris and Staphylococcus condimenti*); and a single strain with a 6% difference in 16S rRNA sequences due to *rrn* operons (observed in *Haloarcula marismortui*) (Coenye & Vandamme, 2003; Conville & Witebsky, 2005; Conville & Witebsky, 2007; Ninet *et al*., 1996). The discrepancy seen between 16S rRNA sequence analysis and DNA relatedness suggests that when highly related strains are being compared additional methods are advisable.

The use of 16S rRNA molecules as phylogenetic markers offers great potential in the study of bacterial assemblages. Apart from the clear disadvantages stated above, the molecule is structurally and functionally stable, with a rapidly growing universal database of considerable size. Some studies have designed shorter nested primers that amplify variable regions shown to represent the largest variability within a given phylum or domain. Examples include archaeal-'specific' primers (Øvreås *et al*. 1997), actinobacterial-'specific' primers (Heuer *et al*., 1997; Stach *et al*., 2003) and Proteobacterial-'specific' primers (Øvreås *et al*., 1997). The need to sequence all type strains for characterization helps in the identification of uncultured sequences, which aren't necessarily possible for other functional genes. Despite the universal application of the 16S rRNA gene for phylogenetic analysis, alternative markers have successfully been employed in the study of bacterial communities. These include *rpoB*, *amoA* (ammonia monooxygenase) and *pmoA* (methane monooxygenase) (Roselló-Mora & Kampfer, 2004).

1.7. Culture vs Culture-independent

Detectable bacterial diversity may be subject to the bias introduced by the method of extraction (be it culture-dependent based or otherwise); by the environment from which it needs to be extracted and the taxonomic ambiguity associated with microorganisms.

1.7.1. Limitations and advantages of phenetic or culture based classification

With regards to culture-based methods, the bias is essentially introduced by the four techniques used to obtain isolates in pure culture: direct plating, enrichment, micromanipulation and extinction culture. These techniques provide a framework for pure, viable culture isolation, but do not account for the fraction of the community that cannot grow on solid media (Amann *et al*., 1995) or are fastidious. It has been shown that slight alterations to growth media may influence the culturability of numerically abundant bacteria (Bartscht & Heribert, 1999). It is estimated that only 0.25% of soil bacteria are accessible through standard culture methods (Amann *et al*., 1995). It is also well established that oligotrophic environments tend to be less amenable to culture-based studies, with a large proportion of the microorganisms entering a viable but nonculturable state. Many isolation techniques fail to dislodge bacteria from the soil particles, or select for subgroups via growth conditions and media (pH, temperature, pressure, oxygen saturation, micro/macronutrients and N/C-sources). In the soda lake environment, for example, *Proteobacteria* may hinder the isolation of other bacterial species due to their affinity for nutrient rich (>8g organic compound) media (Duckworth *et al*., 1996). In a similar way colony-colony inhibition, be it through antibiotics or shear film formation, limits the isolation of slow growing species such as *Mycobacterium*.

The accompanying limitations of culturing bacteria can be remedied depending on the research question. For example, if numerically abundant bacteria are known to grow slower, they can be isolated if nutrient limiting media (5-80mg organic compound per litre) is used and fast growing colonies are excised early on (Fry, 2004). Longer incubation periods of up to 84 days provide a larger pool to work from, micromanipulation has proven to be a less conventional investigative tool. This method utilises original environmental extracts to culture and isolate micro-colonies by means of a soil substrate membrane system (Ferrari & Gillings, 2009). The 'extracts' however contain mixed micro-colonies, which still relies on traditional means of separation.

Community level physiological profiles provide a different approach to culture based studies of bacterial diversity. It makes use of the traditional categorisation of bacteria according to carbon source utilisation by means of spectrophotometric detection of the reduction of tetrazolium dye (Preston-Mafham *et al*., 2002). The BIOLOG systems (96-well microtiter plates with 95 C-sources) and newer eco-plates (31 environmentally applicable C-sources) provide a high throughput means of screening environmental samples for gross functional diversity (Kirk *et al*., 2004; Classen *et al*., 2003).

1.7.2. Culture-independent approaches to phylogenetic diversity

Culture-independent approaches can be subdivided into molecular and biochemically-based techniques. Each approach has its advantages and disadvantages depending on the depth of resolution that needs to be attained for reliable deductions. A relatively wide range of techniques have been adapted to study bacterial diversity, not all equally suited for each purpose. Techniques such as fatty acid methyl ester (FAME), chemotaxonomic marker (polyamines and quinines), diaminophosphates (DAP) and cell wall sugar analysis can only reach up to species level. Conversely, rRNA/DNA sequencing and probes can be employed to delineate at the order, family, genus and species level, but these methods are costly. The accepted methodology for analysing community diversity combines low resolution screening methods with sequencing to circumvent the problems associated with some of these approaches. Table 5. outlines the advantages and disadvantages, as well as the suitability of each technique in terms of its application. Several methods have been reviewed in detail elsewhere (Zhang & Xu, 2008; Hill *et al*., 2000; Nocker *et al*., 2007) and will not form part of this review.

Table 5. Advantages and disadvantages associated with various culture-independent techniques.

Method	Advantages	Disadvantages	Suitability	Reference
G+C content	Not influenced by PCR biases Includes all DNA extracted Quantitative Includes rare members of community	Requires large quantities of DNA. Dependent on lysing and extraction efficiency. Coarse level of resolution		(Tiedje *et al*. , 1999)
Nucleic acid reassociation and hybridization	Total DNA extracted Not influenced by PCR Biases. Study DNA or RNA Can be studied in situ	Lack of sensitivity Sequences need to be in high copy number to be detected; Dependent on lysing and extraction efficiency		(Torsvik *et al*., 1990a; Torsvik, *et al*., 1990b)
DNA microarrays and hybridization	Same as nucleic acid hybridization Multiple genes can be analyzed If using genes or DNA fragments, increased specificity	Only detect most abundant species Need to be able to culture organisms Only accurate in low diversity systems	High throughput screening of metabolically active communities	Hubert *et al*., (1999), Cho and Tiedje, (2001), Greene & Voordouw, (2003)
Denaturing and Temperature Gradient gel electrophoresis (DGGE and TGGE)	Large number of samples can be analyzed simultaneously Reliable, reproducible and rapid	PCR biases Dependent on lysing and extraction efficiency Sample handling can influence community, i.e. if stored too long before extraction, community can change. One band can represent more than one species (co-migration) Only detects dominant species	For samples with a limited number of abundant members	Muyzer *et al*., (1993), Duineveld *et al*., (2001), Maarit-Niemi *et al*., (2001)
Single strand conformation Polymorphism (SSCP)	Same as DGGE/TGGE No G+C clamp No gradient	PCR biases Some ssDNA can form more than one stable. conformation	High sensitivity without restriction analysis or G+C	

Continue Table 4.

Method	*Advantages*	Disadvantages	Suitability
Amplified Ribosomal DNA Restriction analysis (ARDRA) restriction fragment length Polymorphism (RFLP)	Detect structural changes in microbial community	PCR biases Banding patterns often too complex	Low resolution screening Simple community composition
Terminal restriction fragment length polymorphism (T-RFLP)	Simpler banding patterns than RFLP Can be automated; large number of samples Highly reproducible Compare differences in microbial communities	Dependent on extraction and lysing efficiency PCR biases Type of *Taq* can increase variability Choice of universal primer and enzymes will influence community fingerprint	Samples with higher species richness Time-scaled experiments Large sample numbers
Ribosomal intergenic spacer Analysis (RISA) Automated ribosomal Intergenic spacer analysis (ARISA)	Highly reproducible community profiles	Requires large quantities of DNA	Subtle differences between species
Denaturing high performance Liquid chromatography (D-HPLC)	High throughput High sensitivity No sample manipulation	Separation parameters needs to be optimised for different samples	Automated and fast analysis
Cloning and Sequencing	High phylogenetic resolution Allows species identification/determination of closest phylogenetic neighbor	Although the sequencing can be automated, the cloning is time-consuming	When phylogenetic assignment is relevant high-throughput sequencing service is available
Single strand conformation Polymorphism (SSCP)	Same as DGGE/TGGE No G+C clamp No gradient	PCR biases Some ssDNA can form more than one stable. conformation	High sensitivity without restriction analysis or G+C clamps
Amplified Ribosomal DNA Restriction analysis (ARDRA) Or restriction fragment length Polymorphism (RFLP)	Detect structural changes in microbial community	PCR biases Banding patterns often too complex	Low resolution screening Simple community composition
Terminal restriction fragment length polymorphism (T-RFLP)	Simpler banding patterns than RFLP Can be automated; large number of samples Highly reproducible Compare differences in microbial communities	Dependent on extraction and lysing efficiency PCR biases Type of *Taq* can increase variability Choice of universal primer and enzymes will influence community fingerprint	Samples with higher species richness Time-scaled experiments Large sample numbers
Ribosomal intergenic spacer Analysis (RISA)/ ARISA-Automated ribosomal Intergenic spacer analysis	Highly reproducible community profiles	Requires large quantities of DNA	Subtle differences between species
Denaturing high performance Liquid chromatography (D-HPLC)	High throughput High sensitivity No sample manipulation	Separation parameters needs to be optimized for different samples	Automated and fast analysis
Cloning and Sequencing	High phylogenetic resolution Allows species identification/determination of closest phylogenetic neighbor	Although the sequencing can be automated, the cloning is time-consuming	When phylogenetic assignment is relevant and high-throughput sequencing service is available

Source: *Modified from Kirk (2004)

Denaturing gradient gel electrophoresis

Denaturing gradient gel electrophoresis (DGGE) was originally developed to detect point mutations in short gDNA sequences. Muyzer *et al.,* (1993) was the first to see its potential for studying bacterial diversity. It has been successfully used to study bacterial communities in a variety of settings such as rhizosphere soil (Niemi *et al*., 2001), acidic river sediment (Aguilera *et al*., 2006), hot spring water and sediment (Lulustyaningati *et al*., 2008), heavy metal contaminated soil (Gremion *et al*., 2004), meromictic lake sediment (Øvreås *et al*., 1997), solar salterns (Casamayor *et al*., 2002) and Antarctic sediment (Smith *et al*., 2006). The technique is however only as accurate as the resolution of the images. An increased resolution can be attained by limiting the range of the gradient or using a more sensitive staining method i.e. silver staining instead of Ethidium bromide. However, one should refrain from using band intensities to determine relative abundance. As the amplicons for divergent species may co-migrate (two organisms but one band) qualitative data based on band intensity may be misleading (Aguilera *et al*., 2006). Apart from these factors DGGE has been successfully used to track changes within the same habitat (Joynt *et al*., 2006) and can provide reliable and reproducible results from a small amount of starting material (less than 1g of soil) (Kirk *et al*., 2004). The ~300 bp fragment of the amplified V3 region (within the 16S rRNA gene) has been shown to produce an accurate phylogenetic tree corresponding to the full gene tree within a selected phylum and has therefore been adopted for most phylogenetic studies (Øvreås, 2000). It is however a PCR-based technique and falls subject to all the biases introduced by nucleic acid extraction and PCR amplification. Gelsomino *et al.,* (1999) examined the discrepancies introduced by each aspect by profiling of a range of soil structures from geographically dispersed locations (Gelsomino *et al*., 1999). They concluded that all DGGE-profiles were reproducible and that the largest discrepancies existed between indirect and direct nucleic acid extraction from the matrix, in which case the faint banding patterns differed slightly. The study was however conducted on sandy, loam and clay soils without noting the concentrations of the most common inhibitory compounds in such soil environments.

Cloning and amplified ribosomal DNA restriction analysis (ARDRA)

Amplified ribosomal DNA restriction analysis is a method by which a single or multiple restriction enzymes are used to fragment an amplified 16S rRNA gene into bands of various sizes. The principle is based upon the selective restriction pattern of the 4 or 6 bp restriction enzyme which is used. It has been successfully employed in the separation of various *Mycobacterium* (Vaneechoutte *et al*., 2000), *Mycoplasma* (Stakenborg *et al*., 2005)*, Ralstonia* (Segonds *et al*., 2003) as well as *Acinetobacter* species (Koeleman *et al*., 1998), and is readily used as an alternative to the more costly pyrosequencing approach. However, computational studies have shown that in order for ARDRA patterns to form reliable clusters, four tetrameric

restriction enzymes would be required and would still only resolve differences of more than 96% 16S association (Gevers *et al*., 2005). A similar study used sequences from the ribosomal database project's cultured strains and *in silico* modelling of ARDRA patterns to determine the reliability of the technique (Maidak *et al*., 2001). With regards to the ten enzymes used (which included *Alu*I, *Rsa*I and *Taq*I), it was concluded that for a resolution down to genus level, two or more enzymes would be needed. However, if the aim was to distinguish between species, three was the minimum number of enzymes required. ARDRA has been employed to screen 16S rRNA gene libraries from various environments (Sjoling & Cowan, 2003; Khetmalas *et al*., 2002; Dees & Ghiorse, 2001; Maturrano *et al*., 2006).

Bias introduced by nucleic acid extraction

Nucleic acid extraction has been optimised for a large variety of environments; however the extraction of gDNA directly from soil poses several additional hurdles. These restraints are compounded in extreme environments, such as soda lakes, where the coextraction of humic acids, high concentrations of salt and sulphur complicates downstream processing. The few microbial diversity studies conducted on similar environments utilised a range of crude DNA extraction methods (Moore *et al.,* 1999; van Elsas *et al.,* 1996; Chen & Kuo *et al.,* 1993) followed by column purification. Krsek and Wellington (1999) examined the effects of buffer and lysis composition on PCR amplification with various universal primer sets. They found that depending on buffer composition there can be up to a twofold difference in genomic DNA yield. Only two out of 14 buffer/extraction method combinations were able to produce amplified products with all primer sets (Krsek & Wellington, 1999). Several commercial kits are also available, of which the MoBio Powersoil DNA isolation kit is most often used. Direct DNA extraction uses enzymatic and chemical or mechanical cell disruption, without removing the cells from the matrix. This method may lyse more of the cells, but has the added disadvantage of coextracting inhibitors. Indirect DNA extraction uses blending, cation-exchange or super paramagneric silica-magnetite nanoparticles to remove cells from soil before extracting the nucleic acids (Robe *et al.,* 2003). Although these methods circumvent the problem of inhibitors, the extraction may be slightly biased towards a subpopulation of the community. This however does not imply that the direct approach lysis all cell types at the same efficiency. Physical techniques such as sonication, bead beating and freeze thaw cycles tend to discriminate to a lesser degree. Due to the difference in cell wall structure and composition chemical and enzymatic digests favour either Gram-positive or Gram-negative morphotypes depending on the harshness and length of digestion. Janssen *et al.,* (2002) noted that the bias introduced by nucleic acid extraction methods, might account for the under representation of Actinobacteria in clone libraries even when they are prominent in culture-based studies.

The discrepancies between culture-based and culture-independent approaches are evident in the minimal overlap between species observed. A recent review of diversity studies which employed both approaches highlighted these differences in sequence overlap (Donachie *et al.,* 2007). One example was used in which a Hawaiian lake diversity study detected 147 unique cultured taxa and 159 unique OTU's (Fig.6). In that one study, no overlapping taxa were found even though the numbers were comparable, suggesting a large proportion of the bacteria would have been overlooked if one or the other approach was removed. A similar phenomenon was observed in the other studies. It is apparent that these approaches are complementary and that by using both, it increases the likelihood of correcting certain biases.

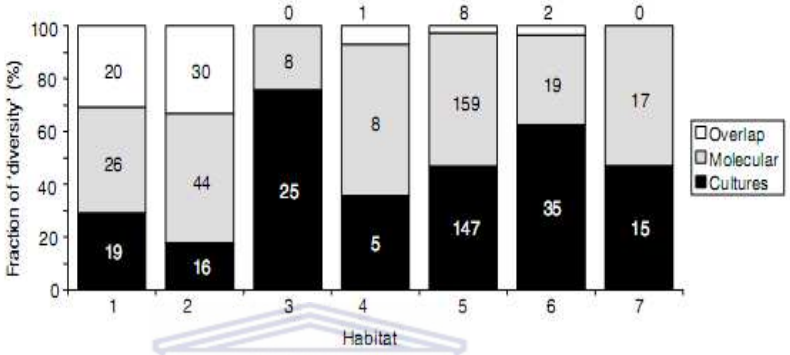


Figure 6. A "Comparative analysis of diversity detected by…culture-dependent and culture-independent approaches" between several studies (Donachie *et al*. 2007). The numbers indicate the unique sequences that represent each technique and those that were detected by both. The seven habitats shown are (1) endodontic pathogenic oral biofilms (2) exodontic oral biofilms (3) Hawaiian lakes and Loihi (4) rhizoplane samples (5) bird feathers (6) ocean (7) hypersaline salterns.

1.8. Biotechnological and Industrial application

Following the discovery of streptomycin in 1943, considerable effort was invested into screening unexplored actinobacteria for secondary metabolites. Since these early discoveries over 12 000 antibiotics have entered our markets, of which 55% are produced by *Streptomyces* species. Actinobacterial secondary metabolites, excluding antibiotics, amount to 42% of the 23000 products on the market today. These include compounds associated with antibacterial, antifungal, antiviral, antiparasitic, immunosuppressive, anti-inflammatory, antitumor, insecticidal, antioxidant, diabetogenic, transplant rejection, inhibition of HIV, enzyme inhibition and high cholesterol treatments. Due to the wide range of possible applications we will discuss three major categories: pharmaceutical and cosmetic, industrial, and biotechnological applications.

The pharmaceutical industry is often the hardest market to penetrate and it remains the market

with the greatest financial rewards. *Streptomyces* species are notorious producers of antimicrobial substances. This may be due to their need to proliferate in competitive environments (mostly terrestrial). The main antimicrobial classes characterized thus far include aminoglycosides, anthracyclines, glyco- and polypeptides, beta-lactams, macrolides, nucleosides, polyenes, polyesters, polyketides, actionomycins and tetracyclines. Additional classes of compounds used in the human health industry include anthraquinones, angycyclinone, phenazine, piercidins, octaketides and lactones (Lam, 2006; Goodfellow & Fiedler, 2010). The source and compatible action is denoted in Table 6.

Table 5. Secondary metabolites produced by actinobacteria which are currently on the market.

Metabolite	Class	Function	Target species	Source	Reference
Fogacin	Octaketide	antibacterial	*Bacillus subtilis*	*Streptomyces* strain TU-6319	(Radzom *et al.* 2006)
Chandramycin	Polypeptide	antibacterial	*Clostridiums*	*Streptomyces lydicus*	(Singh *et al.* 2009)
Rosaramicin	Macrolide	antibacterial		*Micromonospora*	(Horan & Brodsky 1986)
Oktilactomycin	Lactone	antibacterial	*S. aureus* and *S. pyogenes*	*Streptomyces griseoflavus*	(Imai *et al.* 1987)
Chimomycin	Augucyclinone	antibacterial	*Bacillus subtilis*	*Nocardia mediterranei* subsp. *kanglensis* 174764	(Kekuda *et al.* 2010)
Capreomycin sulphate	Polypeptide	Anti-mycobacterial	*Mycobacterium Tuberculosis*	*Streptomyces capreolus*	(Skinner & Cundliffe 1980)
Cycloserine		Anti-mycobacterial	*Mycobacterium Tuberculosis*	Various *Streptomyces* sp.	(Kekuda *et al.* 2010)
Brasilinolide A	Macrolide	antifungal	*Aspergillus niger*	*Nocardia brasiliensis* IFM0406	(Kekuda *et al.* 2010)
AB023	Polyene mix	antifungal	*Botrytis cinerea*	*Streptomyces* SD581	(Kekuda *et al.* 2010)
Coronamycin	Complex	antiprotozoal	*Plasmodium Falciparum* (malarial parasite)	verticillate *Streptomyces* sp.	(Kekuda *et al.* 2010)
Doramectin	Avermectin	Anthemintic	Gastrointestinal nematodes	*Streptomyces avermitis*	(Kekuda *et al.* 2010)
Pentalactones	N/A	Antiviral	DNA viruses	*Streptomyces* strain 606	(Kekuda *et al.* 2010)
Ehrlichin	N/A	Antiviral	Influenza A and B	*Streptomyces lavendulae*	(Kekuda *et al.* 2010)
Actinohivin	N/A	Antiviral	Human immunodeficiency virus(HIV-1)	Actinobacteria strain K97-0003	(Chiba *et al.* 2001)
Tomaymycin	N/A	Anti-phage	T-series E.coli phages	*Streptomyces achromogenes*	(Kekuda *et al.* 2010)
Retymicin, Galtamycin B	N/A	Cytotoxic	cytostatic effects to various human tumor cell lines	*Micromonospora* strain TU6368	(Kekuda *et al.* 2010)
Algacidins A and B	N/A	Cytotoxic	Yoshida sarcoma cells in culture	*Streptomyces* sp. RK-1339	(Kekuda *et al.* 2010)
Carquinostatin A	N/A	Antioxidant Anti-inflammatory	Brain protection	*Streptomyces exfoliatus* *Streptomyces aureofaciens*	(Kekuda *et al.* 2010)

N/A : Not available

Actinobacteria produce a myriad of pigments which are of particular interest to the cosmetics industry. Since synthetic pigments pose a toxicity problem, most manufacturers are searching for natural equivalents in plants and microorganisms. Microbial pigments have the added benefit of light and heat stability, and are often more water soluble than their vegetative counterparts. Compounds associated specifically with cosmetics include carotenoids, melanins, flavins, quinones, prodigiosins, monascins, violacein or indigo (Venil & Lakshmanaperumalsamy, 2009). The *Microbacteriaceae* family exhibit yellow, orange and red pigmentation due to the overproduction of isoprenoids (carotenoids) and several other heterocyclic compounds (Venil & Lakshmanaperumalsamy, 2009) which are utilised by both the cosmetic and food industries (Margesin & Schinner, 2001). Some coloured pigments may even perform a secondary role as a biosurfactant or antioxidant such as in the case of group B Type *Streptococcus* species (Deepika & Kannabiran, 2010).

Several industrial processes including the manufacturing of cement, lye treatment of hides, food processing, wastewater treatment (especially farm effluents), mining operations as well as paper and pulp manufacturing, have a common processing stage i.e. alkaline enzyme treatment (Ulukanli & Digrak, 2002). Alkaliphiles are notorious extracellular enzyme producers of which serine proteases commands the widest market shares (Grant, 1990). Serine proteinases are endopeptidases with reactive serine molecules in their active sites which bind and hydrolyse protein residues. Common products include Alcalase®, Esperase®, Maxatase® and Maxacal® which are used in the detergent industry for the removal of blood and grass stains. Since these enzymes are adapted to function at 50-60°C in an alkaline environment they prove to be more compatible with the non-ionic surfactants and sequestering agents present in detergents (Grant, 1990). In addition to dehairing hides, proteases are also used to remove the gelatinous layer from x-ray film (Fujiwara *et al.,* 1993). Some amylases (i.e. from *Nesterenkonia halobia*), lipases and cellulases derived from similar alkaliphiles have penetrated the same markets, although to a lesser extent (Horikoshi, 1999). Existing alkaline proteases require the presence of Ca^{2+} ions to function effectively and these in turn are scavenged by the sequestering agents in detergents. It therefore follows that isolates from haloalkaline environments (such as the Ethiopian Rift Valley lakes) which contain less than 0.001mg/L of Ca^{2+} might produce an Ca^{2+}-insensitive equivalent to these serine proteases. The rest of the industrial enzyme repertoire includes xylanases (for bleaching in paper manufacturing and Rayon modification), pectinases (for wastewater treatment), mannases (food processing) and cyclodextrins (as emulsifiers or foaming agents and solubilisation of vitamins and hormones) (Ulukanli & Digrak, 2002). To date no halophilic bacterial isolate has shown xylanase activity, but all seem to be present in one or more alkaliphiles (Martı *et al*., 2003).

The industrial utilisation of osmolytes has proved of some importance in the maintenance and storage of a range of proteins; especially in the biotechnological industry. Such osmolytes are commonly produced by extremophiles in haloalkaline environments to balance cell osmotic pressure. These osmotic shift principle which is initiated by an increase in NaCl causing water to withdraw from proteins; thereby helping to maintain the weak interactions that are necessary to stabilise structural integrity. In the same process an increase in potassium leads to the overproduction of the osmolytes which are then excreted into the medium following a sudden dilution event. These osmolytes can then be removed from the medium for use in commercial storage buffers. Some osmolytes can be catabolised or polymerized to an osmotic inert state which complicates its extraction. The production of osmolytes, although widespread, can only be effective up to 1.5 M salt from then on it becomes energetically unfavourable (Bowers *et al*., 2009). Similar biopolymers have been utilised as biosurfactants in oil recovery (Deepika & Kannabiran, 2010).

CHAPTER 2: MATERIALS AND METHODS

2.1. Environmental sample collection

Composite, subsurface (10 cm below the surface) soil samples were collected during November 2009 (Fig.7). All of the samples were collected aseptically in 1L polypropylene bottles or 50 ml falcon tubes and kept as cool as possible during the sampling period. Samples were stored at 4°C following the sampling expedition as well as during shipment to South Africa where it was stored at -80°C. *In situ* GPS coordinates, elevation, pH, temperature (soil and water), conductivity (ms/cm) and total salts (%) were measured (Table 7.). All the samples were taken in the rainy season littoral zone, limiting human contamination by avoiding the benthic mats that form at the waters edge.

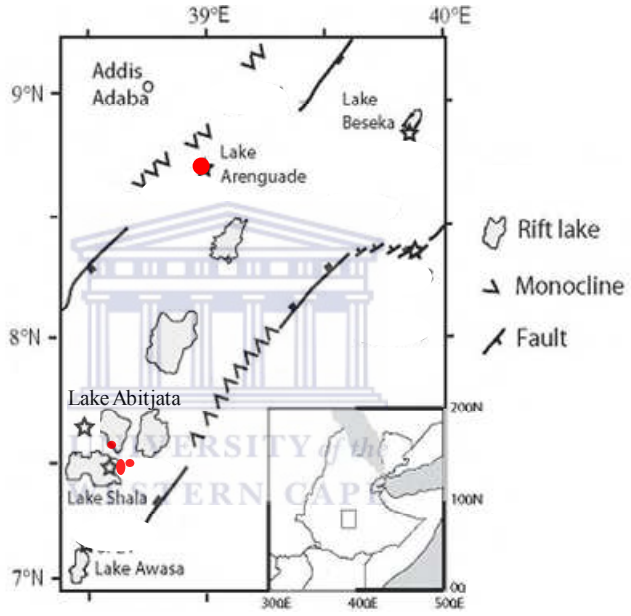


Figure 7. Sampling sites (indicated by red points) on which diversity studies were conducted.

Table 6. Sampling site parameters and readings.

Name	GPS N	GPS E	GPS m.o.s.	Temperature °C		pH (digital)	conductivity (ms/cm)	Salt w/v
				Water	Sediment			
Arenguade /Hora	8° 41, 841	38° 58,795	1858	24	N/A	9.77	10.2	2.50%
Shala hot spring_1	7° 28, 696	38° 38, 096	1572	63	92	8.2	10.2	0.60%
Shala hot spring_2	7° 28, 699	38° 38, 097	1572	71	95	8.5	10.21	0.50%
Shala hot spring_3	7° 28, 697	38° 38,100	1570	70	N/A	8.26	11.52	0.60%
Lake Abitjata	7° 42,305	38° 36,670	1570	29	N/A	9.81	41.66	3.50%

N/A Not able to reach sediment

2.2. DNA Extractions

2.2.1. Extraction of DNA from sediment samples

Three protocols were used in order to compare the implications of bias introduced through isolation. The first is a procedure described by Zhou *et al.,* (1999) with additional modifications added by Stach *et al.,* (2001), the second is a procedure by Wang *et al.,* (1996) slightly modified by including an overnight lysis instead of 2 hours, and the third is a modified hard lysis protocol. The hard lysis protocol was also used in bacterial culture DNA extraction and is described in section 2.2.2.

2.2.2. Preparation of genomic DNA from pure bacterial cultures

A new protocol (hard lysis) was designed for the actinobacterial isolates due to their resistance to lysozyme treatment. Overnight liquid cultures (2 ml), were harvested by centrifugation (Eppendorf™ microcentrifuge) at 10 000 x *g*. The cell pellets were washed twice with 0.1 M Tris-EDTA (TE) buffer at pH 8.0 and centrifuged at 10 000 x *g*. The pellet was then resuspended in 100 µl breaking buffer solution (2% Triton X100; 1% SDS; 100 mM NaCl; 10 mM Tris-HCl pH 8.0; 1 mM EDTA pH 8.0) vortexed and added to a 2 ml eppendorf tube containing 100 ng of 50-70 mesh particle size seasand (SIGMA™). One hundred microlitres (100 µl) of PCI (Phenol: Chloroform: Isoamyl alcohol 25:24:1) was added and the tubes were vortexed at full speed for 3 min. Afterwards, 200 µl of TE buffer was added and vortexed briefly before centrifugation at 10 000 x *g* for 5 min. The supernatant was removed to prechilled 1.5 ml tubes followed by the addition of 1 ml ice cold 100% EtOH. This was left on ice for 10 min followed by centrifugation at 10 000 x *g* for 5 min. The pellet was then resuspended in a solution containing 1ml 70% ice cold EtOH and 10 µl 4M NH_4OAc, left on ice for 10 min and centrifugation at 10 000 x *g* for 5 min. The resultant DNA pellet was airdried and resuspended in 50 µl TE buffer pH 8.0.

2.2.4. Determination of DNA concentration

The concentration of DNA in suspension was determined spectrophotometricly using a NanoDrop™ ND-1000 UV-Vis spectrophotometer (Thermo SCIENTIFIC, USA). The purity of the sample was evaluated by the A_{260}/A_{280} and A_{230}/A_{260} ratios. A 260/230 ratio of 0.3-0.9 and a 260/280 ratio of 1.6-1.8 was considered optimum for DNA purity indicating low levels of inhibitors. A 5 µl aliquot of genomic DNA was loaded onto a 1.0% LE grade agarose gel containing 0.5 µg/ml ethidium bromide alongside a 0.3 µg/µl FERMENTAS™ 100bp Plus ladder unless otherwise stated. The chemiluminescence of the genomic DNA was then matched to the ladder and its corresponding concentration was determined as per manufacturer's specifications.

2.3. Polymerase chain reaction (PCR)

2.3.1. Oligonucleotide primers and protocols

A series of primers were used during the course of this study (Fig.8 and Table 8). Table 8 gives a summary of the primer specifications and function. Genomic DNA was diluted for optimal amplification which ranged from 1- 50 ng depending on the concentration of inhibitory compounds. A standard PCR reaction contained 0.2 mM each of dATP, dCTP, dGTP and dTTP, 2 mM of each primer, 2 mM $MgCl_2$ and 0.25u (total volume) polymerase enzyme. In all soil genomic DNA amplifications, 0.4 ng/µl Bovine serum albumin (Fermentas©) solution was added. In all general PCR reactions Fermentas DreamTaq™ polymerase was used. The only two exceptions were with DGGE amplifications where the primer set 341F-GC and 534r was used, and M13 amplifications using Mf & Mr primers. KAPA 2G Robust™ *Taq* (KAPA Biosystems™) was used for PCRs with the DGGE primers and the pAC-S vector primers (Mf & Mr) were used with DyNazyme™ Polymerase (FINNZYMES©). All PCR reactions were carried out on an Eppendorf Multigene™ thermal cycler at 50 µl reaction volumes.

2.3.2. Purification of PCR products

PCR purifications were carried out using ExoSAP-IT™ (Affymetrix©) as per company specifications. Gel purifications were carried out using the Qiaquick™ Gel extraction purification system (Qiagen©).

Table 7. Oligonucleotide primers used during the course of this study

Primer	Forward/ reverse	Primer target	Primer function	G+C clamp	Sequence	Cycling parameters	Predicted size	Reference
F1	Forward	General Bacterial 16S rRNA	General sequencing	None	5- AGA GTT TGATCI TGG CTC AG -3	95°C/4 min 30 x(95°C/30s-55°C/30s-72°C/90s) 72°C/10 min	1500bp	Modified (Weisburg *et al.*, 1991)
R5	Reverse			None	5- ACG GIT ACC TTG TTA CGA CTT -3			
F3	Forward	General Bacterial 16S rRNA	General sequencing	None	5-GCC AGC AGC CGC GGT AATAC-3	96°C/1 min 25x (96°C/10s-50°C/5s-60°C/4 min) 4°C/hold	800bp	
341F-GC	Forward	DGGE 16S rRNA	DGGE	Yes	5- CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GCC TAC GGG AGG CAG CAG-3	94°C/4 min 20 x(94°C/45s-65°C/45s-72°C/60s) 20x (94°C/30s-55°C/30s-72°C/60s) 72°C/10 min	200bp	(Muyzer *et al.*, 1993)
534r	Reverse	DGGE 16S rRNA	DGGE	None	5- ATT ACC GCG GCT GCT GG -3			
Act235r	Forward	Actinobacterial-specific 16S	Clone Library/DGGE	None	5- CGC GGC CTATCA GCT TGT TG -3	94°C/4 min 10 x(94°C/45s-65°C/45s-72°C/60s) 25x (94°C/30s-57°C/30s-72°C/90s) 72°C/5 min	297bp	(Stach *et al.*, 2003)
341F-GC								
A3Fa	Forward	Archaeal-specific primers	16S	None	5- TCCGGTTGATCCYGCC GG-3	94°C/4 min 40 x(94°C/60s-55°C/60s-73°C/3 min) 72°C/10 min	200bp	(Jurgens *et al.*, 1997)
Ab927r	Reverse			None	5- CCCGCCAATTCCTTTA AGTTTC-3			
M13f	Forward	Vector pGEMT Easy	Clone Library	None	5- CCC AGT CAC GAC GTT GTA AAA CG -3	94°C/4 min 10 x(94°C/30s-65°C/30s-72°C/90s) 25x (94°C/30s-55°C/30s-72°C/90s) 72°C/5 min	1650bp	pGEM-T Easy primers
M13r	Reverse			None	5- AGC GGATAA CAATTT CAC ACA GG -3			

Continue Table 8.

Primer	Forward/ reverse	Primer target	Primer function	G+C clamp	Sequence	Cycling parameters	Predicted size	Reference
Mf	Forward	Vector (pSC-A)	Clone Library	None	Not provided	94°C/5 min 25 x(94°C/30s-55°C/30s-72°C/60s)	450bp	pAC-S vector primers
Mr	Reverse			None		72°C/7 min		

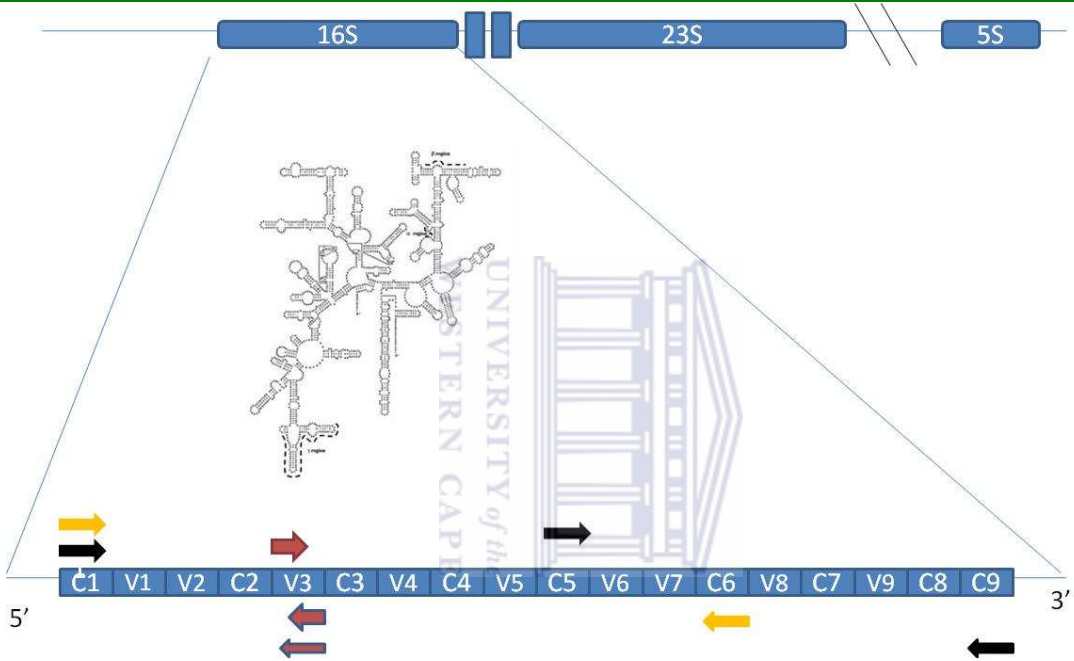


Figure 8. Primer annealing region within the 16S rRNA gene. Black denotes universal primers (F1, F3, R5); Red denotes actinobacterial and general DGGE primers; Yellow denotes Archaeal primers.

2.4. Cloning

2.4.1. Preparation of *Escherichia coli* competent cells

Chemically competent Genehog *E.coli* (Invitrogen©) cells were prepared by inoculating a single colony into a 5ml ψ-broth (2.0% Tryptone; 0.5% Yeast extract; 0.4% $MgSO_4.7H_2O$; 10 mM KCl) and culturing overnight at 37°C with agitation (120 rpm). This was then subsequently used in the process as stipulated by (Sambrook & Russell, 2001).

2.4.2. Ligation of PCR products

In the construction of the eubacterial 16S rRNA gene libraries, the pGEM-T Easy (Fermentas©) vector system was used as per manufacturer's instructions and ligation was performed overnight at 4°C. The actinobacterial-specific 16S rRNA gene libraries were made using the STRATAclone™ PCR cloning system (Agilent technologies©) and the ligations were performed at room temperature for 5 min as per manufacturer's specifications. In all reaction mixtures a 1:3 ratio (vector to insert) was used.

2.4.3. Transformation of competent cells

Aliquots (50 µl) of competent cells were thawed on ice and 2 µl of the ligation mixture was added to the cells and incubated on ice for 20 min. The tubes were then heat shocked at 42°C for 45s and placed on ice for another 2 min before adding 1ml of SOB media. The *E. coli* was allowed to recover for for 2 hours at 37°C. Cells were diluted with LB broth and plated onto Luria agar containing 100 µg/ml ampicillin, 50 µg/ml X-Gal.

2.4.4. Blue/white selection and size screening for recombinant plasmids

Between 300 and 400 white colonies were picked per library. Template DNA was prepared by inoculating single tranformants into 10 µl of double distilled water and using a subsequent 1 µl of the 10 µl dilution in a M13 PCR reaction (see section 2.3.1). All clones which where found to contain a corresponding ~1700 bp fragment on a 1% LE grade agarose gel were considered positive and 100 positives clones from each library were randomly selected for further screening.

2.4.5. Storage of clones

All clones were grown overnight at 37°C (120 rpm agitation) in a 96-well microtiter plate containing 100 µl Luria broth supplemented with 100 µg/ml ampicillin. Overnight cultures were supplemented with 30 µl of 80% glycerol and stored at -80°C.

2.5. 16S rRNA gene sequencing

2.5.1. Thermocycler protocol

The pSC-A selected M13 full amplicons (~550bp) of the actinobacterial libraries that represented unique ribotypes were sequenced using a BigDye™ sequencing (Applied Biosystems©) kit as per manufacturer

specifications. All unique ribotypes were sequenced using the STRATAclone M13 forward primer (Mf) provided by the manufacturer and 100ng of purified template was used.

2.5.2. Purification of extension products

All PCR products were cleaned up using ExoSAP-IT™ (Finnzymes®) as per manufacturer specifications and diluted to 100ng/µl for further use.

2.5.3. General sequencing and primers

Sequencing was done using an Applied Biosystems 3100 sequencer and a 27F primer (also known as F1) for non-clone 16S amplicons. All clones were sequenced using the appropriate M13 forward primer complimentary to the vector in question and the F3& R5 primer (see Table 8). The Applied Biosystems Big Dye terminator v3.1 Cycle Sequencing kit was used as per manufacturer's protocols with all reactions performed according to the manufacturer's instructions and cycle sequenced on a GeneAmp PCR System 9700 (Perkin Elmer, Applied Biosystems©). All actinobacterial library clones were sequenced as a service by the UIB DNA sequencing facility. All other sequencing runs were performed as a service by the UCT sequencing facility.

2.6. Amplified ribosomal DNA restriction analysis (ARDRA)

2.6.1. Preparation and protocol

All 16S rRNA gene clones were amplified using the respective vector primers (refer to section 2.3). The amplified product (10 µl) was used in an overnight double digest at 37°C with 0.25u of *Rsa*I and *Alu*I (Fermentas©) in a 21 µl reaction. The enzymes were deactivated by incubation for 15 min at 80°C. The digested products were loaded onto a 3.5% agarose gel containing 0.75 µg/ml ethidium bromide or Gelred™ and visualised with an Alphaimager 3400 Imaging System UV transilluminator (AlphaInnotech Corporation™ San Leandro, CA).

2.6.2. Ribotype selection

Unique ribotypes were selected using the freeware Gel2K (Norland, 2004), GELCOMPARE II© v 6.0 (Applied Maths NV®) and via visual confirmation when digested bands were too light for detection (i.e. lower resolution).

2.7. Denaturing gradient gel electrophoresis (DGGE)

2.7.1. PCR protocol

All DGGE primer sets are indicated in Table 8. Eubacterial DGGE products were amplified in a two step process. Firstly, universal eubacterial primers (F1 and R5) were used to amplify the entire 16S rRNA gene followed by a nested PCR reaction to amplify the 200bp V3 region and attach a GC-clamp (341F-GC and 534r).The actinobacterial DGGE products were amplified in a similar two step process by which the initial

universal bacterial amplicon (spanning F1-R5) was used as a template for the nested actinobacterial DGGE primer amplification (341F-GC and Act235r). This fragment also spanned the variable 3 region within the 16S rRNA gene (region 205-502bp *Streptomyces lividans* 16S rRNA gene numbering). The Archaeal DGGE products were firstly biased with the Archaeal primer (A3Fa and Ab927r) set amplifying directly from the genomic DNA followed by the universal DGGE primer set (341F-GC and 534r) to attach a GC- clamp.

2.7.2. Preparation and visualisation

The DGGE system was set up as per manufacturer's instructions. A 1xTAE (40mM Tris acetate and 1mM EDTA) buffer was used as a running buffer in a DCode™ Universal mutation system (BIORAD®) preheated to 62°C. A 9% (w/v) polyacrylamide, urea-formamide denaturing gel (16.5mm x 16.5mm x 1mm thick) was constructed by using a BIORAD model 475 gradient former (Bio-Rad, Hercules, USA). The individual 40/80% denaturing solutions were constructed as per manufacturer suggestions and 0.5% (w/v) APS as well as 0.02% (v/v) TEMED were used as catalysts for polymerization. All gels were loaded with a mixture containing 40 µl of PCR product (containing roughly 500ng DNA) and a 30% glycerol loading dye. The gels were then run at 100V for 16 h at 60°C. Staining of individual gels were conducted in TAE (40mM Tris acetate and 1mM EDTA) buffer containing 0.5 µg/ml EtBr for approximately 20 min. The initial staining was followed by destaining in TAE buffer for 10 min prior to visualisation with an Alphaimager 3400 Imaging System UV transilluminator (AlphaInnotech CorporationTM San Leandro, CA).

2.8. Bioinformatics

2.8.1 Sequences editing and alignment

Output .abi sequence files were edited using CHROMAS Lite v2.01 (Technelysium Pty Ltd., Australia). All isolates were identified by BLASTn alignment and selection of the closest cultivated strain. All sequences were also analysed with CHIMERA_CHECK for possible artefacts prior to alignment. The alignment was done by creating a single fasta file with BioEdit Sequence alignment editor v7.0.0 (Copyright© Tom Hall) and uploading to the Sina internet based aligner (http://www.arb-silva.de/aligner/) and the ClustalW aligner of MEGA v4.0. The alignment was crosschecked with Geneious™ v5.1.7 and ClustalX v2.0. In all cases (clone sequences) 2-5 closest match sequences were added from the NCBI database with at least one type strain.

2.8.2 Tree construction

Tree construction was done through either Geneious™ v5.1.7 (Biomatters Ltd.) or MEGA 4 construction by using a Tamura-Nei distancing model and a neighbour joining tree building model unless otherwise stated. The resultant rooted tree topologies were re-evaluated by Bootstrap analysis with 100 000 seed counts and 1000 resamplings. A maximum likelihood Heuristic method with a nearest neighbour-interchange was used for the construction of the novel isolate phylogenetic tree based on full length sequences (1300 bp). In all except the novel isolate tree, the archaeon *Methanocalculus halotolerans* was used as an outgroup whilst

Bogoriella caseolytica was used as an outgroup for the two novel *Streptomyces* isolates.

2.8.3 DGGE profile analysis

DGGE profiles were analysed by using GELCOMPARE II© v 6.0 (Applied Maths NV®). Ethidium bromide stained gels were photographed using an Alphaimager 3400 Imaging System UV transilluminator (AlphaInnotech CorporationTM San Leandro, CA). The16-bit TIFF files were subsequently subjected to background subtraction to limit background noise. Band intensity was defined by an arithmetic average and least square filtering, while identification of bands were done by applying a minimum profiling of 5%, gray zone of 1% and shoulder sensitivity of 2%. In all cases sample LA was used as a standard reference. Cluster analysis was done by using a curve based Pearson correlation similarity coefficient unless otherwise stated.

2.8.4 Biogeochemical data analysis

A geographical distance matrix was constructed using decimal degrees and a perpendicular distance calculator with a UE3 spherical function (Ersts, 2011). The resultant matrix was used along with an OTU overlap matrix. Both matrices were compared in a Mantel test using XLSTAT software. The null hypothesis stated that there existed no relationship between the datasheets whilst the alternative hypothesis suggested that a negative relationship existed. The p-value was calculated by using the distribution of r(AB) estimated from 10000 permutations.

All physico-chemical data collected *in situ* were compared using The Unscrambler® (Camo software, AS) and principle component analysis (PCA).

2.8.5 Diversity, Richness and Evenness estimators

An operational taxonomic unit (OTU) was defined as a sequence with less than 99% 16S rRNA sequence similarity to any other clone sequences from the same 16S rRNA library. The phylotype richness, S, was calculated as the percentage of the total number of distinct ARDRA patterns to clones/OTUs (Dunbar *et al*., 2000).

The Shannon-Weiner diversity index (Shannon, 1948) was calculated as follows:

Equation 2 $H = -\sum(\rho i)(\log_2 \rho i)$,

where ρ is the proportion of an unique ARDRA pattern relative to the sum of all patterns. Evenness (Krebs, 1989) was calculated from the Shannon-Weiner diversity function as follows:

Equation 3 $E = H/H_{max}$ where $Hmax = Log_2(S)$

The Simpson Dominance index (Buckley, 1998) was calculated by using the following equation:

Equation 4 $SI' = n(n-1)/N(N-1)$

2.9. Actinobacterial isolation from soil

2.9.1 Media and growth conditions

Four selected media were used to isolate actinobacteria including: ISP2 (yeast extract-malt extract agar) (Shirling & Gottlieb, 1966), Medium A (Sato *et al*., 1983), M233 Hi-nutrient media (High media Ltd.) and Cellulose-casein-multisalts media (Tang *et al*., 2008). A pH range of 7-10.1, NaCl range from 1-5% and temperature range of 28°C, 30°C, 37°C, 45°C and 55°C were tested for each solid media type. All initial isolation media contained 50 µg/ml Cycloheximide to limit fungal growth. One gram (1g) of soil from each site was serially diluted in 1x PBS (phosphate buffered saline) and dilutions 10^{-2}-10^{-7} were plated. All plates were incubated for 8-10 weeks. A further 2g of soil was inoculated into liquid broth (medium A pH 8.0, Luria broth pH 8.0) and incubated at 37°C for 2 days before plating.

2.9.2 Storage of pure cultures

All cultures are stored in 20% glycerol stocks in their respective growth media at both -20°C and -80°C. Further stab cultures were kept at 4°C and room temperature.

2.10. Characterization of novel isolates

2.10.1 Growth conditions

All novel isolates were grown in Medium A pH 10.1 without antibiotic supplementation at 37°C for 4 days, 120 rpm shaking unless otherwise stated.

2.10.2 Storage

All strains were grown for 2 days (see 2.10.1) before addition of 20% glycerol and storage at -20°C. Stab cultures of all strains were kept at room temperature on ISP2 medium.

2.10.3 Degradation of compounds

As per International *Streptomyces* Project (ISP) guidelines, strains were tested for enzymatic activity. In all cases a single colony was streaked on the characterization media. In the case of pectinase activity, pectin hydrolysis medium was used at pH 7.0 and 8.0. Lipolysis and lecithinase activity was verified with egg yolk medium. Several additional compounds (1% w/v) were analysed for degradation including: gelatin, xylan, adenine, guanine, hypoxanthine, cellulose, elastin, starch, L-tyrosine, casein, Tween 80 and urea.

2.10.4 Pigment production

Pigment production was tested by means of ISP guidelines on ISP 5, 6 and 7 at both pH 7 and 8, with and without the addition of Na_2CO_3 and trace elements.

2.10.5 Antibiotic susceptibility

All antibiotics were tested at standard concentrations (see Table 9) in Bennett's glycerol medium agar at pH

8.0 with and without trace elements. In all cases the agar was tested after pouring with universal pH strips to ensure no change in pH had occurred.

Table 8. Antibiotic concentrations at which resistance was tested.

ANTIBIOTICS	FINAL CONCENTRATION
Cephaloridine	100 µg/ml
Gentamicin	100 µg/ml
Lincomycin	100 µg/ml
Neomycin	50 µg/ml
Oleandomycin	100 µg/ml
Penicillin G	10 I.U.
Rifampicin	50 µg/ml
Streptomycin	100 µg/ml
Tobramycin	50 µg/ml
Vancomycin	50 µg/ml
Ampicillin	100 µg/ml
Tetracycline	25-50 µg/ml
Cycloheximide	50 µg/ml
Chloramphenicol	50 µg/ml

2.10.6 Spore characterization

All mycelial and spore structures were observed after cultures were incubated on ISP4 for 2 weeks at 37°C (pH 9.0). Photographs were taken at 400x and 1000x magnification on a light microscope.

2.10.7 Resistance against inhibitory compounds

Resistance against phenol (0.001g/L) and crystal violet (0.001g/L) was tested on Bennett's medium pH 8.0, with and without trace elements and 10% Na_2CO_3 over a period of 21 days.

2.10.8 TLC of Cell wall sugars, phospholipids and DAP

A 500ml culture of medium A pH 10.1 was inoculated with a water spore suspension corresponding to 4 x 10^8 (ERV7) and 6 x 10^8(ERV8) cfu. After a two week period the cells were harvested and washed with TE buffer pH 8.0 to remove all excess media components. Cells were then freeze dried in a FreeZone 1 Liter Benchtop Freeze Dry Systems (Labconco©) before further processing. Cell wall sugars and diaminophosphates were extracted according to Hasegawa *et al.,* (1983) and Staneck and Roberts (1974) and the thin layer chromatography done as stipulated in each article. The standards included diaminopimelic acid, glycine, glucose, mannose, ribose, galactose, xylose and arabinose at a concentration of 1% w/v. The phospholipids extractions were done according to Minnikin *et al.,* (1984) and separated on a 10 x 10 cm silica gel 60 F_{254} TLC plate with a chloroform-methanol-water (65:25:4, v/v) liquid phase for the first dimension; and chloroform-acetic acid-methanol-water (40:7.5:6:2, v/v) liquid phase for the second dimension. The silica plate was then sprayed with α-Naphtol and ninhydrin solutions followed by charring at 140°C for 10 min to visualize the spots. *Dietzia maris* cell wall extracts were used as a control.

2.10.9 Xylanase characterization

2.10.9.1 Xylanase plate assays

Both strains were inoculated onto 4-O-methyl-D-glucurono-D-xylan-remazol brilliant blue (RBB-xylan; Sigma™) plates containing 1% w/v birchwood xylan on a basal media of Bennett's medium pH 10.0 incubated at 37°C for 24 hours. Activity zones were measured by subtracting the radius of the colony from the radius of the clearance zone.

2.10.9.2 Growth conditions and crude extracts

Spore suspensions of both strains were inoculated into Erlenmeyer flasks containing 400 ml of Bennett's medium pH 10.0 and 0.1 or 1% w/v birchwood xylan. After an incubation period of 24 and 48 hours a 1 ml and 10 ml aliquot was taken. The first was used as is in the following assays whereas the 10 ml aliquot was subjected to three cycles of 30 sec sonications. The resultant suspension was then centrifuged and the supernatant used for further analysis.

2.10.9.3 DNS assays

The *3,5-Dinitrosalicylic acid* (DNS) assay was performed as described in Saleem *et al*., (2008) by using pure and diluted crude extract (see section 2.10.9.2). The standard curve was constructed by using 99.9% pure xylose dissolved in the following buffers: MES pH 5.5, PIPES pH 6.5, HEPES pH 7.9, Tris pH 8.5 and CAPS pH 10.0. Samples were incubated for 30 and 60 min periods before measurement on a spectrophotometer at 595nm. Xylose concentrations were determined via a polynomial equation derived from standard curves whereby activity was measured in IU units (mM xylose liberated/min). Standard deviations were obtained through triplicate repeats.

CHAPTER 3: CULTURE-INDEPENDENT ANALYSIS OF EUBACTERIAL AND ACTINOBACTERIAL DIVERSITY: RESULTS AND DISCUSSION

3.1 Introduction

Extreme environments represent the ultimate boundary in which life can exist. Several parameters have been shown to limit bacterial growth including hypertonic environments (high salt or mineral content), alkaline or acidic extremes and temperature extremes. Each extremophilic organism can therefore be classified according to its ability to tolerate one extreme. Polyextremophilic environments present a new problem, whereby cellular mechanisms aren't solely devoted to counterbalance one abiotic stress but several (Bowers *et al*., 2009). In the case of the Ethiopian Rift Valley lakes, there exist both extremophilic and polyextremophilic environments. Although most of the lakes are classified as having high alkalinity and a moderate salt content (0.5-3.5% NaCl), the addition of hot springs creates a hyperthermophilic environment with altered monovalent cation equilibriums. Nevertheless, these extreme environments have proven to be one of the most productive ecosystems in the world averaging between 8-10 g cm^{-2} $day^{-1}$of biomass, sixteen times higher than the average lake system (0.6 g cm-2 day^{-1}) (Belachew, 2010; Rees *et al*., 2004).

A high productivity rate, under normal conditions (25°C, pH 6-7.5, 20% oxygen, atm pressure), would translate into increased biomass production or biodiversity (Mesbah & Wiegel, 2008). However, the increase in primary productivity in soda lakes has not been linked to the same phenomenon.

3.2 Culture-independent analysis of microbial diversity

3.2.1. Extraction of genomic DNA

Three separate protocols were used to extract metagenomic DNA from the soda lake sediment. The DNA concentration was measured spectrophotometrically at A_{260} as well as determined using gel electrophoresis. Genomic DNA shown in Fig.9 was clear with limited impurities with a yield estimated at between 1 and 5 µg/µl total DNA per gram wet weight sediment (Table 10). Although concentrations of up to 34 µg/µl have been obtained from similar soda lake environments, concentrations of 1-10 µg/µl are regarded as a relative average for hot spring environments (Baumgarte, 2003; Herrera & Cockell, 2007).

Table 9. Metagenomic DNA extraction yields from sediments using three different extraction protocols.

Sample ID	µg/µl	A260/A280	A260/A230
LA a	1.068	1.54	1.18
LA b	3.932	1.64	1.28
LA c	3.756	1.5	1.36
LAb a	2.868	1.74	1.3
LAb b	2.655	1.72	1.29
LAb c	2.829	1.74	1.31
HS1 a	3.355	1.66	1.23
HS1 b	3.874	1.63	1.9
HS1 c	4.875	1.76	1.54
HS2 a	4.137	1.74	1.29
HS2 b	5.504	1.34	1.31
HS2 c	4.033	1.65	1.81
HS3 a	4.119	1.82	1.38
HS3 b	4.339	1.78	1.6
HS3 c	4.505	1.67	1.6

a Wang DNA extraction method
b Hard Lysis DNA extraction method
c Modified Zhou DNA extraction method

Overall, the 260/230 ratios were significantly higher than would be expected for DNA of optimum purity (0.3-0.9) indicating contamination by organic compounds and salts. The 260/280 ratios ranged from 1.64 to1.8 and showed that the DNA was relatively pure, except for LA a, LA c and HS2 b where the lower values may be an indication of protein contamination.

Several DNA extraction protocols have been employed for metagenomic DNA isolation, all of which result in a crude extract with some degree of contamination (Herrera & Cockell, 2007; Krsek & Wellington, 1999; Miller *et al*., 1999; Robe *et al*., 2003). A large variety of DNA purification methods exist to remove inhibitory compounds including several commercial kits (i.e. MoBio© Powermax®/Ultraclean® , Epicenter Meta-G-nome ®and Norgen© soil DNA isolation kit) and PVPP treatment (Kirby *et al*., 2011; Nannipieri & Smalla, 2006). Direct isolation methods, such as the ones employed in this study, result in extraction of more inhibitory compounds, but also increase the DNA yield. Three purification methods including PVPP columns, MoBio© purification columns and agarose plugs were used to minimize humic acid contamination, however, due to the resultant DNA loss, were abandoned (results not shown).

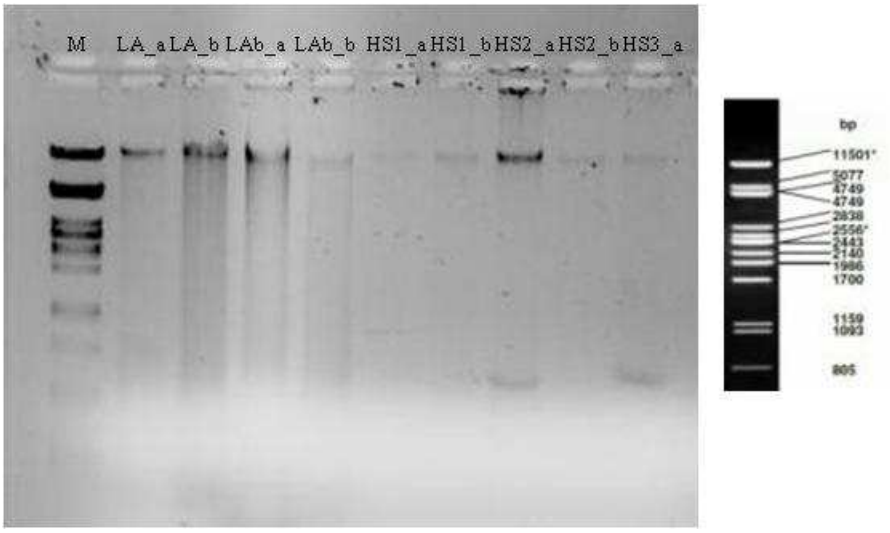


Figure 9**.** A 1.0% agarose gel of the separate genomic DNA extractions. Each well contains a 5µl aliquot of DNA corresponding to 0.01-0.1µg of DNA. The DNA marker is a *Pst* I digest of lambda DNA.

Humic acids are the most common inhibitory compounds that are co-extracted during nucleic acid extraction from sediment (Wintzingerode *et al*., 1997). Humic acids are a heterogenous group of high molecular weight compounds created during the polycondensation of aromatic and nitrogenous compounds (Pospisilova & Fasurova, 2009). As a by-product of organic matter decomposition they are ubiquitous in terrestrial environments and are resistant to microbial degradation. In sandy, leached soils the largest proportion of organic matter is in the form of fulvic acids, whereas alkaline and neutral soils tend to contain larger quantities of humin and humic acids. Each compound may contain several phenolic groups. These may either bind directly to amide groups within biological molecules or bind DNA covalently following oxidation to quinines. The binding to DNA inhibits several downstream applications, including the reliable quantification of DNA due to the co-absorbance of light at 260 nm (Wintzingerode *et al*., 1997).

There are several methods to limit the co-extraction of humic substances. However, most of the known techniques do introduce a certain amount of bias to the process. Indirect approaches require the removal of cell matter from soil particles (Robe *et al*., 2003). These methods have been shown to limit humic extraction in dry soils compared to 'wet' samples prior to nucleic acid isolation. The adsorption of cells and so-called 'free' DNA onto soil particles makes this a less favourable approach due to the loss of a large proportion of the metagenomic DNA. Direct extraction, on the other hand, increases DNA yield, but coextracts more inhibitory compounds. Since soft-lysis methods are ineffective for Gram-positive bacteria, fungi and algae, hard lysis methods are often employed (Robe *et al*., 2003; Miller *et al*., 1999).

The three extraction techniques employed in this study make use of both direct soft and hard lysis methods with variations in extraction buffers. During soft lysis, lysozyme is used to disrupt the cell wall by cleaving the β-1, 4-glycosidic linkage between the N-acetymuramic acid and N-acetylglucosamine repeating unit of

the peptidoglycan layer of the cell wall. Proteinase K, a serine proteinase that cleaves the peptide cross-linkages, breaks down the cell wall further. The Wang method uses these two enzymes with a basic EDTA buffer, whereas the modified Zhou method uses a CTAB-sodium phosphate-NaCl EDTA buffer (Wang *et al*., 1996; Stach *et al*., 2001). The last mentioned buffer is slightly different due to the addition of a cationic detergent (CTAB), which denatures and precipitates lipopolysaccharides and cell wall proteins. Polysaccharides act as an inhibitor in enzyme based downstream applications and are more abundant in Gram-positive bacteria and fungi (Niemi *et al*., 2001).

3.2.2. PCR amplification of the 16S rRNA gene

Crude genomic DNA extracted from all five Ethiopian Rift Valley sites were used as template DNA in 16S rRNA gene PCR assays. Published universal bacterial and actinobacterial-specific primers were used in combination with several DNA polymerases (see section 2.3.2). A range of DNA concentrations were tested for PCR amplification from undiluted to between 1-10ng/µl. A 10^{-3} dilution of genomic DNA resulted in amplification from all templates at a concentration of 1-5 ng (Fig.10).

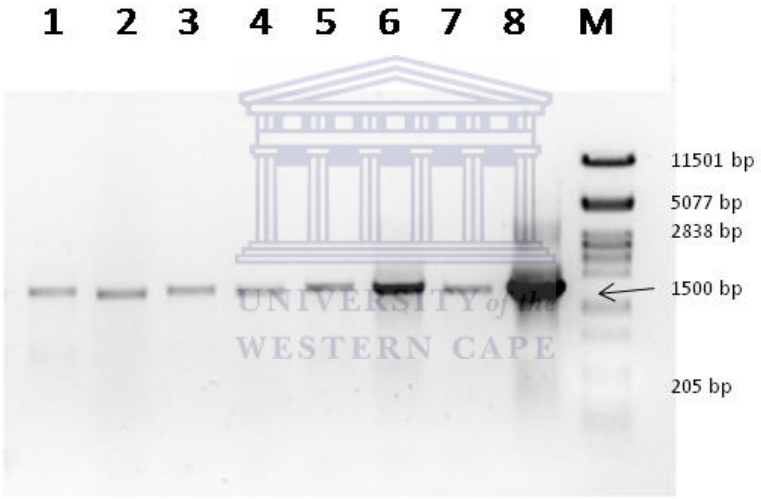


Figure 10. PCR amplification of 16S rRNA gene of approximately 1.5 kB in size. Lane 1: Lake Arenguade, Lane 2: Lake Abitjata, Lane 3-5: Lake Shala hot spring1, 2, 3. Lane 6: *E.coli* positive Lane 7: *Streptomyces* sp. positive with BSA, Lane 8: *Streptomyces* sp. positive, M: Lambda *Pst* I ladder.

PCR fragments of approximately 300 bp were amplified with the actinobacterial-specific primers for all DNA samples (Fig.11).

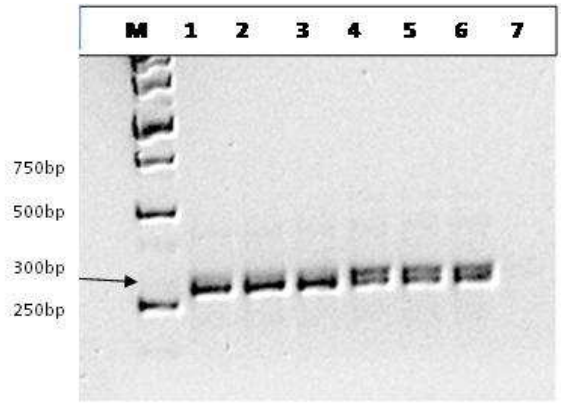


Figure 11. Actinobacterial-specific 16S rRNA gene fragment amplification. Lane 1: Lake Arenguade, Lane 2: Lake Abitjata, Lane 3-5 Lake Shala hot springs 1, 2, 3, Lane 6: Positive control, Lane 7: No DNA control, M: Generuler 1kb PLUS ladder.

3.2.3. PCR amplification of DGGE fragments

A nested PCR was used to amplify the ~200 bp 16S rRNA gene fragment from the initial full 16S rRNA gene for DGGE analysis (343-523bp *E. coli* numbering). Due to the large impact of inhibitors on PCR amplification, nested PCR was performed on the full 16S rRNA amplicons using the primer set 341F-GC and 534r (Muyzer *et al.,* 1993). This drastically improved the visibility of amplicons on DGGE gels (data not shown). Initially, three polymerases were tested and the best amplification was obtained with KAPA2G Robust *Taq* polymerase (two appear in Fig.12). For DGGE gels two 25 µl reactions were pooled to increase the detection of faint bands.

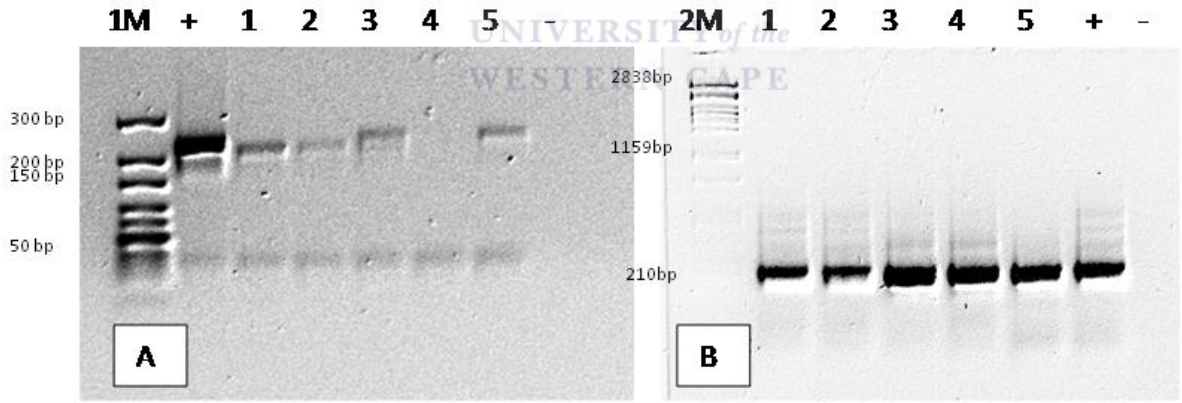


Figure 12. Amplification of 16S rRNA gene fragment with primers containing a GC-clamp for DGGE analysis. Figure A: Dreamtaq amplification of nested PCR, (~200 bp fragment) 1M: Generuler ULTRA low range ladder. Figure B: Robust *Taq* amplification of nested PCR (~200 bp fragment) 2M: Lambda *Pst*I ladder, Lane 1: Lake Arenguade, Lane 2: Lake Abitjata, Lane 3-5 Lake Shala hot springs 1,2,3, +: Positive control, -: No DNA control

3.2.4. Community analysis via denaturing gradient gel electrophoresis

3.2.4.1. Optimising DGGE profiles by limiting PCR bias

DNA from all sampling sites amplified with actinobacterial-specific primers. Amplicons were separated on a 40-80 and 60-80 % urea formamide denaturing gradient gel (see section 2.7). Both gradients generated the same banding patterns (data not shown). The reproducibility of the DGGE fingerprint profiles, as well as the effect of metagenomic DNA template concentration was tested. Amplification with diluted and undiluted template DNA using DreamTaq DNA polymerase failed to provide distinctive banding patterns at this resolution (Fig.13). As all samples amplified with Robust KAPA2G, a high fidelity *Taq* polymerase resistant to low amount of inhibitors, this polymerase was chosen for all subsequent PCR amplifications.

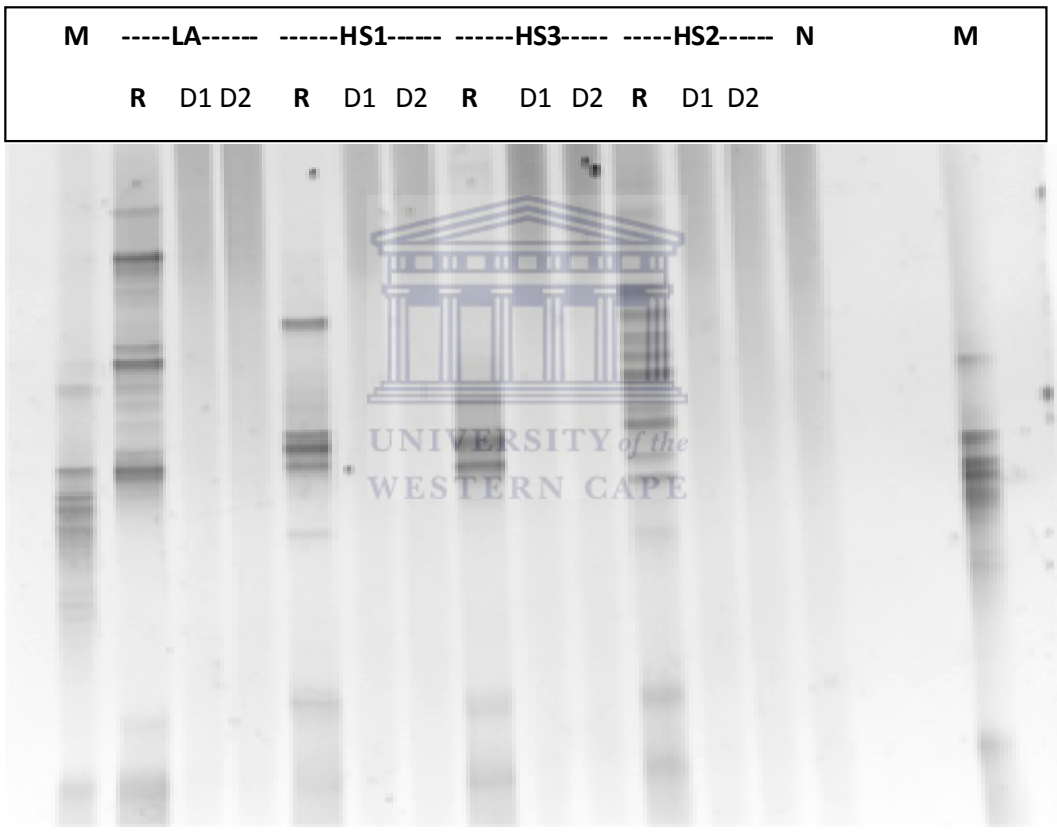


Figure 13. A denaturing gradient gel with a 30-80% denaturing gradient of the actinobacterial-specific 16S rRNA gene amplicons from four sampling sites. M: Ladder N: Negative control (no DNA). LA: Lake Arenguade, HS1-3: Lake Shala Hot spring 1, 2, 3.R: PCR was conducted with KAPA 2G Robust *Taq* polymerase. D1: Undiluted DNA template amplified with DreamTaq DNA polymerase. D2: Diluted 10^{-2} DNA template amplified with DreamTaq DNA polymerase.

Two dimensional analysis of the EtBr stained DGGE gels was conducted to compare the reproducibility of banding patterns from the three different metagenomic DNA extraction methods (see Fig. 14,15). In all cases only the dilution factor differed with regards to PCR conditions. Metagenomic DNA isolated by means of the modified Zhou (Stach *et al*., 2001) and Wang *et al.,* (1996) methods were diluted 1000 times (1-5ng/µl) due to the inhibitory compounds co-extracted. Metagenomic DNA extracted via the hard lysis method (see

section 2.2.1) was used undiluted or in a 10 fold dilution (10-500ng/µl). Initial concentrations of genomic DNA were relatively similar between samples for all isolation methods except for Lake Arenguade which had a considerably lower genomic DNA concentration from the Wang *et al.,* (1996) method (see Table 10). A Universal 16S rRNA gene amplification followed by a nested actinobacterial-specific PCR with a GC-clamp was conducted to compare triplicates.

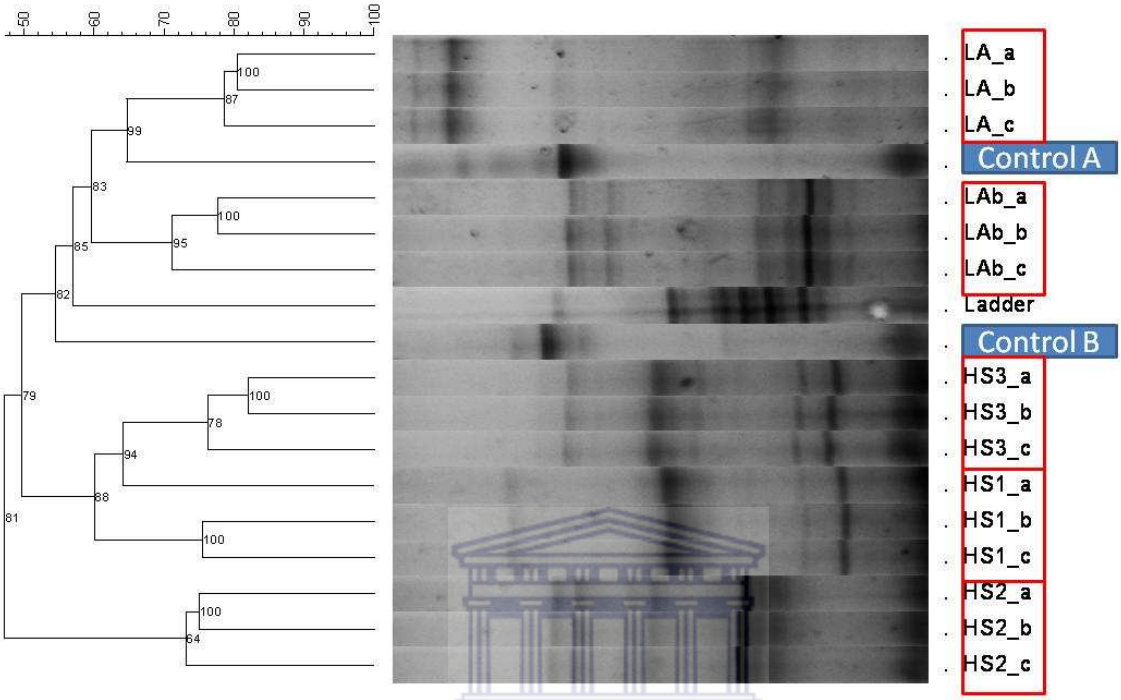


Figure 14. Cluster analysis of the DGGE profiles indicating the differences between the three metagenomic DNA extraction methods and the influence on community profiles. The dendogram is based on UMPGA and DICE based analysis. LA: Lake Arenguade, Lab: Lake Abitjata, HS1-3: Lake Shala Hot spring 1, 2, 3. _a: Modified Zhou DNA extraction. (Stach *et al.,* 2001) DNA extraction, b_: Wang *et al.* (1996) DNA extraction, _c Hard Lysis method DNA isolation. Control A: *Streptomyces* sp.ERV8 Control B: *Streptomyces* sp. ERV7.

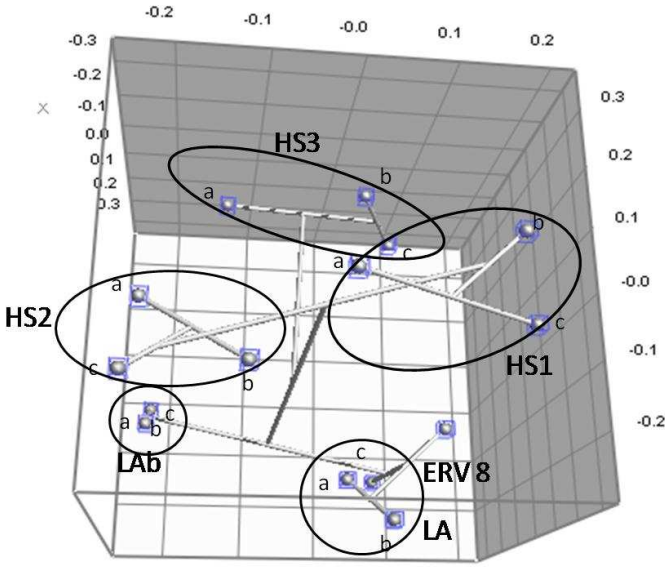


Figure 15. MDS plot of the Actinobacterial-specific DGGE profiles.

Although all replicates within a sampling site grouped together in the cluster analysis (Fig.14), slight differences were detected between banding patterns. The MDS plot (Fig.15) clearly illustrates the grouping of NA extraction methods within each site, however none were identical. Previous studies comparing different metagenomic DNA extraction methods from soil indicated that a combination of bead beating and SDS treatment gave the purest DNA with the most discernable DGGE profiles (Aguilera *et al*., 2006). Due to the high similarity between the DGGE patterns, the genomic DNA isolated with the Hard Lysis method (using SDS and bead beating) were used for all further analyses.

The novel isolates obtained via culture-dependent analysis was included as controls. Isolate ERV8 (see section 4.2), grouped with the Lake Arenguade and Lake Abitjata samples. This was to be expected since it was isolated from Lake Abitjata. ERV7 was isolated from Lake Chitu, a lake located adjacent to Lake Shala and Lake Abitjata, and this isolate formed a separate branch.

3.2.4.2. DGGE analysis of Eubacterial diversity

Once the reproducibility of DGGE profiles was established, the universal eubacterial amplicons were analysed on a range of different denaturing gradients. Since there exists a general consensus regarding the role of both high and low G+C bacterial phyla in soda lakes, a wide and narrow gradient was studied in order to analyse the community profile (Duckworth *et al*., 1996; Jones *et al*., 1998; Baumgarte, 2003; Rees *et al*., 2004; Zhilina & Zavarzin, 1994). Analysis on a 30-80% denaturing gradient provided ample resolution to identify between 22-39 OTUs (with varying levels of brightness). Each distinct band was defined as an operational taxonomic unit (OTU). Figure 16 depicts a dendogram showing the grouping of the OTUs from five sites based on the eubacterial subpopulation. It is evident that the Lake Shala hot springs are similar with HS2 and HS3 grouping together with HS1 being the closest branch. Lake Abitjata and Arenguade, on the other hand branched completely separately. The reliability of the dendogram is indicated by the cophenetic correlation values at each node. In the current study a value of >75% was used to indicate a significant cluster. All cophenetic correlations in Fig.16 were higher than 96%.

Multi dimensional scaling (MDS) was used to visualize similarities between sites based on the clustering in 3-dimensional space. The clustering patterns for both the MDS plot (Fig.17) and the dendogram were similar suggesting that all of the sites are similar in community composition, although the three separate lakes are more dissimilar than the three hot springs situated within the same lake.

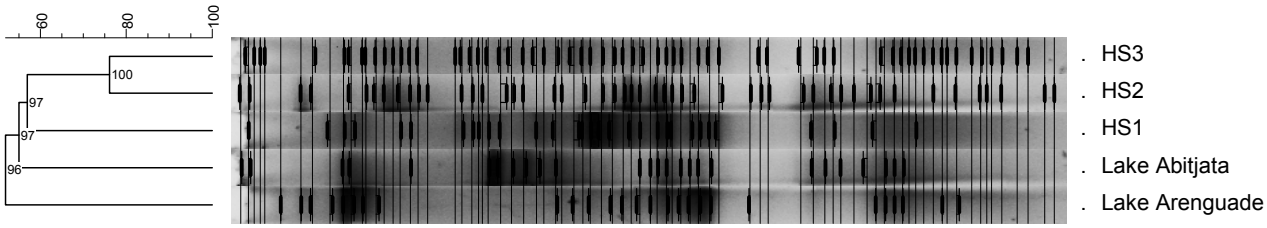


Figure 16. Band matching across the eubacterial community fingerprints. Each band in a fingerprint appears as a bold broken line. Band classes, matching bands across fingerprints, are represented by a solid line. HS1-3: Lake Shala Hot spring 1, 2, 3.

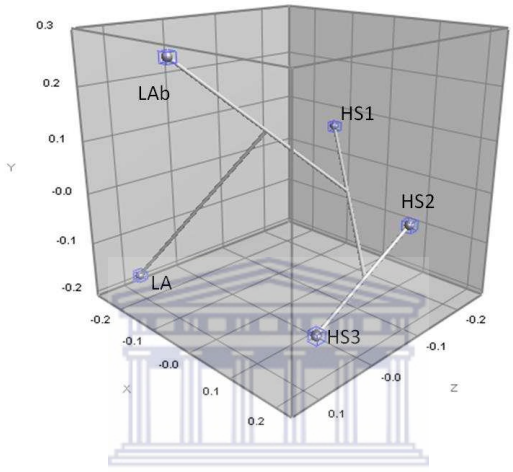


Figure 17. MDS plot of the eubacterial DGGE profiles as shown in fig.16

Separation of eubacterial amplicons on a 60-80% denaturing gel repeatedly resulted in a low resolution smear (at 60% denaturant) and faint bands at higher denaturing gradients (See Fig.18). Therefore, relatively few high G+C OTUs (such as expected for actinobacteria) could be clearly identified using universal primers 341F-GC and 534r.

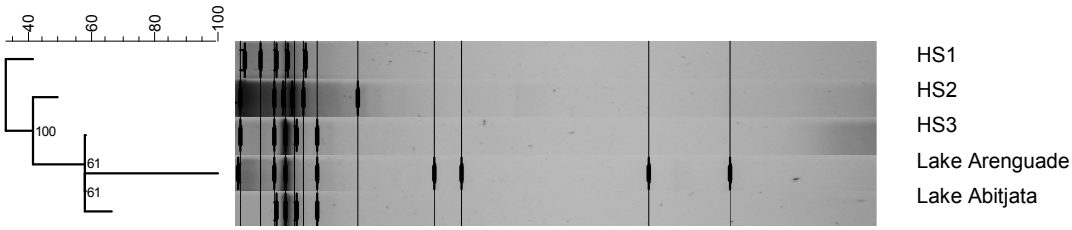


Figure 18. The DGGE profiles of all eubacterial amplicons on a 60-80% denaturing gradient. HS1-3: Lake Shala Hot spring 1, 2, 3.

The eubacterial community similarities between the Rift Valley lakes seem to remain similar when

comparing the two dendograms (Fig. 16 and 18). Although the second dendogram is by definition non-conclusive (due to the limited gradient and low resolution), the separation of Lake Shala from Lake Arenguade and Lake Abitjata is evident. The correlation between Lake Arenguade and Lake Abitjata will be discussed further in section 3.2.4.3

3.2.4.3. DGGE analysis of Actinobacterial diversity

Previous studies investigating the microbial diversity within soda lakes have generally detected low levels of actinobacteria when using 'universal' bacterial primers (Jones & Grant, 1999; Baumgarte, 2003; Belachew, 2010). This has been attributed to the fact that actinobacterial species are a minor component within soda lake soil communities, even though the class is ubiquitous in terrestrial environments (Heuer *et al*., 1997). Therefore, in the current study specific primers were employed in order to detect members of the class actinobacteria. A nested primer set amplifying the highly variable γ region (variable 3 region) of the 16S rRNA gene was used in order to bias the PCR towards the amplification of actinobacteria.

DGGE analysis of actinobacterial-specific amplicons (see section 2.7) was conducted using the same approach as was used with the eubacterial amplicons (see section 2.3.2). UPGMA analysis of the actinobacterial DGGE profiles showed a different clustering pattern to the eubacterial dendogram. Lake Arenguade and Abitjata had a higher similarity whereas all of the Hot spring sites branched separately suggesting a completely different actinobacteria population (see Fig.19). All of the cophenetic correlation values were above 96% indicating confidence in all clusters.

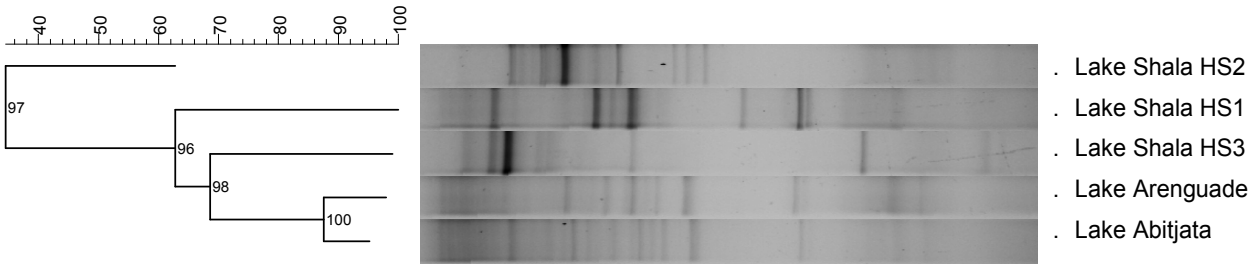
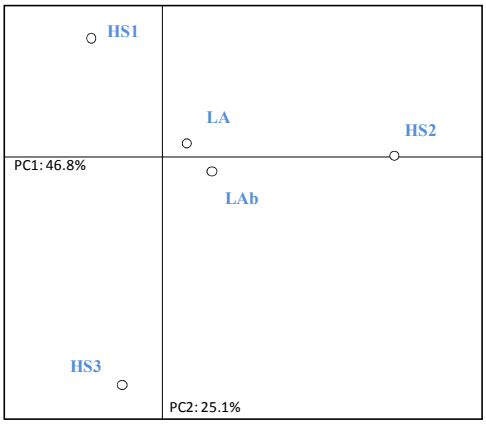


Figure 19. A dendogram of the Actinobacterial DGGE profiles of all sites as derived by UMPGA and DICE band-based analysis.

A PCA plot of the above mentioned profiles (Fig.20) indicated the same clustering of Lake Arenguade and Abitjata, with the hot springs separate from each other and from the mesophilic lake samples. The first two components represent 71.9% of the variance between the samples, clearly showing no clustering of the three hot springs substantiating the dendogram analysis.



Component	Contribution
1	46.8%
2=Y	25.1%
3=X=Z	21.2%
4	6.0%
5	1.0%

Figure 20. Principle Component Analysis (PCA) graph of actinobacterial DGGE profiles. LA: Lake Arenguade, LAb: Lake Abitjata, HS1-3: Lake Shala Hot spring 1, 2, 3.

3.2.4.4. DGGE analysis of Archaeal diversity

The lack of complete archaeal 16S rRNA gene sequences has hampered the design of 'universal' archaeal primers for phylogenetic studies. Published universal archaeal primers are therefore extremely biased towards classes that have been cultured and may fail to detect unique species. This also accounts for the fact that most PCR-based studies have grossly underestimated the diversity of archaea in extreme environments. Microscopy and isolation studies have confirmed that archae are in fact present and flourish in thermophilic and halophilic environments (Malkawi & Al-omari, 2010; Oren, 2002). The archaeal primers used in this study were designed to span the V3 region of the 16S rRNA gene.

DGGE only identified 6-11 bands for each site, where Lake Arenguade and Lake Abitjata showed the least diversity with 6 and 8 OTUs, respectively (Fig.21). The cluster analysis showed a larger similarity between Lake Arenguade and Lake Abitjata, with a vastly different and separate branch for each of the Hot springs.

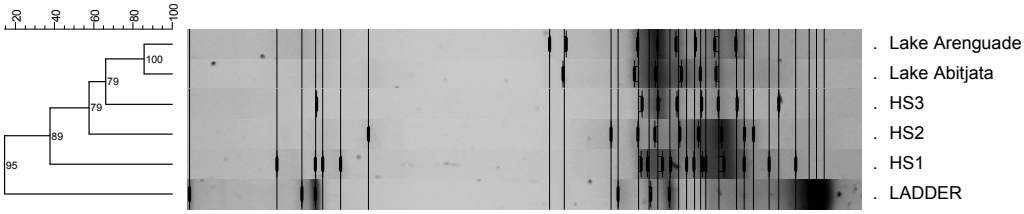


Figure 21. A dendogram of the archaeal DGGE profiles from all the sites.

3.2.5 Cloning and amplified ribosomal DNA restriction analysis (ARDRA)

Several fingerprinting techniques exist of which DGGE, ARDRA and T-RFLP are the most widely published techniques. Whilst the DGGE profiles can provide an idea of the differences between sites it cannot be used to extrapolate the diversity within each site. Clone libraries provide a vastly higher resolution image of the

diversity within the lakes when coupled with 454 sequencing and screening methods such as ARDRA or T-RFLP. ARDRA profiling can be used at both the species and strain level to distinguish ribotypes or isolates based on unique restriction patterns. During the course of the past decade ARDRA has been applied to the characterization of several bacterial strains (Mañes-Lázaro *et al*., 2008; Fortina *et al*., 2009; Nemec *et al*., 2009) and has also found widespread recognition in screening of various sediment derived libraries (Baumgarte, 2003; Torsvik *et al*., 1998; Sjoling & Cowan, 2003; Islam *et al*., 2011).

In this study all clones underwent a double digest with *Rsa* I and *Alu* I in a complimentary buffer to increase the sensitivity of each pattern. These restriction enzymes have been used with other endonucleases to distinguish between bacterial communities and actinomycete subpopulations (Vaneechoutte *et al*., 2000; Khetmalas *et al*., 2002; Messaoudi & Wagenlehner, 2010).

Three hundred colonies from each library were screened to confirm insert size (via M13 colony PCR) prior to ARDRA (see Fig.22). A hundred to a 150 positive clones containing a ~1500 bp M13 amplicon (eubacterial libraries) were selected for ARDRA screening. Similarly, all amplicons of 450bp-550bp from each actinobacterial library were selected for further analysis (see Fig.23).

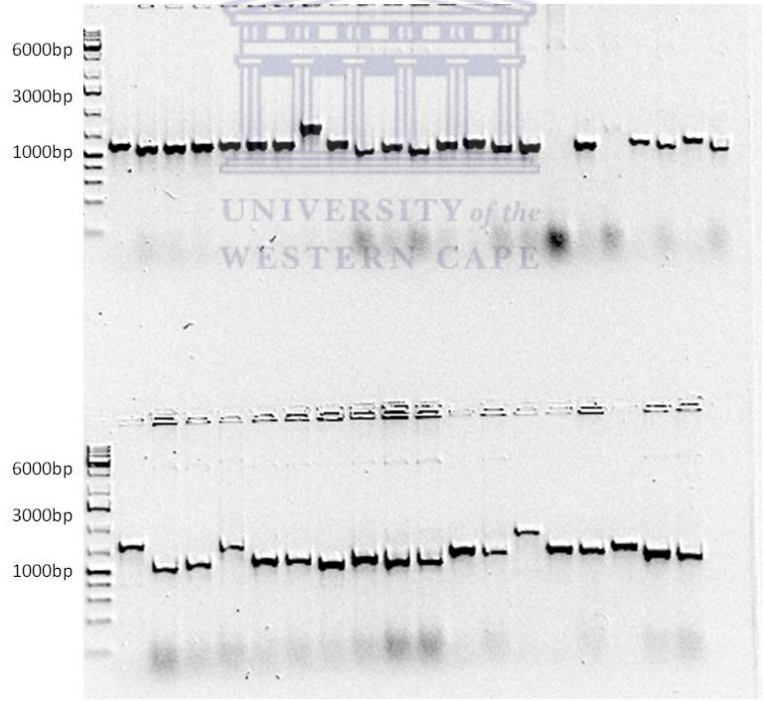


Figure 22. Insert size verification of Lake Arenguade eubacterial library clones using M13 colony PCR. 1.5% Agarose gel depicting the expected insert size of ~1250-1500bp. Marker: 1kb Fermentas Ladder

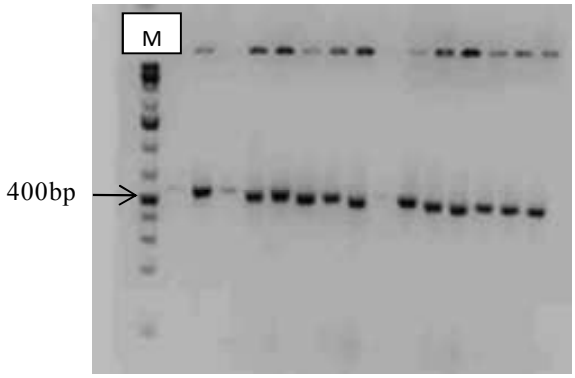


Figure 23. Insert size verification of actinobacterial library clones using M13 colony PCR. The expected size was 500bp. Marker: 1kB Fermentas Ladder

Amplicons containing the expected size inserts were subjected to ARDRA (Fig.24). Unique clones were identified by the size and number of fragments in each profile. An uncut vector was estimated to be in the order of 3.95 kB (Actinobacterial libraries) or 4.55 kB (Eubacterial libraries) and all partial digests were repeated.

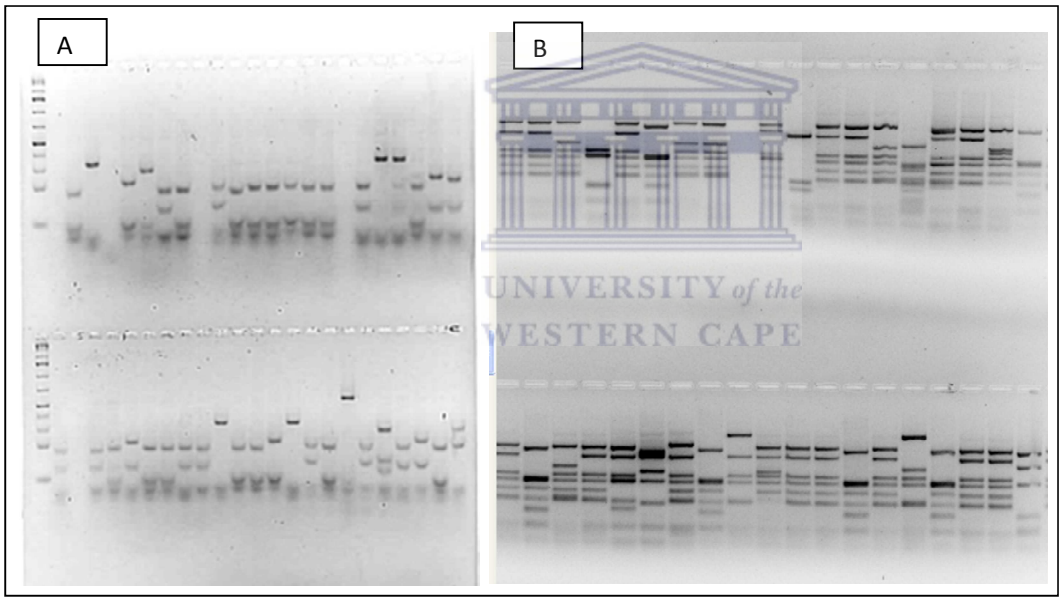


Figure 24. ARDRA profiles of 16S rRNA gene fragments generated by *Rsa*I and *Alu*I digestion. The digested products were separated on 3.5% agarose gels and visualised with GelRed or EtBr staining. A: ARDRA profile of the ~500bp actinomycete-biased libraries. B: ARDRA profile of the eubacterial libraries.

Of the 300 colonies screened from each sample site, 96 were selected for ARDRA, except for the eubacterial libraries from Lake Arenguade and Lake Abitjata where 150 colonies were screened as a larger number of single occurrence unique phylotypes were present (i.e. 9-12 phylotypes). Both *Rsa*I and *Alu*I were chosen due to their compatibility, 50% G+C content and short recognition sequence. Initial screening with *Taq*I produced banding patterns that could not distinguish between clones at the same resolution as *Rsa*I and *Alu*I therefore it was not used for further analysis. Due to the high G+C content of actinobacteria both enzymes provided suitable restriction profiles.

In total 480 clones were analysed for the actinobacterial libraries and 588 clones for the eubacterial libraries. Table 11 lists the number of unique phylotypes observed per sample site.

Table 10. The number of unique phylotypes for the individual libraries identified by ARDRA.

Site	Library	Nr. Of Phylotypes
Lake Arenguade	Actinobacterial	14
Lake Abitjata	Actinobacterial	14
Lake Shala HS1	Actinobacterial	19
Lake Shala HS2	Actinobacterial	14
Lake Shala HS3	Actinobacterial	12
Lake Arenguade	Eubacterial	23
Lake Abitjata	Eubacterial	20
Lake Shala HS1	Eubacterial	17
Lake Shala HS2	Eubacterial	16
Lake Shala HS3	Eubacterial	16

3.2.6. Composition of Eubacterial 16S rRNA gene libraries

A total of 94 unique OTUs were identified using ARDRA, of which 92 were confirmed as unique phylotypes (sequences <99% similarity). As seen in Fig.25 eubacterial composition of each site varied dramatically. The hot springs were mainly dominated by members of the phyla α-Proteobacteria and Firmicutes (predominantly *Bacillus* species), whilst the other two lakes were colonized by both Gram-positive and Gram-negative bacteria. A multivariate statistical analysis (PCA) was conducted on the number and occurrence of each genera or class listed in Fig.25. This was done to highlight the main discrepancies (i.e. variance) between the individual sites based on overall eubacterial composition. Statistical analysis of the microbial diversity of each sampling site revealed a clear difference between sites (see Fig.26) regardless of proximity. A PCA of the physico-chemical parameters measured *in situ*, divided the five sites into three groups representing the three lakes.

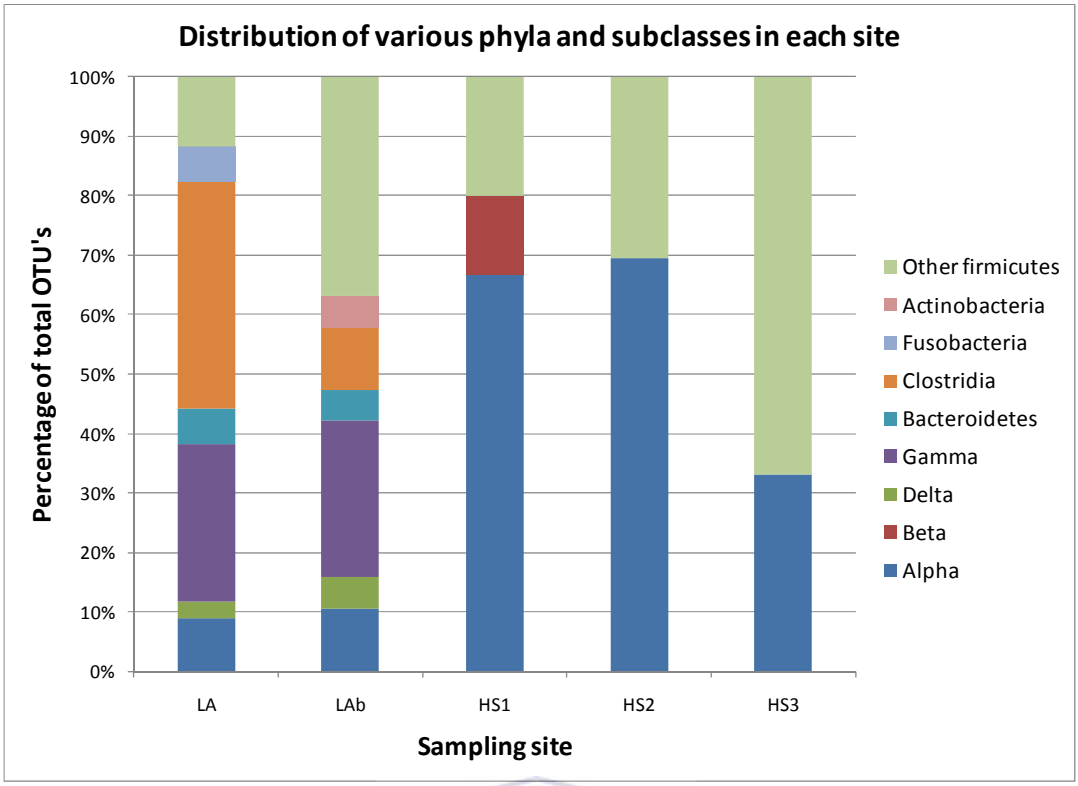
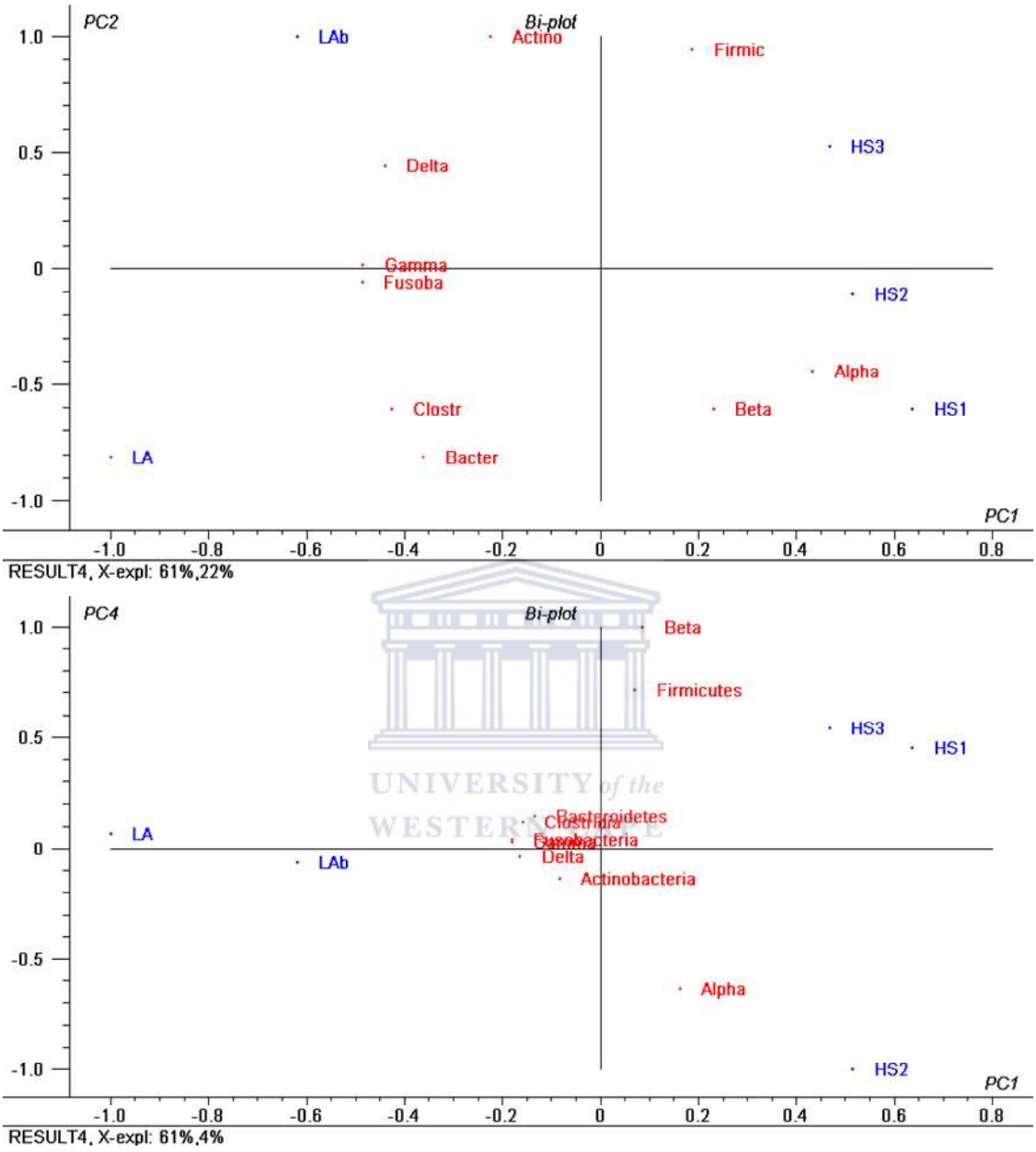


Figure 25. Eubacterial library composition of each sampling site. Each colour represents a different group of organisms as the percentage of total OTUs from that specific site.

Fig.26A is a bi-plot of the scores and loadings representing the various groups as they are separated based on the occurrence of a specific class. Lake Abitjata is clearly separated based on the occurrence of actinobacteria and δ-Proteobacteria, while Lake Arenguade is separated due to the abundance of *Clostridia* a lack of actinobacterial OTUs. The hot springs, on the other hand, are separated based on the variance in PC2 (22%) which was mainly due to the inverse correlation of α- and β Proteobacteria with other Firmicutes. Fig.26B clearly illustrates the separation of hot spring 2 based solely on α-Proteobacterial OTUs.

The diversity of each site was determined based on the number of occurrences of a specific OTU in each eubacterial library as calculated by the Shannon diversity index (Shannon, 1948). The index is highly informative when comparing similar environments, due to the universal indicator of both richness and evenness. It has been employed to study molecular data from environmental samples such as USA soil samples (Dunbar *et al.,* 2000), soda lake sediments in India (Wani *et al.,* 2006), Antarctic mineral soils (Smith *et al.,* 2006). Both the Shannon-Wiener (H') and Simpson Dominance (D) indices can be applied as a semi-quatitative estimate based on species accounts (which refers to OTUs in the current context). This is then used in conjunction with the qualitative estimators (i.e. Chao1, ACE or rarefraction analysis) to determine the full extent of the sample diversity. A value of 1.5 is indicative of a low diversity whilst 3.5 is a high diversity (MacDonald & MacDonald, 2003). All Shannon values for the selected sites indicate low diversity (1.02-1.679), with the exception of Lake Arenguade which has a Shannon value of 2.643 (Table 12). However, the species evenness is close to 1.0 for both Lake Arenguade and Lake Shala Hot spring 3,

suggesting a higher heterogeneity at these sites. The evenness values merely highlight slight differences in species composition as seen in the stacked bar chart in Fig.25. The phylotype richness (Table 12) varied between the different lakes, with Lake Arenguade (20%) having the highest followed by Lake Abitjata (10%) and



Lake Shala HS1 the lowest (6%).

Figure 26. Principle component analysis of the eubacterial composition of each sampling site. Above (A): PC1(61%) vs PC2 (22%) Below (B): PC1 (61%) vs PC4 (4%).

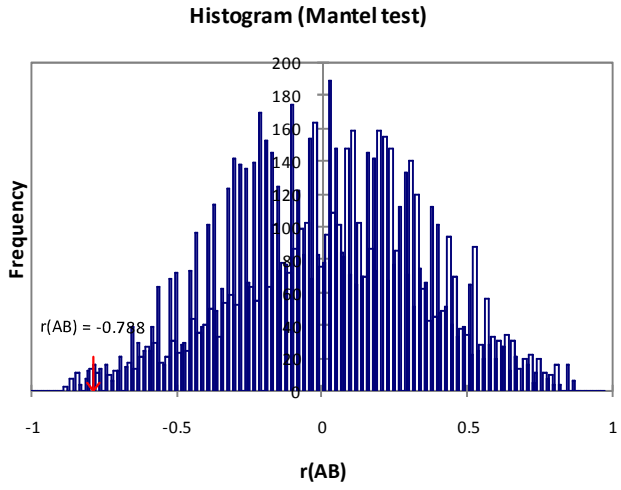
Table 11. Diversity, evenness, phylotype richness and library coverage estimators for the eubacterial libraries.

Index	16S rRNA clone Library				
	LA	LAb	HS1	HS2	HS3
S_a	21.00	10.00	8.00	6.00	7.00
H_b	2.643	1.435	1.441	1.024	1.679
E_c	0.868	0.623	0.693	0.571	0.863
Good's Coverage	0.971	0.762	0.376	0.631	0.753
S_{ACE}	22.95	28.898	55.062	34.085	21.12
S_{CHAO1}	22.21	25.64	38.913	26.395	20.49

a Phylotype richness, S, was calculated as the percentage of the total number of distinct ARDRA patterns to clones.
b Shannon-Weiner diversity index (Shannon, 1948)
c Evenness

3.2.6.1. Biogeochemical profiling between sites

Geographical distance has been proposed as a possible contributory factor behind microbial community divergence, however data supporting this theory is lacking (Green *et al.,* 2006). A Mantel test was conducted based on the geographical distances between Ethiopian Rift Valley lakes and the OTU overlap between the sites (see Fig.27). The proposed alternative hypothesis stated that a relationship exists between geographical distance and community similarity. As the computed p-value (p= 0.006) was determined to be lower than the significance level (alpha=0.05), the alternative hypothesis was accepted with an error of 0.61%. The high negative r value (-0.788) suggests a strong negative correlation between the geographical distance and community similarity. In other words, an increase in distance correlates with a decrease in community similarity. Fig.28 clearly illustrates the lack of a significant overlap between the Lake Arenguade (the northernmost lake) and Lake Abitjata communities, and conversely, the overlap increases between lakes closer together.



Mantel test	
r(AB)	-0.788
p-value	0.006
alpha	0.05

Figure 27. Mantel test r-distribution for matrixes A (geographical distance) and B (OTU overlap)

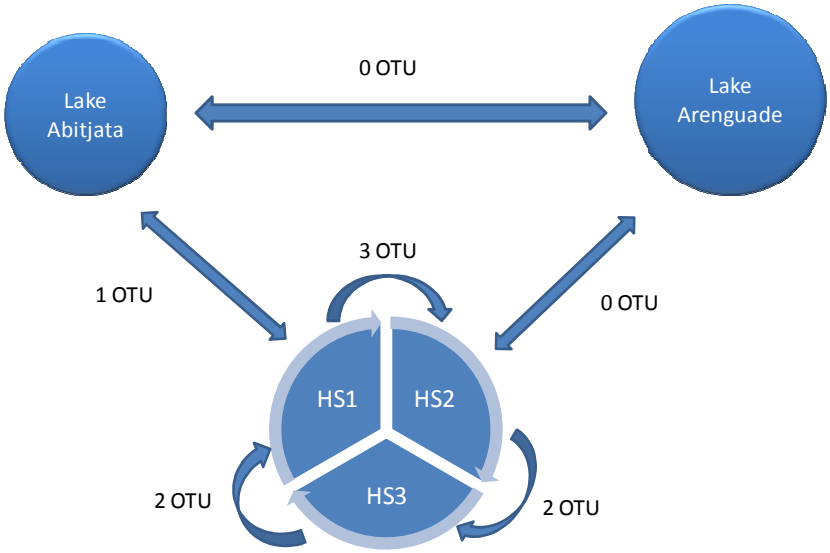


Figure 28. Operational taxanomic unit (OTU) overlap between sites. Each OTU represents a unique species from the eubacterial libraries present in both sites.

By combining geochemical/physico-chemical parameters such as conductivity and temperature with the biological composition of each site, it is possible to identify the factor(s) contributing the most towards variance between individual sites. The PCA plot in Fig.29 indicates the distribution of the sites based on these factors and shows the same clustering pattern observed on the dendograms (Fig.16, 19). The lake groups were differentiated by temperature and pH, which were inversely correlated in PC1, and by the conductivity shown in PC2. A Chi-square test of independence demonstrated that the composition of the microbial communities was dependent on the physico-chemical characteristics of the environment at a significance level of $p<0.05$. The data indicated that at high temperatures (HS1, HS2 and HS3), sequences related to the α-Proteobacteria and Firmicutes (mainly *Bacillus* species) were abundant. Lake Abitjata, as a temperate site with a higher conductivity and salt content, was dominated by actinobacteria and δ-Proteobacteria sequences, whereas Lake Arenguade (a less haloalkaline site) contained more *Bacteroidetes* and *Clostridia* type-sequences. These findings are in agreement with similar work done on Tunisian alkaline hot springs (with comparable physico-chemical parameters) and soda lakes in the Kenyan-Tanzanian Rift Valley (Sayeh, 2010; Grant *et al.,* 2004).

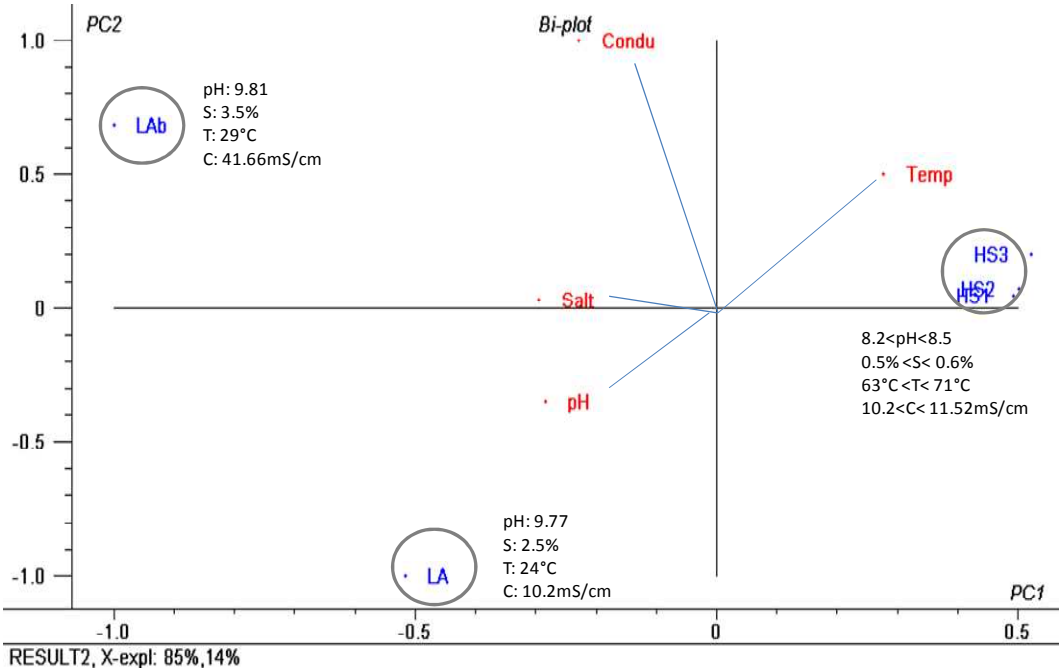


Figure 29. PCA plot of the physico-chemical parameters separating the different sampling sites.

3.2.6.2. Estimating 16S rRNA library coverage

Statistical inference is subject to sampling efficiency. Statistically, therefore it is necessary to estimate at which point a library is sufficiently sampled in order to evaluate the significance. With regards to eubacterial diversity, this is often an impossible task as no technique exists whereby the community in its entirety can be calculated. An estimation of the diversity can, however, be reached if a reference value is used. In this study, a combination of S_{chao1} and S_{ace} was used to determine if the population was successfully sampled. Both richness estimators make use of the occurrence of unique phylotypes, where S_{chao1} is used when a considerable number of rare phylotypes exist while S_{ace} is used in libraries with a larger amount of recurrent phylotypes. Good's Coverage (see Fig.30) is a third estimator which plots the observed phylotype over the predicted S value, which is S_{chao1} in this case. The reason for using S_{chao1} as opposed to S_{ACE} is that a large number of rare phylotypes were identified in all eubacterial libraries.

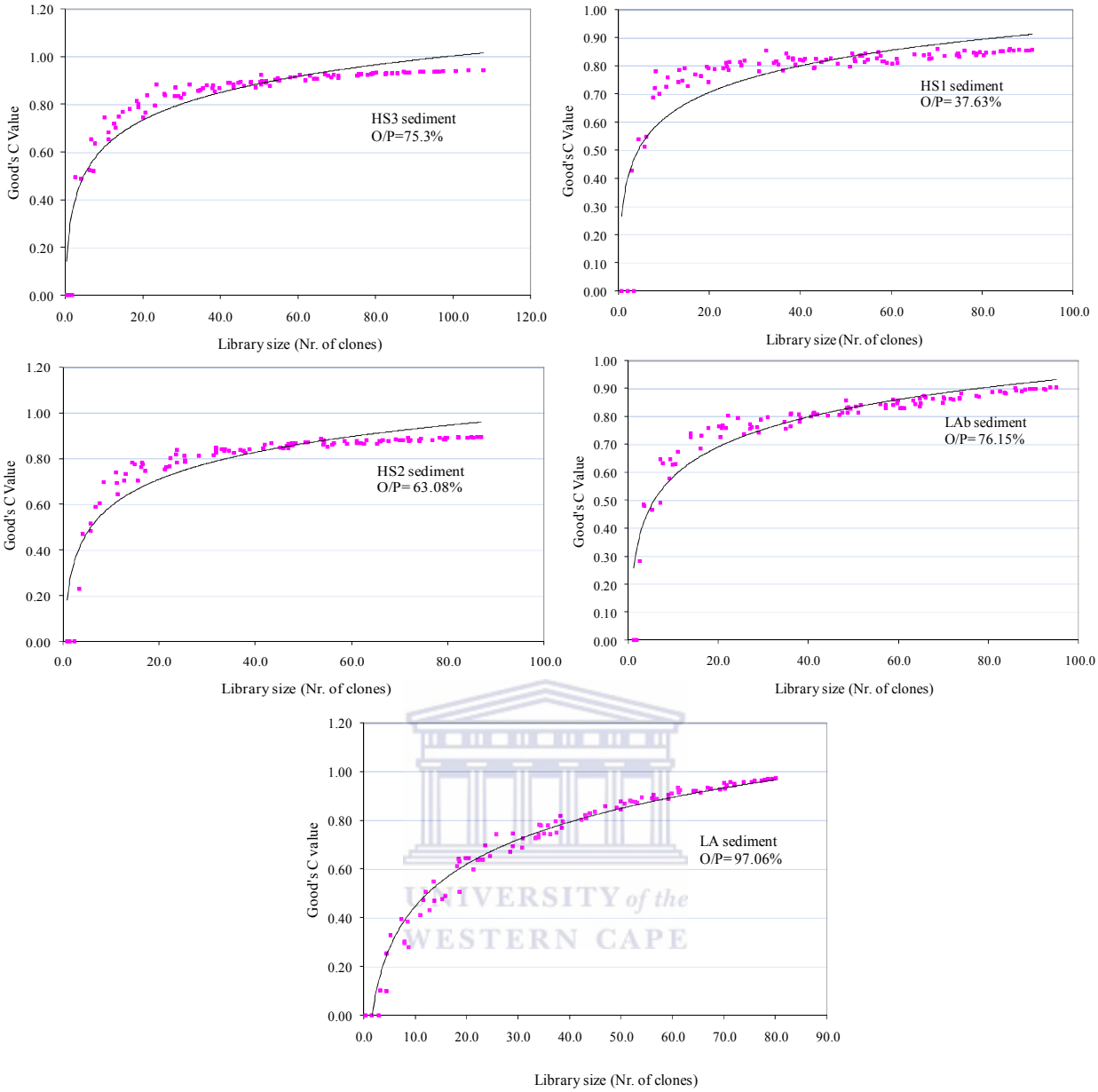


Figure 30. Clone library coverage based on Good's C estimator of the eubacterial clone libraries from the sediment of Lake Arenguade (LA), Lake Abitjata (LAb), Lake Shala hot springs 1, 2, 3 (HS1, 2, 3). O: observed number of phylotypes, P: predicted number of phylotypes.

The Good's C value for the Lake Arenguade eubacterial clone library (97.06%) indicated that it was the only site for which sampling almost reached a plateau. The Good's C value for both Lake Abitjata (76.15%) and Lake Shala hot spring 3 (75.3%) indicated that more than two thirds of these sites had been sampled. The only outlier was Lake Shala hot spring 1 which showed a low count (37.63%), even after screening an additional 50 clones. This could be attributed to either a higher diversity or a lack of sensitivity of the estimator due to recurrent phylotypes. The lower Good's C value for S_{ace} however suggests that a higher diversity may be present at this site.

3.2.7. Phylogenetic analysis of 16S rRNA sequence (OTUs)

All eubacterial library OTUs were grouped according to sequence similarity ('type') and categorized under specific phyla or families. As the overall composition was discussed in section 3.2.7 the predominant subpopulations will be explored in greater detail. Phylogenetic trees were constructed for all subpopulations and the subsequent clone clusters identified (see supplementary material). The actinobacterial subpopulation will be discussed separately in section.3.2.8.

3.2.7.1. Phylogenetic analysis of the 16S rRNA eubacterial library: Bacillus

The eubacterial library clone sequences that represented unique ribotypes were categorized according to the closest 16S rRNA gene relative identified from BLAST analysis. The largest proportion of the clones formed part of four clusters within the genus *Bacillus* (Fig.31). Most clones showed a high degree of similarity (96-99%) except for two; LAbC9 and HS3H1. These showed 93% and 91% similarity to the closest type strain, respectively. In all cases except for cluster 2, clones from Lake Abitjata branched separately from the hot springs suggesting a high degree of homology within the site.

Cluster 1 included both Lake Shala hot spring 1, 2 and Lake Abitjata clones, with the closest matches being three obligate alkaliphilic *Bacillus* species, *Bacillus agaradhaerens* and *Bacillus clarkii* (both soil isolates), and *Bacillus vedderi*. In addition, an uncultured clone from Lake Magadi (Kenyan Rift Valley soda lake) groups within the same cluster suggesting a close relationship between alkaliphilic bacilli. Interestingly, all three species require high sodium chloride concentrations for cell growth.

The second cluster contains two alkaliphilic species, one with a slightly lower tolerance for NaCl. *Bacillus hortii*, a species distantly associated with the above mentioned three type strains, grows optimally between pH 8-10 while tolerating high NaCl concentrations. *Bacillus mannanilyticus,* however, only tolerates up to 3% NaCl, but grows optimally at the same pH (Nogi *et al*., 2005). Both clones (HS1.E12 and LAb.C9) forming part of this cluster, aligned loosely with *B. hortii* (92 and 91%, respectively) whilst both clones showed a high similarity (97%) to *B. mannanilyticus*.

The third cluster of clones grouped with the alkaliphiles *Bacillus cohnii*, *Bacillus halmapalus* and *Bacillus horikoshii.* All three type strains belong to the *Bacillus* 16S rRNA group I. The fourth and last cluster includes *Bacillus akibai* and *Bacillus pallidus,* a *Bacillus*16S rRNA group V member. LAb.C9 was the only clone which showed a high similarity to a common soil bacterium, *Bacillus licheniformis,* and this clone originated from Lake Abitjata.

The predominance of *Bacillus* sequences is to be expected since a considerably large number of *Bacillus* species have been isolated from Rift Valley soda lakes along the Kenyan-Tanzanian border (Duckworth *et al*., 1996). Duckworth *et al.,* (1996) proposed that two different 16S rRNA gene groups colonise two separate

habitats within these soda lakes: littoral mud (fluctuating environment) and the lake water/sediment (less variable). They postulated that the *Bacillus* 16S rRNA gene group 6 were mostly found in littoral mud and dried soil, whereas 16S rRNA group 7 alkaliphilic species require high Na^+ ions and alkaline conditions to survive (such as found in the lake) (Table 13). Grant (2004) speculated that the predominance of *Bacillus* species could be due to the ability of some to enter a dormant stage that can survive drying in marginal areas (Grant, 2004). Interestingly, only a single clone (LAb.C5) from the current study showed any similarity to these two *Bacillus* 16S rRNA groups and most of the clone sequences from the current study associated with 16S rRNA group 1. Given the high proportion of *Bacillus* species identified in this study, further research should investigate the diversity of bacilli colonising the soda lakes.

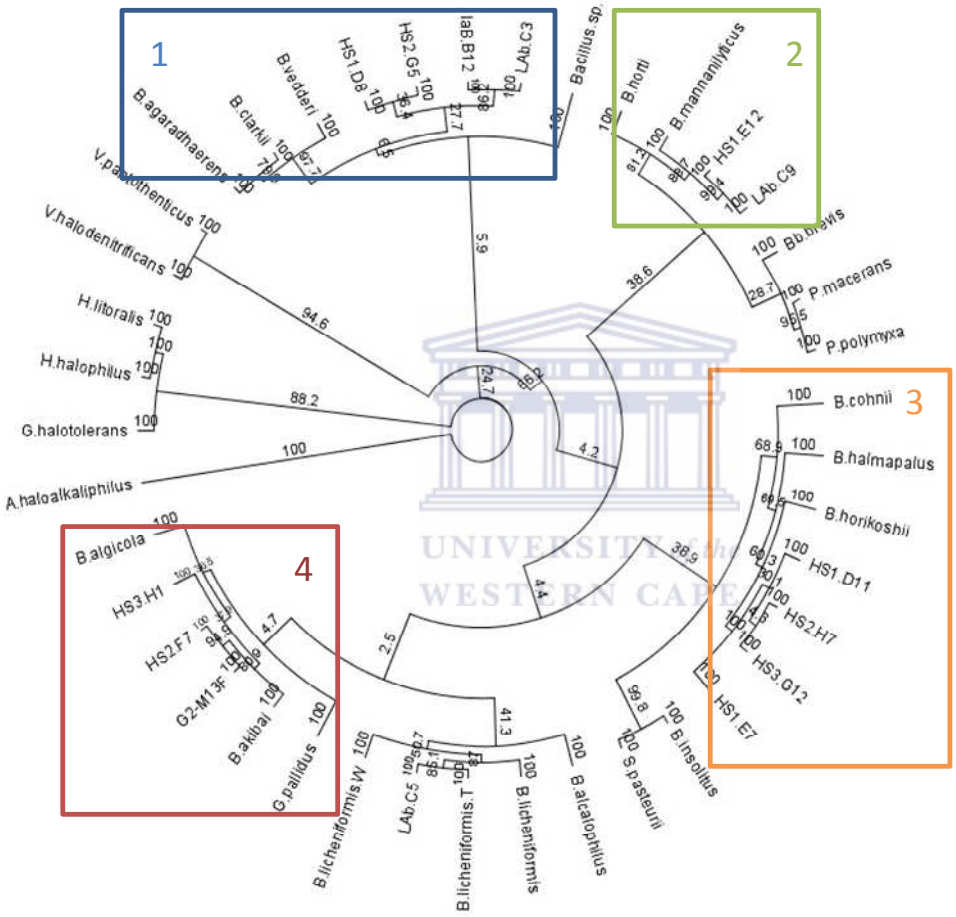


Figure 31. An unrooted phylogenetic tree of all the clone sequences related to *Bacillus* species and the most closely related type strains based on 16S rRNA gene homology. A HKY neighbour joining method with 1000 bootstrap permutations was used to construct the tree.

Table 12. Species used in phylogenetic analysis of clones with the sequence type *Bacillus*.

Species	Group	Source	pH Tolerance	T [°C]	NaCl
Bacillus cohnii	**1**	**Horse meadow Soil**	obligate alkaliphile	**10-47**	**5%**
Bacillus halmapalus	**1**	**Soil**	**pH_{opt} 8.0, pH 7**	**10-40**	**<5%**
Bacillus horikoshii	**1**	**Soil**	obligate alkaliphile	**10-45**	**17%**
Bacillus pallidus	**5**	**Waste water**	**8.0-8.5**	**30-70**	**n.k.**
Bacillus agaradhaerens	**u**	**Soil**	obligate alkaliphile	**10-45**	**16%**
Bacillus clarkii	**u**	**Soil**	obligate alkaliphile	**15-45**	**16%**
Bacillus horti	**u**	**Soil, Japan**	**alkalitolerant**	**15-40**	**10%**
Bacillus vedderi	**u**	**Bauxite waste**	obligate alkaliphile	**45-50**	**7.50%**
Bacillus alcalophilus	**6**	**Soil and faeces**	obligate alkaliphile	**10-40**	**8%**
Bacillus haloalcalophilus	**6**	**Brine/mud, Wadi natrun**	obligate alkaliphile	**15-40**	**25%**
Bacillus halodurans	**6**	**Soil**	obligate alkaliphile	**15-55**	**12%**
Bacillus halophilus	**u**	**Rotting wood, Japan**	**pH 6-8, pH_{opt} 7.0**	**n.k**	**30%**
Gracibacillus halotolerans	**n.k**	**Great salt Lake, Utah**	**5-10, pH_{opt} 7.5**	**6-50**	**20%**
Halobacillus halophilus	**n.k**	**Salt marsh soil**	**7.0-9.0**	**15-37**	**15%**
Halobacillus litoralis	**n.k**	**Sediment, Great Salt Lake ,Utah**	**6.0-9.5, pH_{opt} 7.5**	**10-43**	**25%**

n.k. Not known

3.2.7.2 Phylogenetic analysis of the 16S rRNA eubacterial library: Clostridia

The majority of the clones that showed homology to Clostridial type strains had relatively low similarity scores between 92-95%, suggesting considerable novelty. Only two clones showed a higher similarity, LA.B6 and HS1.C2 (indicated as cluster 3 in Supplementary Fig.1), with a 99% homology to *Clostridium quinii* and *Clostridium corinoforum,* respectively. Both of these type strains fall within the *Clostridia* 16S rRNA Cluster I, which is the equivalent of the *Bacillus* 16S rRNA group I (Collins *et al*., 1994). *Clostridium quinii* was originally isolated from waste water at pH 10.0, but grows optimally at pH of 7.5, between 40-45°C (Svensson *et al*., 1992). Additionally, clone LA.B4 was distantly related to a separate branch which included *Clostridium mesophilum* (90%) and *Clostridium hveragerdense* (94%). The first is an obligate anaerobic mesophile that was originally isolated from an upflow anaerobic sludge basket (UASB) reactor, whilst the second is an anaerobic thermophilic bacterium isolated from a hot spring in Iceland.

A distinct cluster (cluster 1) contained clones from Lake Arenguade, Abitjata and Shala hot spring 1 and was separated from all other families. This phenomenon has been reported in similar fingerprinting studies on

Lake Magadi, where clones were most closely related to uncharacterized cultured isolates from the same lake (Baumgarte, 2003). Although uncultured bacteria related to *Thermoaerobacter*, *Thermoaerobacterium* and *Halanaerobiales* species have been found in the past (Kenyan soda lakes), none were detected in the current study (Mesbah & Wiegel, 2008; Baumgarte, 2003). The fragmented nature of the *Clostridia* hierarchical cluster complicates the allocation of uncultured clones to a specific family/genus. This is the reason behind the proposed reclassification of the Clostridial taxa (Collins *et al*., 1994).

3.2.7.3. Phylogenetic analysis of the 16S rRNA eubacterial library: Alpha-Proteobacteria

Clone sequences associating with the α-subdivision of the Proteobacteria phylum, formed two distinct clusters when compared to other type strains (Fig.32). Cluster 1 was a combination of clones with relatively low (<90% similarity) similarities to alkaliphilic *Sphingobium* species. The cluster consisted of clones from all three hot springs and Lake Abitjata. *Sphingobium* species have been isolated from activated sludge and fly ash, as well as clinical samples and pentachlorophenol contaminated soil (Takeuchi *et al*., 2001). Cluster 5, however, contained nine clones which all showed a high degree (98-99%) of similarity to *Brucella melietensis* and *Ochrobactrum antriopii*. *B. melietensis* is a Gram-negative, aerobic pathogen originally isolated from goats (Pozo *et al*., 2002). Clones related to *B. melietensis* were common in all samples, with the exception of Lake Arenguade. Although *Brucella* species are considered animal pathogens and these clones could be due to contamination from the local community, studies have shown they are present in soil and some strains can grow without an animal host (Scholz *et al*., 2008). As the four sites (HS1, HS2, HS3 and LAb) are commonly used as an animal waterhole, the presence of an animal pathogen is not surprising. However, Lake Arenguade is protected by the local government and is therefore not accessible to farm animals.

The last two clusters (4 & 5) are composed of sequences from Lake Shala hot spring 1 and 2. HS1.E3 and HS1.E10 are both distantly (88%) related to *Ochrobacterium grigonense* while HS2.G4 showed only 81% similarity to *Paenochrobactrum gallinarii.* The low similarity to any type strains makes it impossible to accurately predict to which genus or class of bacteria these clones belong, but rather suggests that several novel organisms might be present in the hot springs.

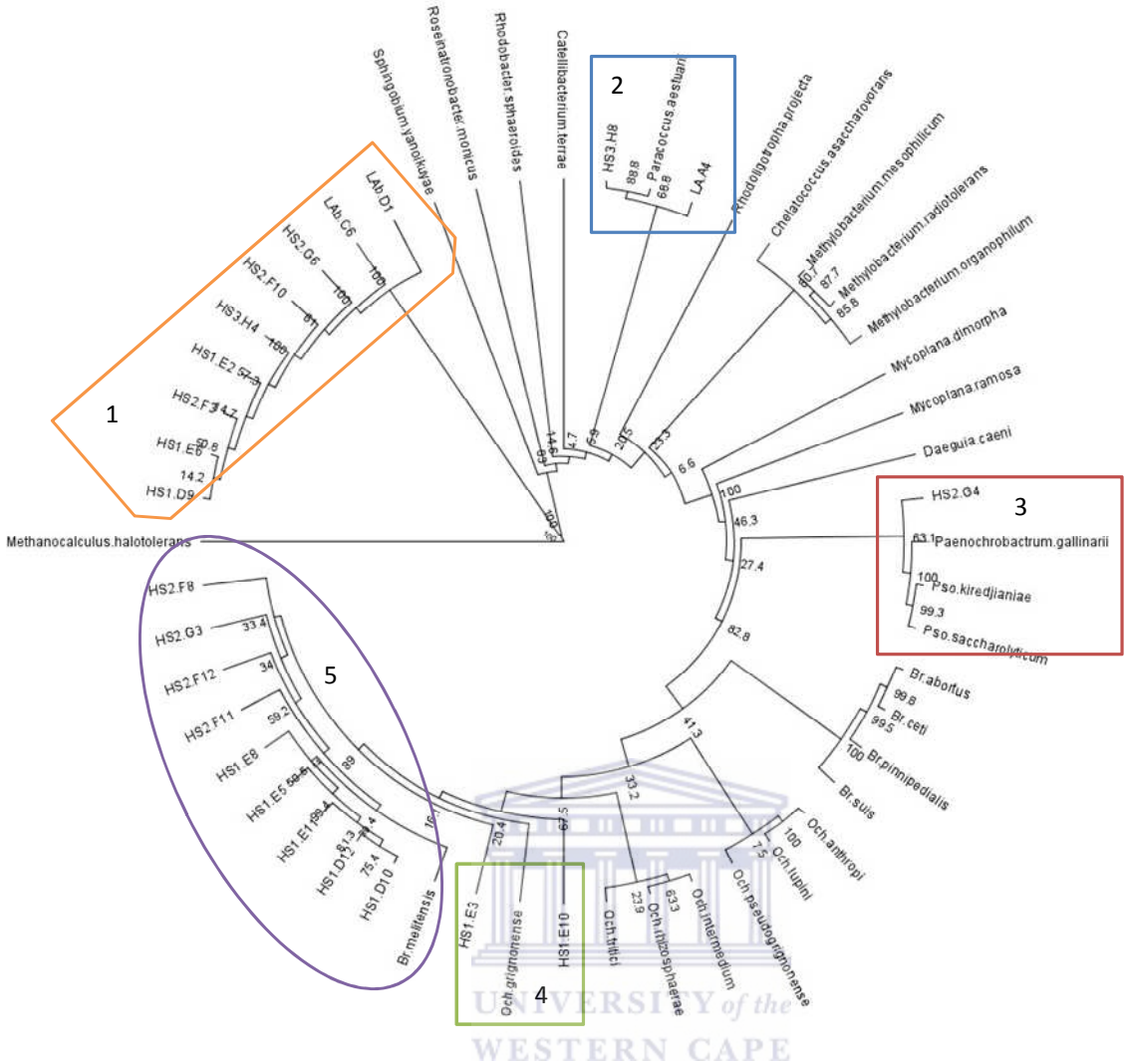


Figure 32. Neighbour Joining tree of the Alpha-Proteobacterial sequence types.

3.2.7.4. Phylogenetic analysis of the 16S rRNA eubacterial library: Gamma-Proteobacteria

All clone sequences closely related to γ-Proteobacteria could be divided into six clusters (see Supplementary Fig.2). The largest group (Cluster 6) showed a low similarity (90%) to *Thiohalorhabdus denitrificans,* a halophilic sulphur oxidizing bacterium isolated from a salt lake in Russia. The second largest cluster (Cluster 2) contained two Lake Arenguade clones (LA.B7 and LA.A11) and one Lake Abitjata clone (LAb.D6) that closely associated with three *Pseudomonas* species (95-96%). All of them fall into the *Pseuomonas*16S rRNA Group I subdivision. This is not surprising since pseudomonads have been isolated from most soda lake environments, with several isolates being partially characterized (Jones *et al*., 1998; Duckworth *et al*., 1996). The closest homolog to B7, A11 and D6 is *Pseudomonas acephalitica* which is an organophosphate degrading bacterium. Clusters 3, 4 and 5 contain clones with a 90% sequence similarity to *Idiomarina* species. Although only a few *Idiomarina* species have been documented in soda lakes, they were present in both temperate lakes (Lake Arenguade and Abitjata). *Idiomarina* type species, including *Idiomarina seosinensis,* have been mainly isolated from deep sea water trenches (Brettar, 2003). All partially

characterized species included in cluster 3 were isolated from a soda lake in the Hungarian lowlands and are defined as alkaliphilic. Cluster 1 includes a single clone which showed a high similarity to *Thioalkalimicrobium sibiricum*, an alkaliphilic thiosulfate reducer that was originally isolated from the Buriatia soda lake in Siberia (Sorokin *et al*., 2001).

3.2.7.5. Phylogenetic analysis of the 16S rRNA eubacterial library: Beta- & Delta-Proteobacteria

Within the five libraries only four clones distributed between Lake Arenguade, Lake Abitjata and Lake Shala hot spring 1 were related to β- and δ-Proteobacterial sequences. Two hot spring clones showed high (99%) similarity to *Variovorax paradoxus*. The genus *Variovorax* contains four species and while *V. paradoxus* has been found in atrizine contaminated soil and cotton, strains have yet to be isolated from soda lake environments. (De Souza, *et al.,* 1998). The two other clones (LA and LAb.C4) grouped with *Pelobacter propionicus*, an obligate anaerobic non-spore forming δ-Proteobacterium known for its utilisation of 2, 3-butanediol as a sole carbon source.

3.2.7.6. Phylogenetic analysis of the 16S rRNA eubacterial library: other uncultured bacteria

Several sequences were only loosely related (85-95%) to any genus and could only be categorized by family affiliations. A total of 15.2% of clones fall within this bracket, of which the nearest type matches are *Fusobacteria* species and members of the phylum *Bacteroidetes*. Three clones formed associations with several type strains (89-92%) and were therefore excluded from the above classifications. Clone LAb.C10 showed an 89% similarity to an uncharacterized Actinobacterium (SCGC) isolate believed to play a role in carbon fixation in deep sea trenches. The other two clones both originated from Lake Arenguade. LA.B10 showed a 91% similarity to *Ilyobacter insuetus* as well as several uncultured clones derived from lake sediment. LA.B1 showed a 92% similarity to *Fusobacterium varium,* and 97% homology to a potentially endophytic *Fusobacteria* sp.

3.2.8. Analysis of Actinobacterial subpopulations

The five libraries contained a relative low diversity within each sampling site ranging from a diversity index of 0.69 to 1.74 (see Table 14). The Simpson's index (D) (Buckley *et al.*, 1998) and Shannon evenness (Shannon, 1948) estimator both indicate a very heterogeneous population of actinobacteria at all sample sites.

Table 13. The evenness, diversity and richness estimators for the five actinobacterial 16S rRNA libraries

	LA	LAB	HS1	HS2	HS3
Shannon-Wiener Diversity Index	0.693	1.735	1.143	1.494	1.099
Species Richness (S)	2.00	6.00	4.00	5.00	3.00
Simpson Diversity Index					
D:	0.50	0.185	0.361	0.250	0.333
1-D:	0.500	0.815	0.639	0.750	0.667
1/D:	2.000	5.400	2.769	4.000	3.000
Evenness	1.000	0.968	0.825	0.928	1.000

a Phylotype richness, S, was calculated as the percentage of the total number of distinct ARDRA patterns to clones.
b Shannon-Weiner diversity index
c Evenness

The unique composition of each site is evident in Fig.33. Lake Abitjata showed the largest diversity (1.74) with OTUs covering the families *Solirubrobacteraceae, Rubrobacteraceae, Acidothermaceae, Streptomycetaceae* and *Micrococcaceae*. The presence of sequences related to the family *Streptomycetaceae* distinguished it from the other sites (Fig.34). This is perhaps unsurprising as of all the lakes studied, Lake Abitjata has physicochemical characteristics closest to known actinobacterial growth requirements. It can therefore be argued that it should be colonised by a more diverse population. The present study found that the hot springs were characterized by a large proportion of *Micrococcaceae* and *Mycobacteriaceae* OTU sequences. Limited research has been conducted on the distribution of thermophilic and thermotolerant actinobacteria. Previous studies have investigated the diversity of extremophilic actinomycetes in Himalayan hot spring soil, lake and river sediment. In all three studies thermotolerant and thermophilic actinobacteria belonging mostly to the genera *Streptomyces* and *Micromonospora* were identified, with smaller numbers of *Saccharopolyspora, Microtetraspora*, *Actinomadura* and *Microbispora* species. Thermophilic *Mycobacterium* species are less common and lacking in hot spring diversity studies, however *Mycobacteria* have been isolated from fresh and marine water, tap water, swimming pools and bottled water. On a smaller scale Fig.35 shows the distribution of sites on PC2 and PC3 which clearly separates the hot springs based on three families. Hot spring 1 associated with *Rubrobacteriaceae*, while hot spring 2 and 3 separated based on *Frankiaceae* and *Microbacteriaceae* sequences, respectively.

The difference in composition at each site and the low similarity of sequences to cultured strains suggests that a unique, uncultured actinobacterial population exists in all of the sites.

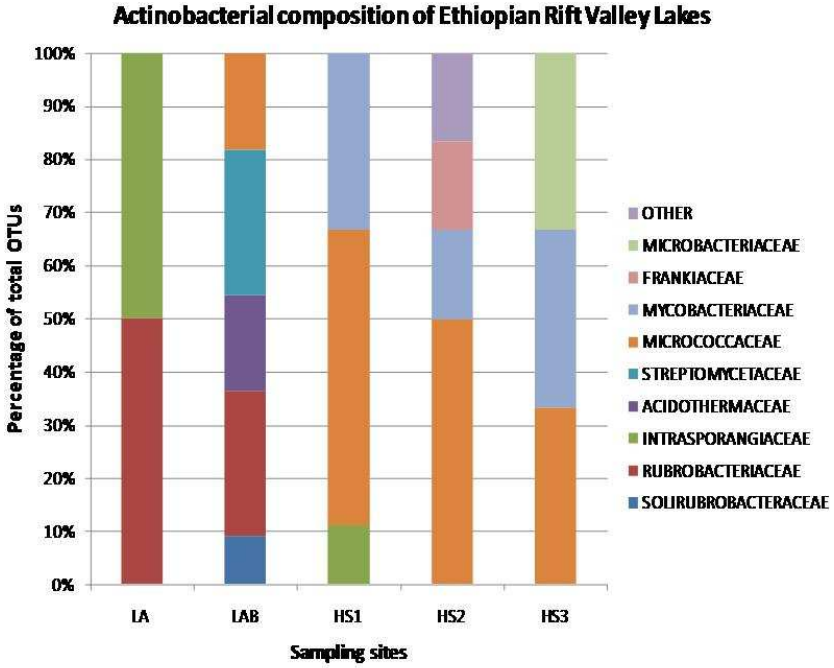


Figure 33. Actinobacterial composition of each sampling site based on OTU percentages.

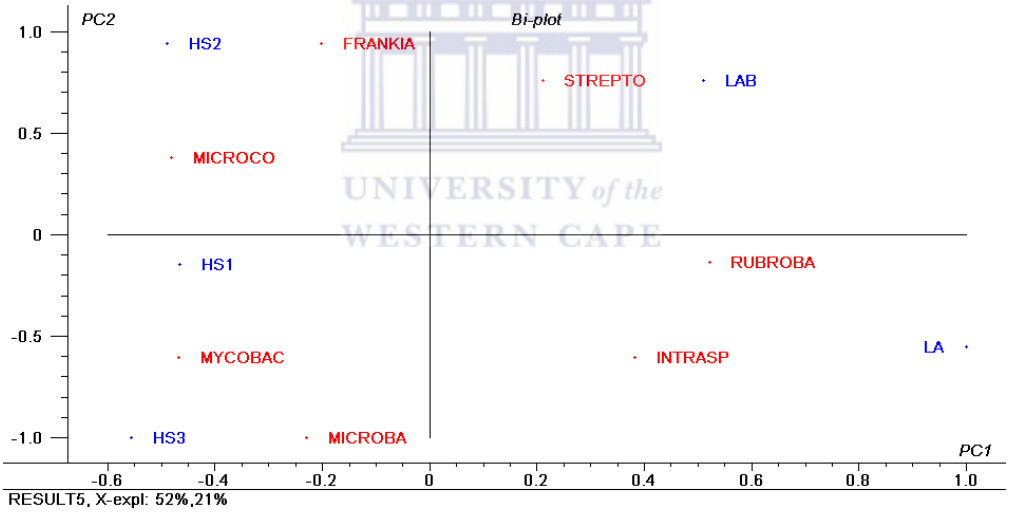


Figure 34. Principle component analysis of the actinobacterial population of each sampling site.

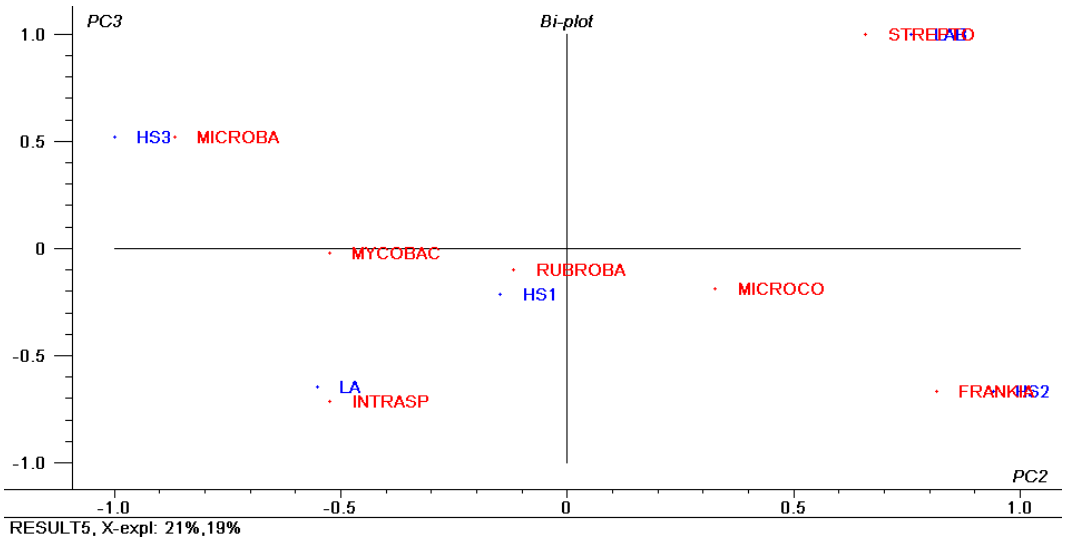


Figure 35. Principle component analysis of the actinobacterial population of each sampling site.

3.2.8.1. Library coverage

The unique phylotypes obtained from each site were evaluated based on the reoccurrence within the 100 clones screened from each library (Fig.36). Based on Good's coverage values and S_{chao1} estimator values all libraries showed high coverage (86-99%). One OTU in particular accounted for 33% of all clones screened, which might have influenced the estimator value. This clone was isolated from Lake Shala hot spring 1 and its closest match was an *Arthrobacter* species (91%). The low similarity however hinders the identification of the clone to species level.

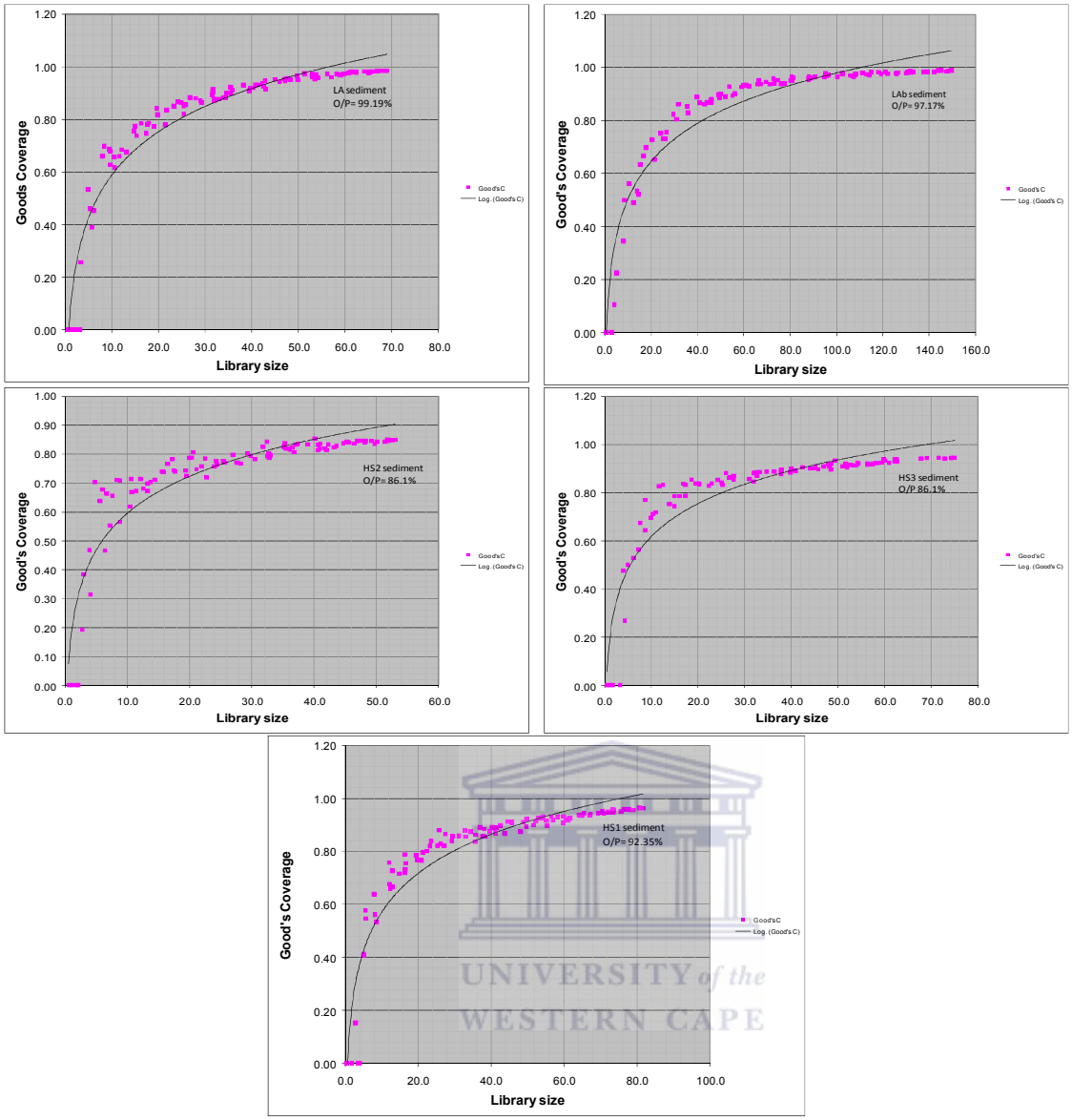


Figure 36. Clone library coverage based on Good's C estimator of the actinobacterial clone libraries from the sediment of Lake Arenguade (LA), Lake Abitjata (LAb), Lake Shala hot springs 1, 2, 3 (HS1, 2, 3). O: observed number of phylotypes, P: predicted number of phylotypes.

3.2.9. Actinobacterial clone libraries of the Ethiopian Rift Valley

Actinobacteria are said to be ubiquitous in all terrestrial and aquatic environments, however limited research has focused on alkaliphilic or thermophilic actinobacteria. The clone sequences retrieved from the actinobacterial library contained clones that mapped to a wide variety of genera (see Fig.37). A small proportion of the sequences showed high homology to other Gram-positive bacteria such as *Bacillus* and *γ-Proteobacteria* as well as a single *Planctomycete* species. Previous studies have reported a similar finding (Jones *et al*., 1998). A possible explanation for the presence of non-actinobacterial amplicons being present in these libraries is that the actinobacterial diversity is low and the more dominant non-target templates are preferentially amplified. Alternatively, primers may lack specificity and unintentionally amplify distantly related organisms. Clonal sequences showed a relatively low to moderate homology to type strains (86-96%)

with a few exceptions. Phylotypes with a 100% identity to *Mycobacterium massiliensis* were detected in all of the Lake Shala hot springs. *M. massiliensis* is a human pathogen, however, a clone sequence with a high similarity to the same species has also been found in a halothermophilic deep surface petroleum reservoir (H. Li *et al*. 2006). It is not possible to determine whether the presence of this organism in the soda lakes is a reflection of human contamination at the site or whether it is part of the microbial community, but considering the temperatures it's likely just a signal and not a metabolizing organism. Two other clones, LAb.A11 and HS1.H9, were 100% homologous to the *Streptomyces pseudovenezuelae* and *Knoellia sinensis* type strains, respectively. *Knoellia sinensis* is an aerobic non-spore forming member of the *Micrococcineae* suborder that was originally isolated from a cave in China (Groth *et al*., 2002). *Streptomyces pseudovenezuelae* is a highly alkaliphilic, halotolerant strain that is able to grow up to pH 12.0 (Shirling & Gottlieb, 1968).

Several clones distributed throughout the hot springs of Lake Shala showed high similarity to the genus *Arthrobacter* and several clones shared 100% similarity to *Arthrobacter globiformis*. This aerobic, non-motile rod has been isolated from Burdwan soil and MuusShadi earth crust (Inner Mongolia). The type strain has a pH and temperature range of 5-8 and 5-37°C, respectively (Conn, 1928). Clones showing a high similarity to *Arthrobacter* species were present in both Lake Shala hot spring 1 and 2, but not detected in hot spring 3. Other clones (distributed between Lake Shala Hot spring 2 and 3) showed 98-99% similarity to two uncharacterized isolates, *Arthrobacter* sp. CY2W2 and TP2MF. The first was isolated from the rhizosphere of *Eupatorium adenophorum* (i.e. snakeroot) and the second was isolated from a membrane bioreactor. One clone, HS2.8C, aligned with an *Arthrobacter pascens* strain isolated from the rhizosphere of *Phyllostachys edulis.*

Five clones derived from hot spring 1 showed low similarity (91%) to an *Arthrobacter* sp. PS29 with the closest type strain being *Arthrobacter agilis* (89% similarity). Strain PS29 was isolated from Phoenix Park soil (Ireland), whilst strains of *Arthrobacter agilis* have been isolated from waste water, soil and concrete (Eppard *et al*., 1996). A similar strain, *Arthrobacter crystallopoietes*, which was also isolated from a concrete structure, was the closest type strain to a clonal sequence from Lake Abitjata LAb.B1. The remaining clones showed relatively low similarities to the closest type strain and will be discussed below according to their respective subclass. A phylogenetic tree (Fig.37) was constructed with the sequences showing the highest similarity values, although the length (~300bp) of the sequences were insufficient to draw definitive high resolution peripheral branches, the family groupings are accurate and will be discussed further.

3.2.9.1 Phylogenetic analysis of the 16S rRNA eubacterial library: Acidothermaceae

Three clones showed high similarity to two *Acidothermaceae* species. Clones LAb.G17 and LAb.G27 were distantly related to *Ferrimicrobium* BGR116. This specific isolate is an iron-oxidizing bacteria found in sulphuric mine waste. The closest type strain was *Acidothermus cellulolyticus*, a hot spring isolate with

cellulolytic capacity that showed 94% and 93% similarity to LAb.G14 and LAb.G27, respectively. The third clone, LAb.G19, showed high similarity (97%) to *Ilumatobacter fluminis,* an alkaliphilic actinobacterium isolated from a marine estuary. *I. fluminis* grows over a pH range of 7.0 to 11.0 and between 26-31°C (Matsumoto, 2009).

3.2.9.2 Phylogenetic analysis of the 16S rRNA eubacterial library: Rubrobacteraceae and Solirubrobacteraceae

Two clones were distantly (89%) related to the thermophile *Rubrobacter taiwanensis*, which was isolated from a hot spring (Chen *et al*., 2004). Both clones originated from the Lake Abitjata library and showed a higher similarity to isolates obtained from soil during a phylogenetic study on mini-colony forming soil bacteria belonging to the *Rubrobacteraceae* family.

3.2.9.3. Phylogenetic analysis of the 16S rRNA eubacterial library: Streptosporangineae

One clone, Lab.G29, showed some relationship (93%) to an uncharacterized thermophilic *Thermomonospora* species isolated from hot environments in Yunnan Province, China. Unfortunately, the strain remains unpublished and the nearest cultured organism to clone LAB.D7 showed 87% similarity. The nearest (97%) uncultured bacterium sequences are from a library of a geothermal site in Yellow Stone National Park.

3.2.9.4. Phylogenetic analysis of the 16S rRNA eubacterial library:Kineosporiaceae

A single clone (LA.G0) from Lake Arenguade showed 89% sequence similarity to *Kineosporia rhizophila*, which was originally isolated from galingale root (*Cyperus nzicroiria*). This mesophilic strain grows optimally at 28°C, pH 7 (Kudo *et al*., 1998). Similarly, clone G15 was 98% similar to several uncultured clone sequences derived from river sediment and wetland soil.

3.2.9.5. Phylogenetic analysis of the 16S rRNA eubacterial library: Frankiaceae

Clone HS2.D4 from Lake Shala hot spring 2 had a low similarity to an uncharacterized *Frankia* isolate (90%), which has been identified as a microsymbiote of a *Ceanothus* species only located in North America. Clone HS2.D4 was also distantly (82%) related to sequences belonging to the genus *Mycobacterium*.

3.2.9.6. Phylogenetic analysis of the 16S rRNA eubacterial library: Microbacteriaceae

Although hot spring 3 showed the least variation with regards to phylotype richness, it did contain one clonal sequence (clone HS3.D5) with a relatively high (96%) similarity to *Microbacterium* sp. YRR08. an endophytic bacterium with antimicrobial activity against plant pathogens.

3.2.9.7. Phylogenetic analysis of the 16S rRNA eubacterial library: Non-actinobacterial sequence types within actinomycete-biased libraries

Non-actinobacterial OTU sequences that were detected in the actinobacterial libraries included sequences related *to Bacillus, Clostridia, Bacteroidetes, Planctomycetes* and *Fusobacteria*. All non-actinobacterial clone sequences, with the exception of three, showed low similarity values in the order of 84-92%. The four OTUs with high similarity values (98%) originated from Lake Abitjata and Lake Shala hot spring 1, 2 and 3, with the closest type strains being *Bacillus. halmapalus, Bacillus licheniformis, Bacillus alkalogaya* and *Bacillus decolourationis.*

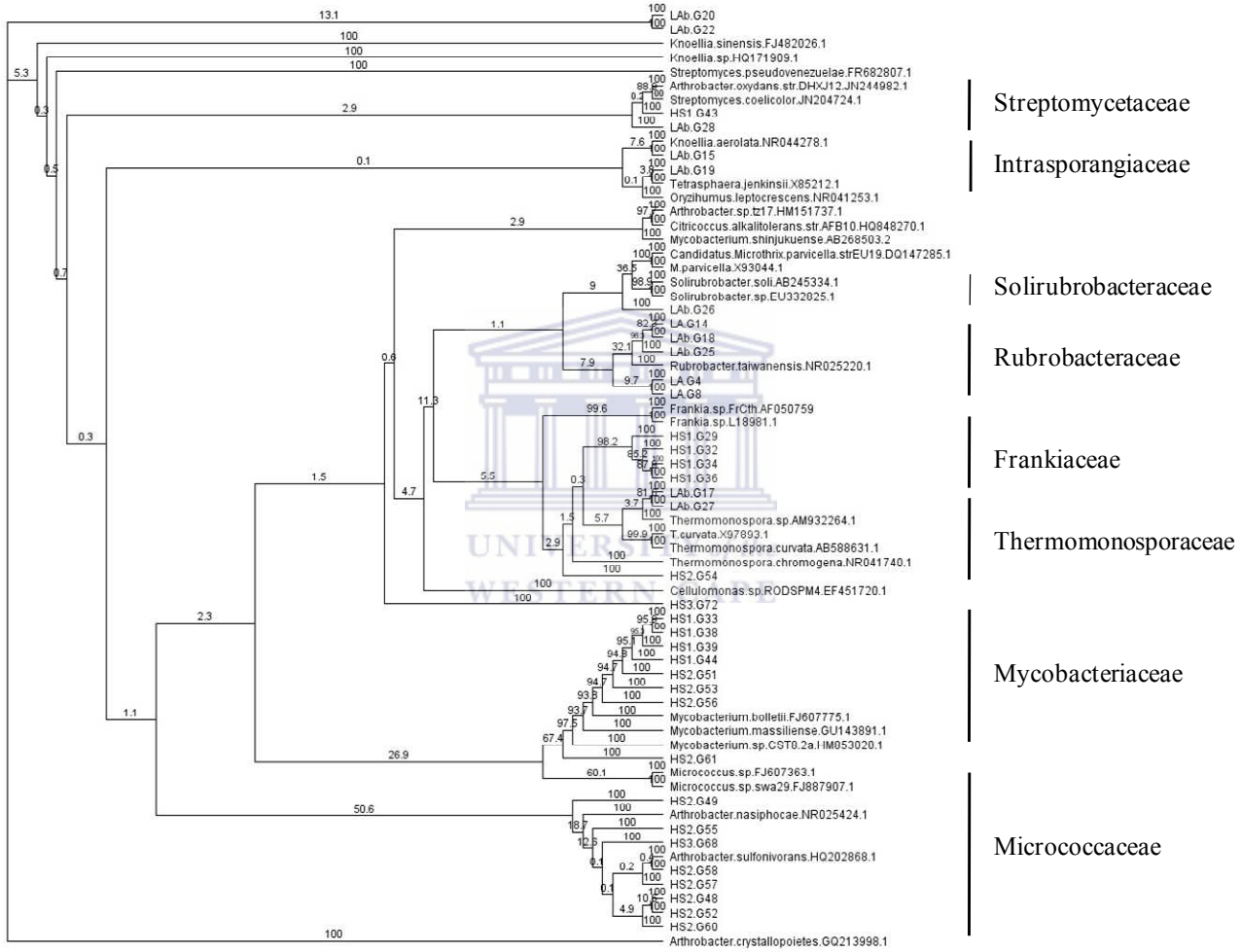


Figure 37. A Neighbour-joining tree of all the short actinobacterial clone sequence. Values on branches indicate Bootstrap values as derived by 1000 permutations.

3.2.10. Subpopulation Diversity: A study between sites

The Ethiopian Rift Valley lakes contained representatives of several actinobacteria subclasses, however relatively few species were found at each individual site. If all the actinobacterial libraries are considered, only 54% of the unique phylotypes showed similarity to actinobacteria. Of these the highest percentage, 22.086%, was from Lake Abitjata, whilst a further 29.43% was distributed between the Hot spring at Lake

Shala (HS1:7.34%, HS2:12.26%, HS3: 7.34%) and a mere 4.86% was from Lake Arenguade. The overlapping phylotypes were unique in that neither Lake Abitjata nor Lake Arenguade shared a single OTU overlap with the Hot springs, whilst the Hot springs shared a single *Mycobacterium* species (see Fig.38). The Lake Shala hot springs were mostly dominated by *Mycobacterium* species, whilst sequences from the Lake Arenguade libraries were closely associated to *Arthrobacter* species. The amplification of non-actinobacterial sequences is an inherent flaw associated with class-specific primers. Due to the vast differences in G+C content of actinobacterial 16S rRNA gene sequences the primers will inherently bias towards a specific families and due to the high similarity between the actinobacterial primers and universal DGGE primers, the detection of non-actinobacterial sequences had to be taken into account.

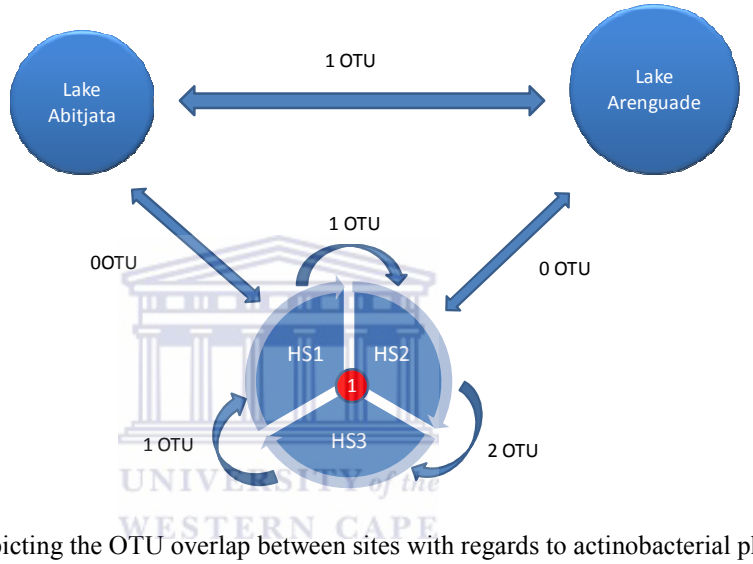


Figure 38. A diagram depicting the OTU overlap between sites with regards to actinobacterial phylotypes.

3.3 Conclusion

The analysis of microbial diversity by means of culture-independent approaches was developed to overcome the limitations of classical culturing. It has provided the means by which extreme environments can be probed in order to investigate the underlying ecological networks. The Ethiopian Rift Valley forms part of these extreme environments, however, the high primary productivity and high diversity complicates the targeting of minor subpopulations. The current study addressed this hurdle by combining two PCR-based approaches, DGGE and clone libraries, to study both the larger community and one of its subpopulations.

DGGE profiles of the five Ethiopian lakes provided a clear pattern of clustering where all five sites separated into three distinct groups (representing the three lakes). The analysis of the eubacterial 16S rRNA gene clone libraries revealed the detection of class-specific clustering, predominantly, with the taxa of Proteobacteria and Firmicutes. Site-dependent variation in the suborders within the phylum Proteobacteria was observed with Lake Abitjata being dominated by δ-Proteobacteria, the hot springs by α- and β-Proteobacteria and Lake

Arenguade by γ-Proteobacteria. The role of Proteobacteria in the anaerobic sulphur reducing and oxidising cycles suggest that these may play an important role within the nutrient cycle of these soda lakes. All lakes contained a large proportion of Firmicutes, mostly *Bacillus* and *Clostridia* related sequences. The reason behind the predominance of *Bacillus* sequences are found in their ability to survive drought in spore form and therefore exist in the intertidal zone, where most of the current studies samples were taken. The small deviations between sites were due to various subpopulations such as *Planctomycetes, Bacteroidetes* and *Fusobacteria* whose numbers increased in the mesophilic sites (Lake Arenguade and Lake Abitjata). All of the sequence-based data therefore suggested that a decrease in temperature, increase in pH and increase in conductivity could be coupled to an increase in community diversity.

The analysis of all five ERV eubacterial and actinobacterial 16S rRNA gene libraries revealed that the actinobacterial subpopulation could not be detected within the limited size of the eubacterial clone libraries. Nevertheless, upon use of actinobacterial-specific primers a diverse subpopulation of actinobacteria was detected in the same samples. Approximately 48% of all actinobacterial OTU sequences showed more than 97% similarity to the closest type strain, while 87.23% could be linked to type species with a similarity of 90% or higher. However, 12.77% of all actinobacterial clone sequences showed a lower than 90% similarity to any known cultured species. This 12.77% was divided between Lake Arenguade and Lake Abitjata, suggesting culturing aimed at these two sites might provide more information regarding the diversity. Each site was dominated by a different family including: *Frankiaceae* (HS2), *Intrasporangiaceae* (LA), *Microbacteriaceae* (HS3), *Rubrobacteraceae* (HS1) and *Streptomycetaceae* (LAb). The large diversity between sites could not be linked to any geochemical property measured in the current study. The large microbial diversity in both Lake Abitjata and Lake Arenguade might be attributed to the constant change in water constituents over the past decade. The direct (LAb) and inverse (LA) correlation of conductivity with both suggest a link between the diversity and water composition, however further studies need to be conducted to establish the impact that the water constituents (not included in the current study) have on microbial diversity.

The three Lake Shala hot spring sites had limited 16S rRNA gene diversity (both eubacterial and actinobacterial), which is to be expected due to the extremely high sediment temperatures (>90°C). Temperatures of this magnitude rarely support bacterial life and such sites would be expected to be dominated by Archaeal phylotypes. However, due to the introduction of lake water a temperature gradient forms in the springs which may potentially allow some bacterial species to survive in this extreme environment. A possible reason behind the composition of the eubacterial communities found in the hot spring fissures is the influence of high silica (Si) concentrations. Several studies have shown that silica adsorption to Gram-positive bacterial cell walls cause the precipitation of microscopic structures (microorganisms encased in silica). When protection mechanisms are in place this process may not necessarily result in cell death and poses a unique method by which selection (possibly against Gram-

positive actinobacteria) may occur.

The resolution of DGGE and clone libraries form part of the hierarchical structure used to analyse the various levels of the community composition. In the current study the lower resolution of the DGGE approach was used to distinguish between sites. The detection limit of DGGE profiles increased with the lower diversity communities. This is evident in the increase in similarity of OTUs detected when considering the actinobacterial diversity. DGGE OTU numbers were in the same order of magnitude as the number of unique phylotypes obtained from the clone library analysis (ARDRA). The small deviation seen between these two techniques can be attributed to a variety of factors, including: differences in 16S rrn operon sequence and number (i.e. multiple bands where there should be one), primer specificity (i.e. faint or no bands), G+C content of species (i.e. one band representing more than one species) or staining efficiency/resolution (i.e. a stain does not allow detection of faint bands on gel). However, the resolution of both techniques gives an idea of the species richness within each site.

A comparison of the short ~300bp sequence amplified with the actinobacterial primers and the full length sequences obtained for the cultured strains revealed that the V3 region of the 16S rRNA gene could only distinguish between sequence at the genus level. In other words, although the sequences may represent different species, the short sequence can only identify the genus to which the phylotype might belong. In order to refine the resolution of the clone libraries a larger region (V3-V6) needs to be used or a combination of the current short sequence with a functional gene. A further complication was that several non-actinobacterial sequences amplified with the actinobacterial primers. In all of the actinobacteria-biased libraries a trend was apparent with regards to the percentage non-actinobacterial sequence types. Where actinobacterial diversity was high, low amounts of non-actinomycete bacteria were detected and visa versa.

The majority of sequences showed a slight similarity to alkaliphilic organisms which suggest that within the Ethiopian Rift Valley there are vast populations of unique species that are yet to be discovered. The large similarities between sequences derived from the current sites and those isolates from Lake Magadi and Bogoria in Kenya, suggest that a fundamental ecological structure may be common between these soda lakes.

CHAPTER 4 CULTIVATION-DEPENDENT ANALYSIS OF THE ACTINOBACTERIA AND THE CHARACTERIZATION OF TWO NOVEL *STREPTOMYCES* SPECIES

4.1 Isolation of strains from lake sediment

The selective isolation of actinobacteria involves the formulation of media containing all the essential nutrients for actinomycete growth, while (where possible) excluding nutrients essential for growth of other microbes. While a large number of media have been developed for the isolation of actinobacteria from terrestrial habitats, several complications may arise during the course of culturing. Four parameters in particular may influence the culturability of actinobacteria from environmental samples. These include: storage conditions of the sample, the physiological state of the organism (cells/spores), the organisms ability to reproduce on agar and (finally) the organism's relative growth rate (Oliver, 2005; Janssen *et al*., 2002).

Storage of samples is done with the environment in mind, therefore *psychrophilic* sites require storage at -20°C, while soil from mesophilic sites can be stored at 4°C. Ideally, samples should be processed within 24-48 hours of collection, unless lengthy transportation prohibits this. The Ethiopian Rift Valley samples used in this study needed to be transported and stored at -80°C for several months before analysis could be done. It must be noted that a temperature drop from as high as 92°C (Lake Shala hot spring 1) to -80°C could potentially result in the loss of viability, while some spore-forming species could survive. Another important consideration is that prolonged storage at -80°C causes cell fragmentation to occur more rapidly when liquid enrichment cultures are used. Accordingly, no actinobacteria were isolated in the current study by employing liquid enrichments.

The state of an organism, be it actively growing or merely present as a spore, can influence whether it will be cultured during an isolation campaign. Similarly to slow growing species (such as *Mycobacterium* species) spores also require a longer initial growth phase. Fast growing bacteria and fungi tend to out complete slow growing colonies, making isolation of slow-growing bacteria problematic. The addition of antimicrobials can limit the growth of fungi, but may cause added biological stress hampering the already slow growing bacteria. In some cases a cell can also exist in a viable but unculturable state i.e. individual cells may lack the ability to reproduce on a solid agar surface and therefore be undistinguishable.

The isolation of novel species by means of recreating extreme environments as well as selection for a specific subgroup requires a combination of different culturing approaches. For example by recreating the initial environment one would bias the isolation towards individuals which flourish under those conditions and since most novel actinobacterial species are infact shown to grow at a slow rate, this would be counter productive. A recent report on the world wide isolation of actinobacteria by the Basic Research Center of Merck Sharp and Dohme in Spain, suggested that most rare actinobacteria can be isolated on nutrient

limiting media and prolonged incubation periods (Genilloud *et al*., 2011). In the current study both nutrient limiting and nutrient rich media were used to isolate actinobacteria. The isolation media for the five sites were chosen based on the culturability of general actinobacterial and the geochemical properties of the soda lakes. Medium A was designed for the isolation of organisms from Kenyan soda lakes, whilst ISP 2 and ISP4 are general actinobacterial media. The five media were used in combination with temperature (28°C-55°C), salt (1-10% w/v) and pH (7.0-10.1) ranges in order simulate the sampling environment. A series of diluted soil suspensions were spread onto selective plates containing Cycloheximide and incubated for up to 4 weeks. Fungal contamination was still relatively abundant at concentrations up to 50 µg/ml and higher concentrations resulted in limited or no growth of the desired phylum. Table 15 summarizes those combinations which did and did not yield isolates. The two temperatures not included in the Table (28 and 55°C) as well as the 5-10% NaCl 45°C plates did not yield any growth. A 10^{-1}-10^{-6} dilution range was plated per combination per sampling site.

In total 21 morphologically unique specimens were isolated from Lake Arenguade and Lake Shala hot spring 1 (Table 16). Both novel strains ERV7 and ERV8 were isolated on Medium A, pH 10.1 at 37°C. The rest of the samples did not show any identifiable actinobacterial growth. The isolates were grouped according to morphological characteristics such as spore colour, colony morphology, pigment production and colony colour on ISP2 (Fig.40). Most of the M233 medium plates were completely covered by a biofilm composed of Gram-negative rods (see Fig.39) within 3 days, even when the media was supplemented with 10% NaCl. Therefore medium M233 was excluded from liquid isolation. No growth was observed in liquid cultures of Casein medium. Liquid isolations in LB and ISP2, as well as Medium A did not result in the isolation of any actinobacteria.

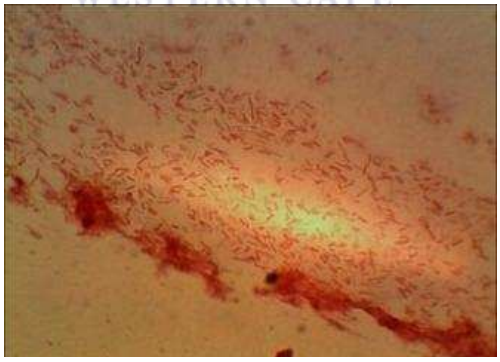


Figure 39. A microscopy photo of the Gram-negative isolate predominant on M233 media plates.

.

Figure 40. Characteristics by which colonies were selected. Left: Gram stain of isolate ERV7. Right: Aerial mycelia give colonies a powdery appearance.

Table 14. Isolation media and parameters used to isolate Actinobacteria with the exception of 28 and 55°C.

Temp	NaCl (%)	Medium A						M233						Casein Medium						ISP2						ISP4						
	pH	7.0	7.5	8.0	8.5	9.0	10.1	7.0	7.5	8.0	8.5	9.0	10.1	7.0	7.5	8.0	8.5	9.0	10.1	7.0	7.5	8.0	8.5	9.0	10.1	7.0	7.5	8.0	8.5	9.0	10.1
30	1	■	■	■																					■	■	■	■	■	■	
	2																														
	3																														
	4																														
	5																		■	■	■	■	■	■							
	10																														
37	1	■	■	■																					■	■	■	■	■	■	
	2																														
	3																														
	4																														
	5																		■	■	■	■	■	■							
	10																														
45	1			■											■	■	■															
	2																														
	3																														
	4																														

*Solid blocks indicate those conditions on which Actinobacteria were found.

Actinobacterial strains belonging to a limited number of genera have been isolated from haloalkaline environments including: *Streptomyces*, *Micromonospora*, *Arthrobacter, Terrabacter, Nesterenkonia* and *Dietzia* (Jones & Grant, 1999; Duckworth *et al*., 1996). All of these genera, with the exception of *Micromonospora*, have been isolated from Kenyan soda lakes.

Table 15. Actinobacterial isolates and the conditions at which they were isolated.

ERV Isolate	Sample site	Media	pH	NaCl % w/v	Temp (°C)
1.1.1-1.1.2	LA	CM[a]	7	5	45
1.4.1-1.4.2	LA	CM[a]	7	5	55
1.14	LA	CM[a]	7	5	37
1.15	LA	CM[a]	7	5	37
1.2.1-1.2.2	LA	CM[a]	8	5	45
1.3.1-1.3.2	LA	CM[a]	8	5	55
1.6	LA	CM[a]	8	5	37
1.7	LA	CM[a]	8	5	37
1.8	LA	CM[a]	8	5	37
1.9	LA	CM[a]	8	5	37
1.10	LA	CM[a]	8	5	37
1.11	LA	CM[a]	8	5	37
1.12	LA	CM[a]	8	5	37
1.16	LA	CM[a]	8	5	37
1.17	LA	CM[a]	8	5	45
1.18	LA	CM[a]	8	5	37
1.20	LA	CM[a]	8	5	45
1.21	LA	CM[a]	8	5	45
1.19	LA	M233[b]	8	5	45
1.13	HS1	CM[a]	9	5	37
ERV 7	LAb	MA[c]	10.1	0	37
ERV 8	LC	MA[c]	10.1	0	37

[a] Casein-cellulose medium
[b] M233 medium
[c] Medium A

4.2. ARDRA of isolates

The ARDRA patterns for ERV 1.6-1.16, with the exception of 1.13, all showed a banding pattern similar to one LAb clone (clone LAb.C1) with a low (86%) similarity to a *Bacteroidetes* species. In a library with such a large diversity (as has been discussed for Lake Abitjata) it is possible for two non-related species to show a similar banding pattern, especially since agarose gels have a low resolution when examining ARDRA patterns. However, all of the isolates were from Lake Arenguade, with the exception of ERV1.21, a slow growing *Streptomyces* species which was isolated from Lake Shala.

4.3. DGGE of isolates

DGGE analysis of isolates ERV 1.4-1.17 (Fig.41) showed a similar banding pattern with at least 3 distinct bands on a 60-80% denaturing gradient. Slight differences in GC content were not observed on this gradient. The two top bands could possibly be due to the amplification of two different *rrn* operons with a different GC content, with the smear lower on the gel being mainly primer dimers formed during the PCR amplification. Inter-gel comparison showed no overlapping bands with any of the community actinomycete-specific profiles generated from the metagenomic libraries (data not shown).

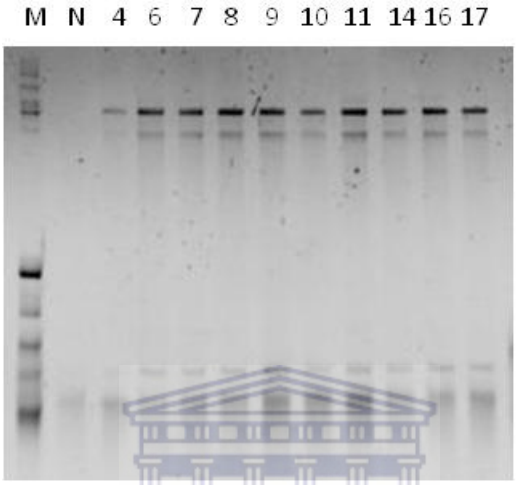


Figure 41. DGGE profile of isolates ERV1.4-1.17 on a 60-80% urea formamide denaturing gel. M: Ladder, N: negative control no DNA

4.4. 16S rRNA sequence determination and morphological characteristics

Universal bacterial primers were used to obtain an approximately 1300-1500bp fragment of the 16S rRNA gene for 18 of the 21 isolates. Despite several alterations to the growth media three of the 21 isolates exhibited extremely slow growth rates (30-40 days for visible colonies to appear) and were therefore excluded from further analysis. Eight (ERV 1.4, 7, 10, 12, 14, 16, 17) of the remaining 18 morphologically different isolates all showed 99% similarity to *Streptomyces griseorubens*. Although the colony morphology of these isolates varied tremendously, the 16S rRNA genes sequences were identical (see Table 17). A comparison of mycelial colour and physiological growth optimums between the different 16S rRNA groups confirmed that they were infact different species (Table 18). The remaining ten isolates all showed 99% similarity to various *Streptomyces* species. All the closest relatives were originally isolated from soil and grow optimally at 26°C, including *S. griseorubens* which was originally isolated from marine sediment.

Table 16. Closest 16S rRNA gene sequence matches of all actinobacterial isolates.

Isolate	Isolation site	length	% Identity	Closest match	Accession nr	Origin
ERV1.1	LA	954bp	99	*Streptomyces griseorubens* strain S5MS3	JN400106.1	Chilika lake India
ERV1.2	LA	912bp	99	*Streptomyces albogriseolus* strain S2NW2	JN400095.1	Chilika lake India
ERV1.3*	LA					
ERV1.4	LA	907bp	99	*Streptomyces griseorubens* strain 3433	JN180193.1	China: Shanxi Province soil
ERV1.6	LA	1283bp	99	*Streptomyces variabilis* strain GSN08	JN314850.1	India
ERV1.7	LA	1375bp	99	*Streptomyces griseorubens* strain 3433	JN180193.1	China: Shanxi Province soil
ERV1.8	LA	1352bp	99	*Streptomyces albogriseolus* strain S2NW2	JN400095.1	Chilika lake India
ERV1.9	LA	1259bp	99	*Streptomyces variabilis* strain GSN08	JN314850.1	India
ERV1.10	LA	1360bp	99	*Streptomyces griseorubens strain 3433*	JN180193.1	
ERV1.11	LA	1374bp	99	*Streptomyces variabilis* strain GSN08	JN314850.1	India
ERV1.12	LA	1444bp	98	*Streptomyces griseorubens* strain 3433	JN180193.1	China: Shanxi Province soil
ERV1.13*	LA					
ERV1.14	LA	1436bp	99	*Streptomyces griseorubens* strain 3433	JN180193.1	China: Shanxi Province soil
ERV1.15	LA	1385bp	96	*Streptomyces variabilis* strain 7525	JN180216.1	China: Shanxi Province soil
ERV1.16	LA	1369bp	99	*Streptomyces griseorubens* strain 3433	JN180193.1	China: Shanxi Province soil
ERV1.17	LA	1359bp	99	*Streptomyces griseorubens* strain 3434	JN180193.2	China: Shanxi Province soil
ERV1.18	LA	1227bp	100	*Streptomyces variabilis* strain GSN08	JN314850.1	India
ERV1.19*	LA					
ERV1.20	LA	1415bp	99	*Streptomyces variabilis* strain 7525	JN180216.1	China: Shanxi Province soil
ERV1.21	HS1	1415bp	98	*Streptomyces variabilis* strain 7526	JN180216.2	China: Shanxi Province soil

* Isolates excluded from further analysis on account of slow growth rates

Table 17. Closest characterized strains to cultured isolates.

	Streptomyces sp. ERV1.1	*Streptomyces* sp. ERV1.8	*Streptomyces* sp. ERV1.21	*S. albogriseolus*	*S. griseorubens*
Isolated from	Soda lake soil	Soda lake soil	Soda lake soil	Mangrove sediment	Sediment
Aerial spore mass colour	Grey	White	Grey	Grey	Silky grey
Substrate mycelia colour	Cream	Cream	Cream	Grey	Cream
pH optima	8.0	8.0	8.0	**N/A**	**7.2**
Growth Temp	37	37	37	26	28

*N/A: Not available

4.5. Characterization of novel isolates

4.5.1. Optimal Growth conditions

As Medium A has an extremely high buffering capacity; Bennett's medium was used to establish the pH growth range. Both isolates were unable to grow at a pH below 7.0 (Fig.43). The growth rate remained relatively constant from pH 10.0 to 11.0. Growth rates were only determined up to pH 11, after which point the sodium carbonate in the growth medium precipitated out of solution. A pH of 12, 13 and 14 supported growth, but was not included in the growth curves due to the precipitation of sodium carbonate during the experiment. Precipitation only occurred after day 4 which coincided with a drop in pH. The precipitated salt crystals were removed, photographed under a light microscope (Fig.42) and placed on agarose plates which resulted in growth suggesting cells were still viable.

Figure 42. A 400x magnification of the ERV7 and 8 cells surrounding the precipitated salt crystals prior to growth on solid media.

The cell weight measured in mg is considerably low considering the length of the growth period however the wet cell volumes were up to 4% v/v at day 6 (data not shown) for both strains. The wet mycelial volume to dry weight ratio, did however follow the same growth curve suggesting that the low mass might just me an anomaly.

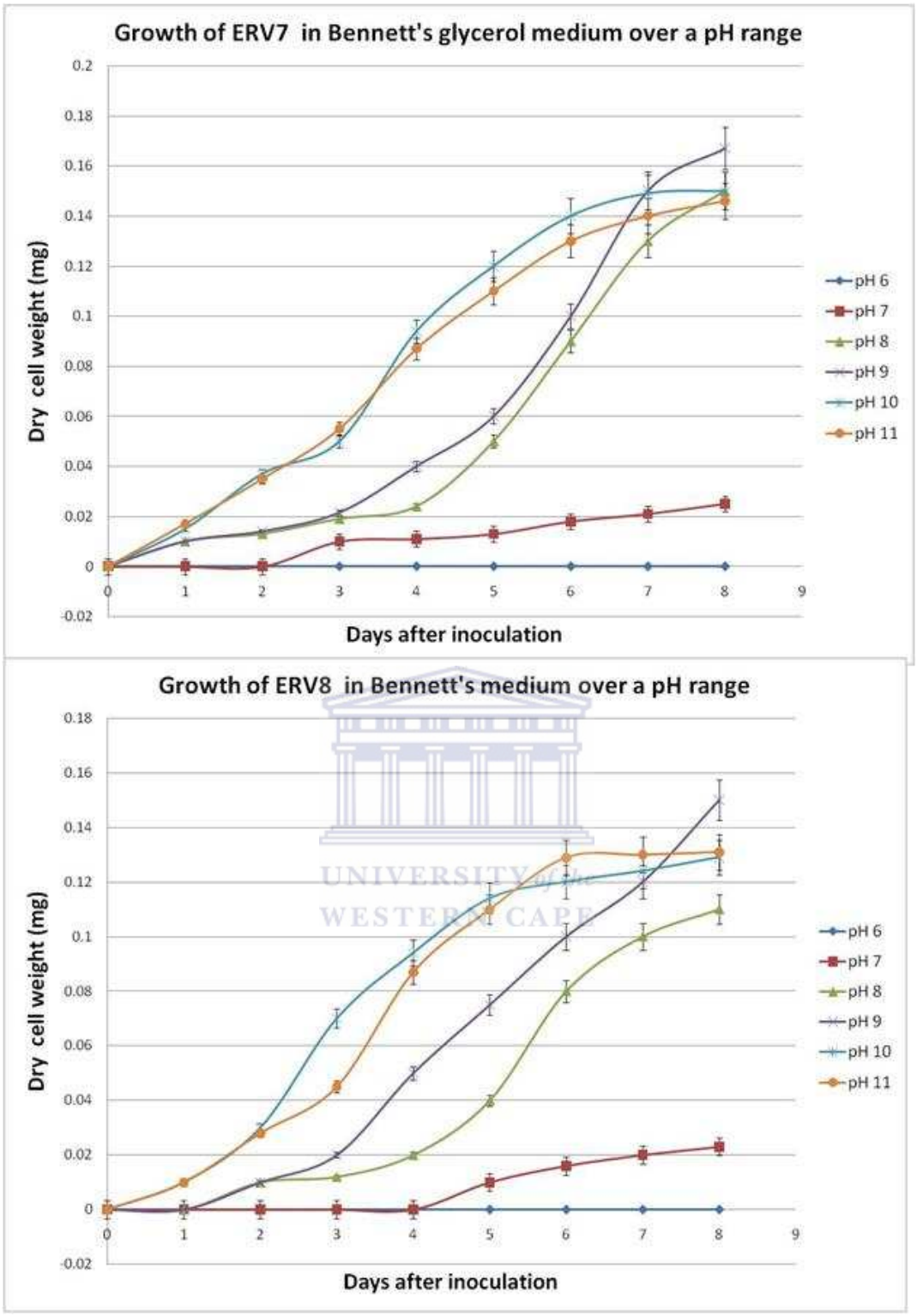
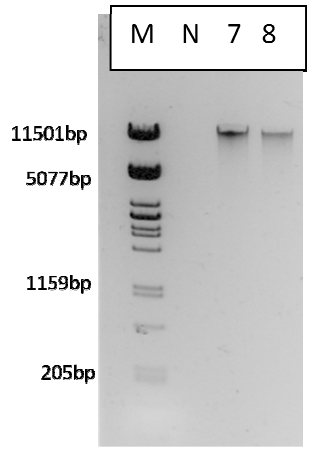


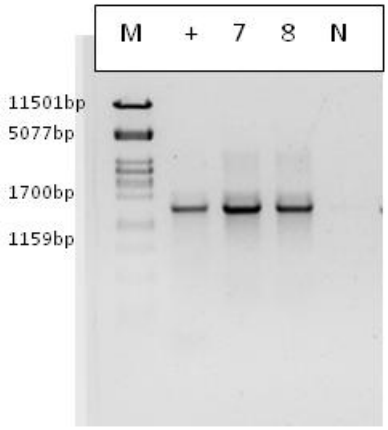
Figure 43. Growth profiles for both *Streptomyces* species in Bennett's glycerol between pH 6.0 and 11.0 at 37°C. Standard deviations indicate technical repeats.

4.5.2. DNA extraction and PCR

Genomic DNA isolation was done using a modified CTAB/SDS method (see section 2.2). Genomic DNA was approximately 11kB in size and of good quality with a concentration of 2 µg/µl (Fig.45). The full 16S

rRNA gene was amplified using primers F1 and R5 (see Fig.44) and sequenced.





the
of a
N:

l depicting the full 16S
the two novel isolates.
da ladder N:No DNA,

4.5.3. Phylogenetic analysis of novel isolates

Sequencing of the 16S rRNA genes with primers 27F, F3 and 1492R allowed for the construction of consensus sequences 1399 bp and 1389 bp in length for ERV8 and ERV7, respectively, both with a G+C content of 60%. BLAST analysis showed that both isolates shared 96% 16S rRNA gene sequence similarity to all validly published *Streptomyces* species. The two isolates shared 99% similarity; however physiological characteristics varied (see section 4.5). A phylogenetic tree comparing the isolates to the nearest characterized strains was constructed (Fig.46 and Table 19). Based on BLAST analysis both isolates were most closely related to two uncharacterized species isolated from an alkaline environment and these strains have been included in the phylogenetic tree.

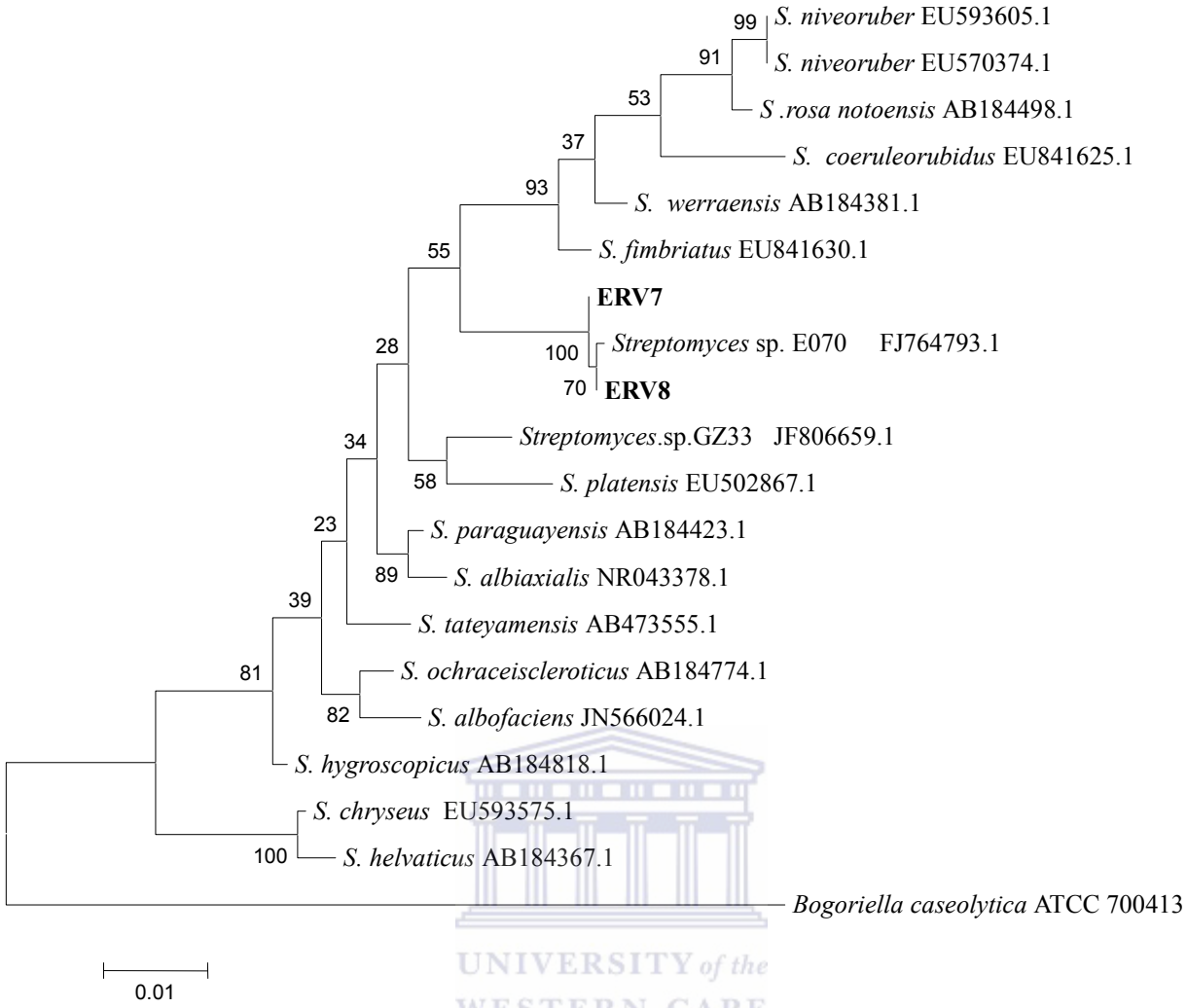


Figure 46. A phylogenetic tree of the aligned partial (~1300bp) 16S rRNA gene sequences of the two *Streptomyces* isolates and the type strains of the most closely related genera. A maximum likelihood Heuristic method with a nearest-neighbour interchange was used. Boot strap values (in percent) are shown at the nodes. *Bogoriella caseolytica* was used to position the root of the tree.

Table 18. Nearest matches to two novel *Streptomyces* species.

Species	Match	Accession	Origin
***Streptomyces* sp. GZ33**	97%	JF806659.1	Seagrasses, *Thalassia hemprichii*
***Streptomyces* sp. E-070**	99%	FJ764793.1	Lake Elmenteita, Kenya
Streptomyces hygroscopicus	96%	AB184818.1	
***Streptomyces platensis* strain F160280**	96%	EU502867.1	
Streptomyces ochraceiscleroticus	96%	AB184774.1	
Streptomyces albofaciens	96%	JN566024.1	
Streptomyces paraguayensis	96%	AB184423.1	
Streptomyces rosa* subsp. *notoensis	96%	AB184498.1	
Streptomyces fimbriatus	96%	EU841630.1	Baoding, China
Streptomyces tateyamensis	96%	AB473555.1	Marine Sponge
Streptomyces werraensis	96%	AB184381.1	
***Streptomyces niveoruber* strain 173757**	96%	EU593605.1	Xinjiang, China
***Streptomyces niveoruber* strain 173756**	96%	EU570374.1	Xinjiang, China
Streptomyces albiaxialis	96%	NR_043378.1	
Streptomyces chryseus	95%	EU593575.1	Xinjiang, China Unpublished
Streptomyces helvaticus	95%	AB184367.1	
***Streptomyces coeruleorubidus* strain HBUM174910**	95%	EU841625.1	Baoding, China

4.5.4. Polyphasic characterization of two novel *Streptomyces* strains

The genus *Streptomyces* has been characterized as a diverse physiological group that contains: mesophilic, neutrophilic, alkaliphilic and thermophilic strains. Alkaliphilic actinobacteria were first detected by Taber *et al.* in 1960. Most alkaliphilic isolates characterized thus far belong to either the genus *Nocardia* or *Streptomyces*. Several studies have isolated alkaliphilic streptomycetes from locations such as Lake Sul'fatnoe (Russia) (Selyanin *et al*., 2005), Ross Links sand dunes (United Kingdom) (Antony-Babu & Goodfellow, 2008), Mongolian steppes (Norovsuren *et al*., 2007), Lake Nakuru (Kenya) (Solingen *et al*., 2001), Lake Loktak (India) (Singh *et al*., 2009) and Lake Chaka (China)(Li *et al*., 2005). Relatively few validly published alkaliphilic *Streptomyces* species have been characterized. The physiological profile of alkaliphilic *Streptomyces* species vary to a large extent. Carbon or nitrogen source utilisation profiles vary between alkaliphic strains and differences in growth patterns are evident when different buffering agents (such as KOH, Na_2CO_3, K_2CO_3 and NaOH) are included in the growth media.

4.5.4.1. Physiological and morphological characterization

Morphological observation of strains ERV 7 and ERV8 after 21 days incubation on ISP4 medium revealed

that both possessed the typical characteristics associated with the genus *Streptomyces*. Substrate mycelia were well developed without any fragmentation. Spore chains were *Rectiflexibles* (ERV7) and *Spirales* (ERV8) (see Fig.47). Light microscopy of the two strains revealed long spore chains that with three to four non-motile spores (Fig. 47). Aerial mycelia colour varied from white to beige, light brown and purple depending on media composition (Table 17). Determining the colour of the substrate mycelia was hindered by the presence of dark brown to black diffusible pigments (Table 20). These diffusible pigments were pH insensitive. For both strains melanin production was observed on ISP6 after 7 days incubation and on ISP7 after 14 days. Mycelial pigment colour changes were more prominent in strain ERV7 than in ERV8 where in liquid medium A ERV7 cells turned grey upon stationary phase whereas ERV8 remained beige, which could not be attributed to spore formation.

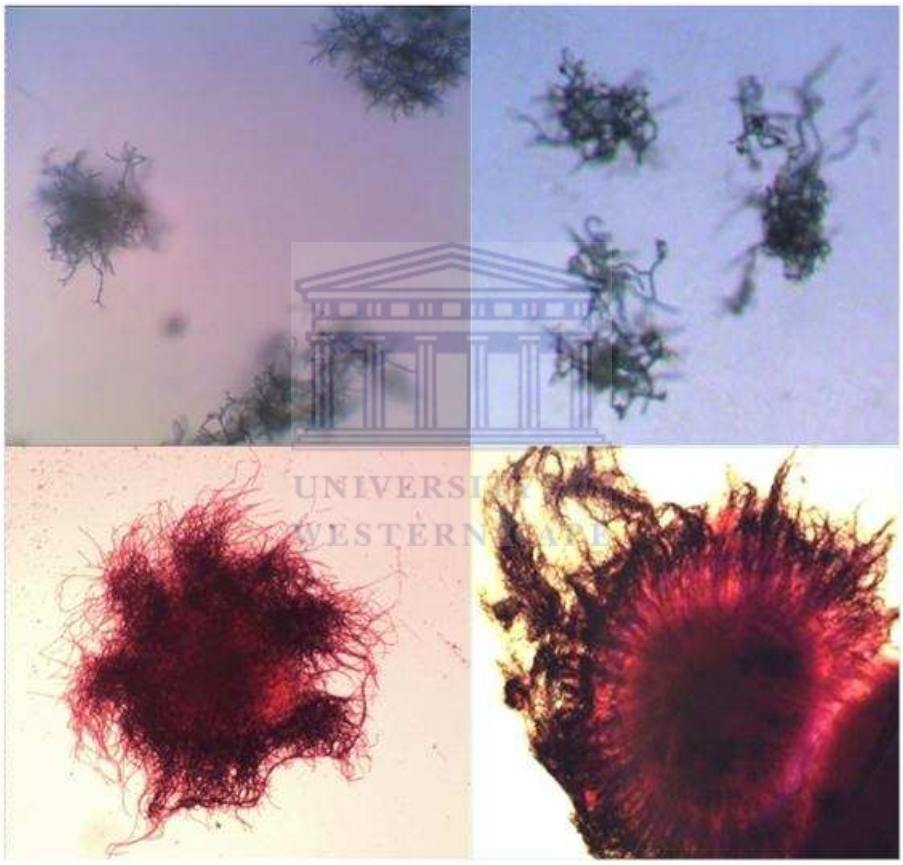


Figure 47. Microscopic enlargement of substrate mycelia at 400x magnification as well as Gram stains of the two novel strains. Top left: Strain ERV7 on ISP4. Top right: Strain ERV8 on ISP4, Bottom right: Gram stain ERV8, Bottom right: Gram stain ERV7.

Table 19. Characteristic properties of mycelia on various media.

Agar medium	Growth	Colour of mycelium			
		Aerial	**Munsell Values**	**Substrate**	**Munsell Values**
Yeast extract Malt extract (ISP2)	**++**	**White**		**Cream**	**N9 9/0**
Inorganic salt/starch (ISP4)	**+**	**White to light brown**		**Cream**	
Glycerol/Asparagine (ISP5)	**+**	**White**		**Cream**	
ISP6	**+**	**Light brown**		**Dark Brown**	
Tyrosine (ISP7)	**+**	**Purple**	**2.5P 6/6**	**Black**	
Bennett's medium	**++**	**White**		**Brown**	
Modified Bennett's medium	**+++**	**Cream**		**Dark Brown**	

Physiological characterization revealed that both strains could grow between pH 7.0 and 12.0, with a pH optima between 9.0–10.0 (Table 21). Both ERV7 and ERV8 grew optimally at 37°C, therefore all subsequent physiological characterizations were conducted at this temperature. Both strains could grow in the presence of 0-10% NaCl, but were inhibited by 12% (w/v) NaCl. Initial growth studies showed that certain compounds were only utilised in the presence of Na_2CO_3 and trace elements. An example is the degradation of hypoxanthine and hippurate (see Table 22). Sole nitrogen and carbon source utilisation patterns are summarized in Table 22. Differences include the finding that ERV8 utilised glycine, histidine, tyrosine and valine as sole nitrogen sources, whereas ERV7 only utilised methionine. In addition, ERV7 was unable to use lactose as a sole carbon source, whilst ERV8 could not use xylose. Interestingly, while both strains were sensitive to crystal violet when Na_2CO_3 and trace elements were added to the growth medium, both strains were able to grow. ERV7 and ERV8 were both resistant to Streptomycin, Chloramphenicol and Ampicillin (at 100 µg/µl).

Physiologically, neither strain resembled their closest phylogenetic neighbours. Based on genetic and phenetic data, strains ERV7 and ERV8 are likely to be strains of a single species and as such they will be discussed as a collective species. Physiologically, *Streptomyces niveoruber* is the closest phylogenetic neighbour to the species. Differences include the ability/inability of the novel species to utilise xylose, arabinose and *meso*-inositol (see Table 21), while the optimal growth and pH requirements are also higher. Additionally, the novel species was compared to three alkaliphilic *Streptomyces* species. All the alkaliphilic strains appear to utilise a wide range of carbon sources, with *S. rochei* being the exception. Similarly, all alkaliphilic strains have relatively high tolerance for NaCl. Conversely, the published alkaliphilic strains have higher temperature optima.

Table 20. Morphological and phenotypic characteristics that differentiate strain ERV7 and ERV8 from closely related species of the genus *Streptomyces*.

	ERV7	***ERV8***	***S. coerule-orubidus***	***S. niveoruber***	***S. fimb-riatus***	***S. decca-nensis****	***S. tana-shiensis****	***S. rochei****
Similarity %			**95**	**96**	**96**	**94**	**94** **94**	
Spore morphology	***Recti flexibles***	***Spirales***	N	N	***Spirales***	***Recti flexibles***	***Recti flexibles***	***Spirales***
ISP4	Light brown	Light Brown	Brown beige	Colourless	Yellow	Beige	Red	Light grey
Melanoid pigment	**+**	**+**	**+**	**-**	**+**	**-**	**+**	
Carbon utilisation:								
D-Glucose	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**
L-Arabinose	**-**	**-**	**+**	**+**	**+**	**+**	**+**	**+**
Sucrose	**+**	**+**	**-**	**+**	**+**	**-**	**+**	**+**
D-Xylose	**+**	**+**	**-**	**-**	**+**	**-**	**-**	**+**
Meso-Inositol	**-**	**-**	**+**	**+**	**-**	**+**	**+**	**+**
D-Mannose	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**-**
D-Fructose	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
D-Raffinose	**+**	**+**	**+**	**-**	**-**	**+**	**+**	**+**
Enzyme activity								
Gelatin	**+**	**+**	**+**	**+**	**-**	**+**	**N**	**+**
Urea	**-**	**-**	**-**	**+**	**-**	**+**	**N**	**-**
Nitrogen utilisation								
Arginine	**+**	**+**	**-**	**+**	**-**	**+**	**N**	**N**
Lysine	**+**	**+**	**-**	**+**	**-**	**-**	**N**	**N**
Optimum Temp	**37**	**37**	**28**	**28**	**28**	**40**	**40**	**30**
NaCl	**10%**	**10%**	**N**	**2.50%**	**0%**	**5%**	**7%**	**7%**
pH range	**pH 7-12**	**pH 7-12**				**pH 9-12**	**pH 7-12**	**pH 6.5-9.5**
Cell Wall sugars	**Ribose,Galactose**	**Ribose,Galactose**	**N**	**N**	**N**	**No diagnostic sugars**	**Galactose Glucose**	**No diagnostic sugars**
DAP	**LL-DAP Glycine**	**LL-DAP Glycine**	**N**	**N**	**N**	**LL-DAP and Glycine**	**LL-DAP and Glycine**	**LL-DAP**

*Alkaliphilic *Streptomyces* sp.

N- Not attainable

Table 21. Differences in sole carbon and nitrogen source utilisation between the ERV7 and ERV8 *Streptomyces* strains.

	ERV7		ERV8	
Utilisation of nitrogen sources:	**+TE**	**-TE**	**+TE**	**-TE**
Asparagine	+	+	++	++
L-serine	++	++	+	+
Arginine	++	++	++++	++++
L-Threonine	++	++	+	+
Potassium nitrite	++	++	+	+
Utilisation of nitrogen sources:				
L-Valine	-	-	+	+
L-Tyrosine	-	-	+	+
L-Histidine	-	-	+	+
L-Methionine	+	+	-	-
Glycine	-	-	+	+
Utilisation of carbon sources:				
D-Cellobiose	+	+	++	++
D-Glucose	+	+	++	++
D-Raffinose	+	+	+++	+++
D-Xylose	+	+	-	-
Sucrose	+	+	++	++
D-Trehalose	+	+	++	++
Lactose	-	-	+	+

+ Weak ++ Good +++Substantial - None

4.5.4.2. Chemotaxonomic characterization

Both strains' cell walls contain LL-diaminopimelic acid and glycine (Fig.49), which matches cell wall chemotype I (Lechevalier & Lechevalier, 1970). From whole cell hydrolysates galactose and ribose were identified to be the cell wall sugars, whilst the common phospholipids detected included phosphatidyl-monoethylethanolamine and cardiolipin (Fig.48) which correspond to those expected for *Streptomyces* species (Lechevalier *et al*., 1977).

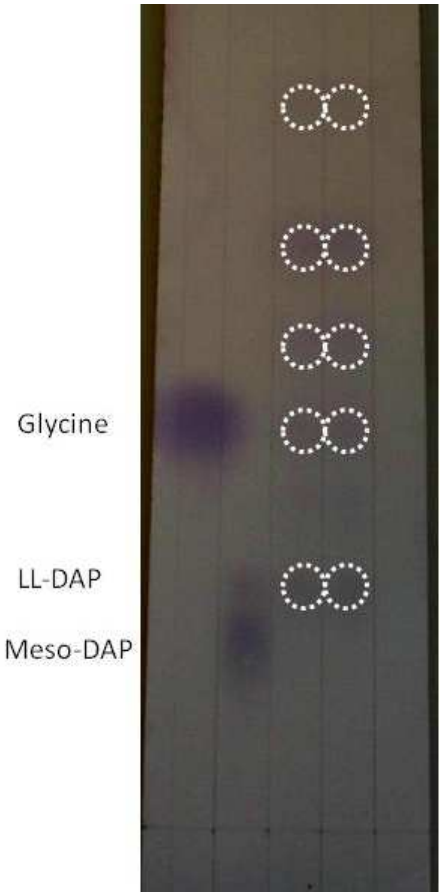


Figure 49. A cellulose aluminium TLC plate indicating the migrated spots of the cell wall sugar extract from both novel isolates. From Left: Glycine standard, DAP standard, ERV7, ERV8.

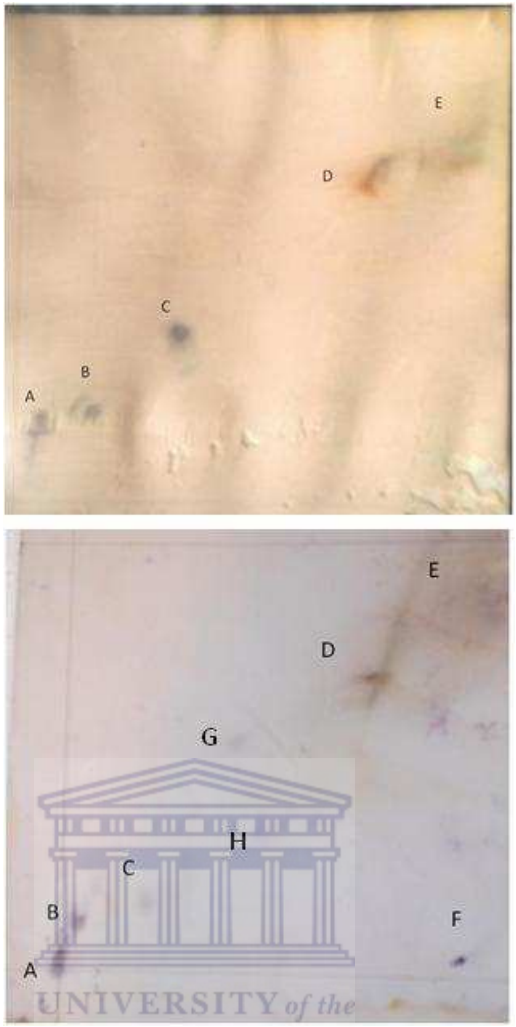


Figure 48. A 2 dimensional TLC plate indicating the migration patterns of the phospholipids extract from both isolates. Above: ERV7 A- B: Unknown C: Phosphatidylserine D:Cardiolipin E: Phosphatidyl monoethylethanolamine. Below: ERV8 A- B:Unknown C:phosphatidylinositol mannoside D: Cardiolipin E: Phosphatidyl monoethylethanolamine F:Unknown G:Lyso-phosphatidylglycerol H:Di-lysocardiolipin

4.6 Partial characterization of the xylanase enzymes produced by ERV7 and ERV8

The production of extracellular enzymes in alkaline environments requires the strain to adapt to the harsh conditions of the surrounding environment. In soda lakes, this encompasses high pH, temperature and NaCl concentrations. Since *Streptomyces* species are notorious extracellular enzyme producers it follows that these enzymes would be naturally adapted for haloalkaline conditions. Both isolates were therefore tested for a range of enzymatic activities of which xylanase activity was the most prominent. Xylanases are hydrolases which depolymerise the xylan-component of plant cell walls into xylooligosaccharides and xylose (Subramaniyan & Prema, 2002). Alkaliphilic xylanases are used in the production of biofuels, agro-waste

treatment, paper and pulp manufacturing to breakdown polymers into various monomers.

The two novel strains, ERV7 and ERV8, both showed extensive xylanase activity on Bennett's glycerol and xylan media, therefore were selected for further characterization. ERV7 and 8 were incubated on Bennett's medium pH 10.1 agar supplemented with 1% (w/v) RBB and 1% (w/v) birchwood xylan. The clearance zones were 5.1 mm (ERV7) and 6.2 mm (ERV8), respectively (see Fig.50).

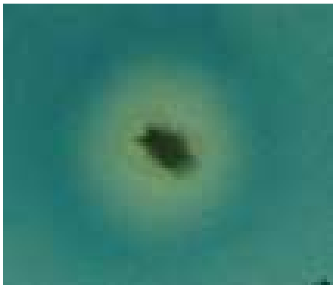


Figure 50. ERV7 cultured on Bennett's medium pH 10.1 supplemented with RBB and birchwood xylan (1% w/v).

Xylanase activity was determined using the DNS assay, and was measured according to their ability to liberate xylose. Since extracellular protein concentrations are often below the detection limit of most protein assays, for the DNS assays the activity is measured by the µM/min xylose liberated. Initially a standard curve was constructed for several buffers, however due to the low degree of variation between the standard curves a pH 7.4 buffer was used for the xylose standards.

Both strains showed xylanolytic activity (Fig.51). Xylanase production was higher when the strains were cultured in 1% w/v birchwood xylan compared to 0.1%. The ERV7 xylanase was most active at pH 6.5 at 37°C, which corresponded to 47.5 µM xylose liberated per min. ERV8 xylanase was most active at pH 8.5 corresponding to 48.3 µM xylose liberated per min. This can be considered as relative high when compared to other cellulose-free xylanases such as those produced by *Streptomyces roseiscleroticus* (16.2 µM/min) and *Bacillus circulans* AB16 (19.28 µM/min) (Subramaniyan & Prema 2002). However, it has been shown that certain cations may play a role in increasing the activity of thermotolerant/thermophilic enzymes (Tan *et al.,* 2001). The effect of the buffer components were however not nested in the current study.

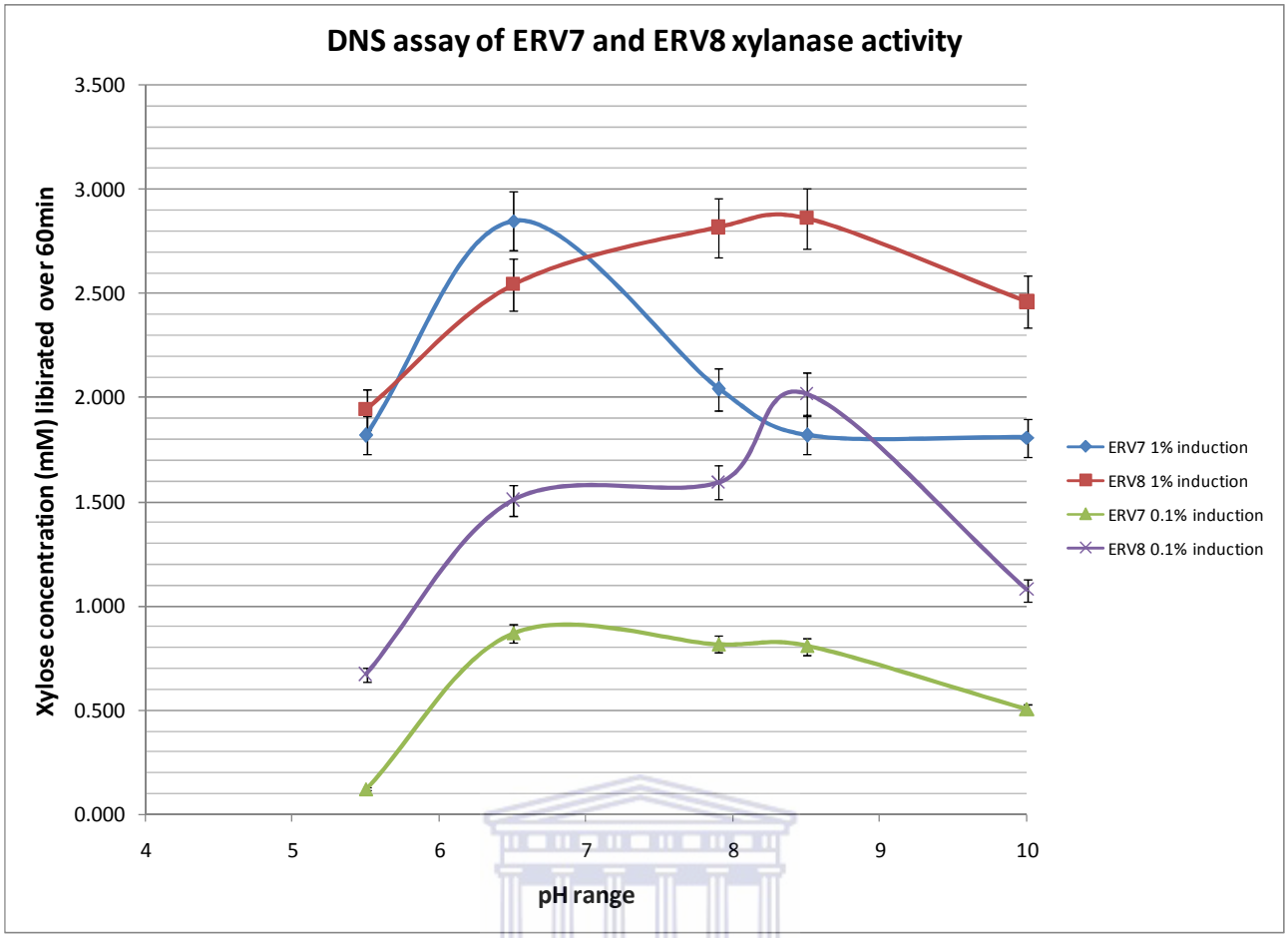


Figure 51. DNS assay of ERV7 and ERV8 crude extract over a pH range at 37°C. The graph compares 0.1% and 1% birchwood xylan as induction parameters. Standard deviations indicate technical repeats.

The temperature range was tested at 30°C, 37°C and 45°C over the pH range of pH 6.5-8.5 ERV7's xylanase activity almost doubled when assayed at 45°C compared to 37°C (Fig. 52), while ERV8's xylanase activity increased from 48-59 µM xylose liberated per min over the same temperature range. Higher temperatures seemed to have a positive effect on enzymatic activity, although the possible explanation for this finding is not clear as both strains were isolated from moderate temperature environments and have a growth temperature optima of 37°C. Further analysis needs to be conducted to establish what effect temperature has on the xylanases activity.

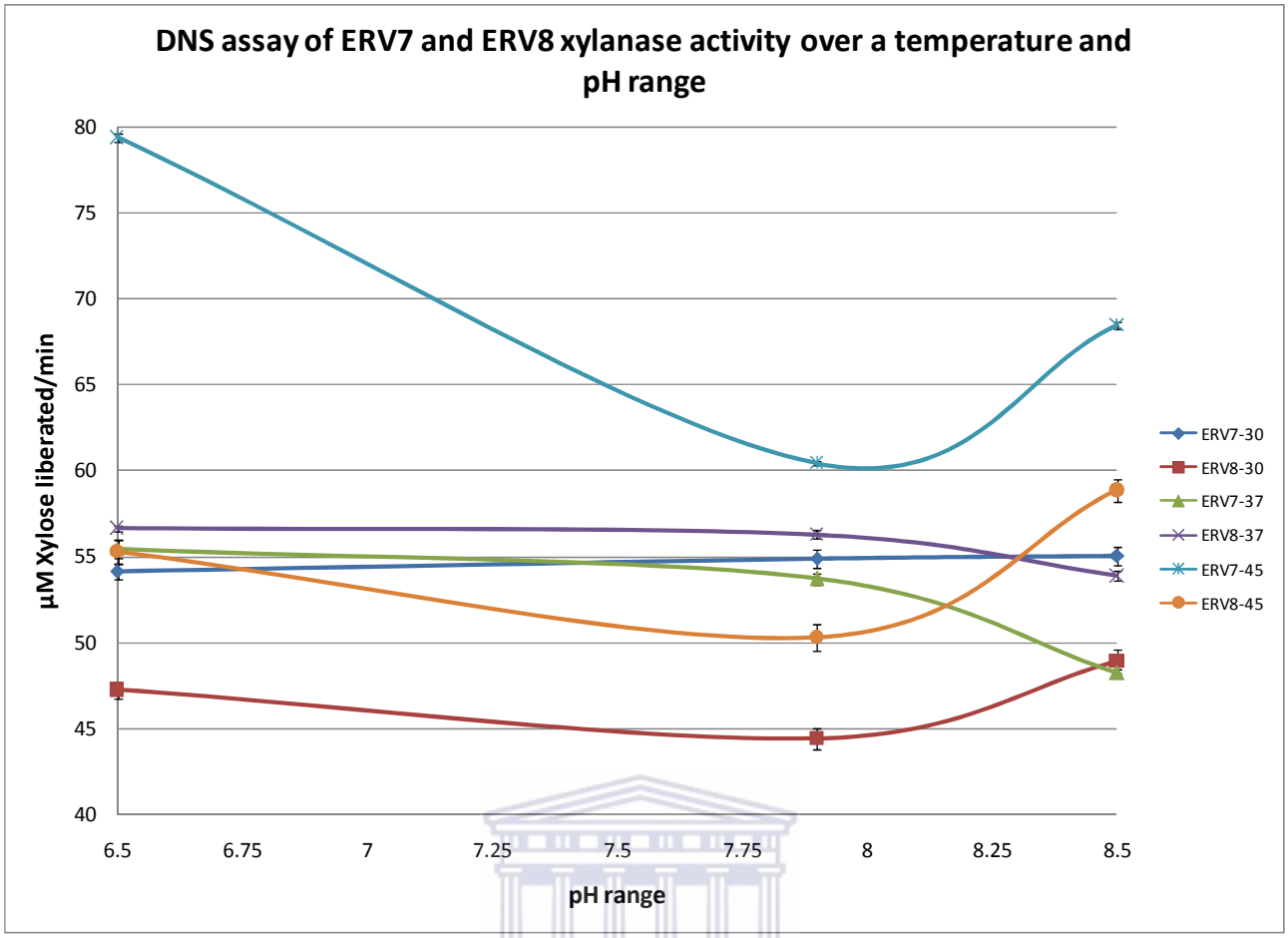


Figure 52. DNS assay of ERV7 and ERV8 determining xylanase activity over a temperature range (30°C, 37°C & 45°C) at pH 6.5, 7.9 and 8.5.Standard deviations indicate technical repeats.

4.7. Conclusion

The culturing of novel or rare organisms is a classical approach by which to study the physiology of an organism as well as its role and abundance in a specific environment. Enrichment for rare actinobacteria, sought after in this study, can be enhanced by the pre-treatment of sediment with various techniques which also limit contamination by fast-growing bacteria. These include phenol treatment, heat pre-treatment, UV-light exposure, low antibiotic supplements and differential centrifugation (Hayakawa *et al.,* 2000, Galatenko and Terekhova, 1990, Kurtböke and French, 2007). Apart from differential centrifugation other pre-treatments require a high viable cell count due to the loss of a large proportion of actinobacteria during the treatment. The extreme conditions in the Ethiopian Rift Valley lakes and the limited publications on the isolation of actinobacteria in the Kenyan Rift Valley lakes, suggested that any pre-treatments might further limit the chances of isolating viable cultures (Jones & Grant, 1999). Therefore, none of the above mentioned treatments were included in the current study.

All isolates were of the *Streptomycetaceae* family, which was to be expected due to the identification scheme

incorporated. The slow growth and low detection numbers suggest that the large variation in storage temperatures and prolonged exposure to subzero temperatures drastically affected the culturability of the soil. The isolation of two novel strains, however suggest that optimum culturing conditions may yield a larger variety of actinobacteria from those sites which yielded no viable colonies.

Chemotaxonomic, physiological and phylogenetic markers suggest that the two isolates, ERV7 and ERV8, are two strains that form a single novel species. As the 16S rRNA gene similarity to the closest validly published species is 96% the novel species can be recognized as a unique genomic species. Additionally, based on phenotypic characteristics the species can be distinguished as two distinct strains. Distinguishing features suggest that both strains belong to the genus *Streptomyces.* The strains ERV7 and ERV8 represent a novel species, for which the name *Streptomyces purpura*, sp. nov. is proposed. The type strain of the species is ERV 7.

4.7.1. Description of *Streptomyces purpura* sp. nov.

Streptomyces purpura (ex. *purpura*. L. adj. purple) strain ERV7.
Colonies appear beige and velvety on most media. Aerial mycelium are well developed and do not fragment. Three spores are borne on each spore chain which appear *rectiflexible* or *spirales* depending on strain. A dark brown non-pH sensitive diffusible pigment is produced on ISP6. Cellobiose, fructose, glucose, glycerol, sucrose, mannitol, mannose, raffinose, starch, trehalose and xylose are utilised as sole carbon sources. Strains are also able to utilise arginine, asparagine, lysine, methionine, potassium nitite, serine, threonine and xylitol as the sole nitrogen source. It is unable to utilise arabinose, lactose, *meso*-inositol, histidine, tyrosine and valine, but degrades casein, gelatine, hippurate, hypoxanthine, starch, tween 80 and xylan. It does not however degrade cellulose. It is resistant to Streptomycin, Ampicillin and Chloramphenicol at working concentrations. Growth occurs on media containing crystal violet and NaCl 0-10% (w/v). The growth range spans pH 7.0 to 12.0, and from 28 to 40°C with optimum growth in media supplemented with Na_2CO_3 at a concentration of 10% (w/v) at 37 °C, pH 10.0. Ribose and galactose are the characteristic whole cell sugars with cell wall LL-DAP and glycine components. Phosphatidyl ethanolamine is the diagnostic phospholipids (Cell wall type I).
Isolated from alkaline habitats (a soda lake in Ethiopia). The type strain is strain ERV7.

CHAPTER 5 CONCLUSION AND FUTURE PROSPECTS

Soda lakes are the dominant naturally occurring alkaline systems in the world, of which the Kenyan and Ethiopian Rift Valley lakes form one of the cornerstone model systems. The low alkali metal composition along with high UV-radiation, pH, salinity and cation concentrations creates a unique niche in which to study the evolution of microbial ecology. Until recently, these lakes have remained unexplored by molecular techniques and the true microbial diversity is unknown. The lakes and hot springs serve as a critical water resource for the local community, and due to rapid geochemical changes in lake compositions it is imperative to source a baseline of microbial diversity to which future diversity studies can be compared. The aims of this study were to establish if five geographically divided ERV sites mirrored five geographically divided microbial communities, and to assess whether two divergent approaches (culture-dependent and culture-independent) could deliver cooperative estimations of the microbial diversity. In addition, the cultured isolates would be subjected to phylogenetic analysis to establish if novel isolates could assist in understanding the underlying ecological structure. All three aims were not mutually exclusive and therefore both of the first two aims could be answered simultaneously.

The analysis of the eubacterial diversity within the Ethiopian Rift Valley lakes revealed a vastly diverse community within each of the soda lakes. This diverse composition corresponds to similar soda lake subpopulations found in Russia and Kenya (Foti *et al*., 2008; Song *et al*., 2009; Rees *et al*., 2004; Jones & Renaut, 1996; Jones & Grant, 1999). However, several subpopulations identified in other soda lake communities (i.e. ε-Proteobacteria, *Halomonadaceae* and *Ectothiorhodospiraceae* species) were absent from the current study and visa versa, suggesting that some heterogeneity in phylotype detection does occur (Jiang *et al*., 2006). The same was apparent for the actinobacterial subpopulations, although actinobacterial isolates from soda lakes are rare and form a relatively small fraction, if any, in published diversity studies. Clone sequences from the genera *Streptomyces*, *Micrococcus*, *Corynebacterium* and *Microbacterium* have been found in soda lake samples from Russia, China and Kenya with clone sequences often mounting to two or three separate species. The combination of both *Rubrobacterales* and *Actinomycetales* sequences in the ERV Hot springs concur with the actinobacterial composition of hot springs in China, Russia and the USA (Song *et al*., 2009). This supports the theory that actinobacterial populations in hot springs are immensely different to marine and freshwater populations. By comparing the eubacterial and actinobacterial populations between the five sites, it became clear that although the eubacterial communities were similar in composition, the actinobacterial compositions were clearly different at the genus level. However, the question remained if the two approaches (culture-based and culture-independent based) which were used to extract the community composition identified identical or different components of the actinobacterial diversity.

Even though a high degree of similarity exists between clone sequences and isolates it is rare to identify the same species using both culture-based and culture-independent based approaches. None of the cultivated strains from the current study were detected in any of the clone libraries. While this is surprising considering

the estimated high library coverage, other studies have reported similar findings (Malkawi & Al-omari, 2010; Jiang *et al*., 2006). Therefore, both approaches target a subset of the diversity present within the soda lake sediment yielding different results. The estimated diversity of both approaches can be seen as cooperative, as it relies on the advantages and limitations attributed to each technique.

The isolation of extremophilic actinobacteria have increased over the past decade, however the detection of novel strains still pose several hurdles. The inability to culture any of the uncultured clones detected in the actinobacterial libraries suggest that a more direct approach targeting the various families might be more effective. Two of the five media used in this study was solely aimed at the *Streptomycetaceae* family, therefore it is to be expected that the majority of the isolates would fall within this family. Another explanation could be the difference in composition between the site and the isolation medium. While other studies have had a degree of success by using water from the site to construct a complex media with low nutrient supplementation, this was not possible in this study due to the limitations in transporting large volumes of water from Ethiopia to South Africa. To bridge the gap between the cultured isolates and the clone libraries it is possible to do a direct examination via the use of species-specific DNA probes (FISH) designed to detect the isolates. When used in combination with viability stains (i.e. baclight) a relatively accurate estimation of the active species abundance/richness can be attained. The only disadvantage is that viability stains only test for membrane integrity and would not indicate viable, cell wall compromised organisms. The less expensive alternative is colony forming unit (CFU) counts, although it has proven to be highly inaccurate given the diverse array of colony morphologies of actinobacteria. Both techniques were considered to be beyond the scope of this project but could be applied in future research.

The culture-independent study detected a single actinobacterial sequence in the eubacterial libraries, while the actinobacterial diversity detected in the actinobacterial-biased libraries was relatively high. This highlights the importance of primers and library coverage in microbial diversity studies. Since the actinobacterial 16S rRNA gene fragments were amplified from the original full length 16S rRNA gene products used for the eubacterial libraries, these ribotypes should be present in the aforementioned library. The absence suggests a larger number of clones should be screened to capture the full diversity of such a library and that the use of class-specific primers gives a higher resolution image of the subpopulation. The low amount of phylotypes detected in each lake and the low similarity to any type strain suggest that the actinobacteria present in the lakes are vastly different from any currently catalogued (Genbank and RDP). This complicates the assumption that each lake may only contain a low actinobacterial diversity, since the actinobacterial primers are based on known/common actinobacterial 16S rRNA genes. Even though the actinobacterial-primers used in this study are a vast improvement on recently published actinobacterial primers (26-213% increase in specificity), it is possible that a larger diversity may exist and that the primers were unable to detect them (Stach *et al*., 2003). This can be remedied by using the sequences derived from cultured isolates in combination with extremophilic actinobacterial 16S rRNA genes to design new improved

primers which could target certain rare actinobacteria. The discrepancy between the cultured diversity and culture-independent diversity emphasizes the need for a multi-faceted approach to community ecology (Head *et al*., 1998; Hill *et al*., 2000).

A novel organism can form the single peripheral focal point that ties the existing physiological capabilities of an organism with the geophysico-chemical environmental pressures, the larger haloalkaliphilic community and the interactive microbial ecology. As Prof. Baas Becking suggested: "alles is overall; maar het milieu selecteert" which translates to: 'everything is everywhere, but, the environment selects' (Baas Becking, 1934). The two schools of thought maintains that (a) all microorganisms have a low degree of endemism and a cosmopolitan distribution or (b) high population numbers cause high dispersal which prevents isolation. Based on the current study, it is evident that the commonly distributed organisms, *Mycobacterium massiliensis* and *Arthrobacter* species, detected in the ERV lakes may support both arguments. The distribution of an animal pathogen in all of the hot springs suggest that organisms may be present in a variety of environments at low numbers and be inactive (i.e. *Mycobacterium massiliensis*). However, it may also be possible for an organism to colonise new environments due to the shear number of spores released into the adjacent locations (i.e. *Arthrobacter* species). Nevertheless, the high occurrence of sequences related to alkaliphilic organisms and correlation of certain physico-chemical parameters with bacterial subpopulations suggest that the environment plays a definite role in shaping the community composition.

If 'everything is everywhere' the next logical argument would be to ask why look for novel metabolites in new organisms if the same metabolites exist in an organism which is more easily accessible. The answer to the question lies in the extraction of novel metabolites. If a metabolite exists with the desired qualities, it's simpler to identify the changes in the molecule and model other metabolites to act accordingly, than it is to construct the multitude of possible variations in a single protein in order to obtain the desired quality. The low magnesium and calcium concentrations in both the lake and sediment surrounding each lake make this the ideal environment for the study of such metabolites (i.e. alkaliphilic enzymes) capable of functioning in such a setting. The novel secondary metabolites produced by the actinobacterial isolates might illuminate a possible mechanism by which such activity is possible. This hypothesis has yet to be tested on the novel *Streptomyces* xylanases partially characterized in the current study. Although both indicate high specific activity rates for *Streptomyces* species, the activity is relatively low compared to other phyla.

This study clearly illustrated the value of a multi-faceted approach to metagenomic studies whereby classical culturing paired with culture-independent community analysis can be beneficial in directing the extraction of industrially relevant compounds. Not only does it deliver the sought after metabolite, but identifies the organism and the role each organism might play in the community ecology.

BIBLIOGRAPHY

Aguilera, A. Gomez, F., Lospitao, E. & Amils, R. (2006). A molecular approach to the characterization of the eukaryotic communities of an extreme acidic environment: Methods for DNA extraction and denaturing gradient gel electrophoresis analysis. *Syst Appl Microbiol* 29(1):593-605.

Alemayehu, T., Ayenew, T. & Kebede, S. (2006). Hydrogeochemical and lake level changes in the Ethiopian Rift. *J Hydrol*, 316(1-4): 290-300.

Amann, R.I., Ludwig, W. & Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59(1):143-169.

Anderson, A.S. & Wellington, E.M.H. (2001). The taxonomy of *Streptomyces* and related genera. *Int J Syst Evol Micr* 51:797-814.

Antony-Babu, S. & Goodfellow, M. (2008). Biosystematics of alkaliphilic streptomycetes isolated from seven locations across a beach and dune sand system. *Anton Leeuw J Int G* 94(4): 581-591.

Baas-Becking, L.M.G. (1934). *Geobiologie of inleiding tot de milieukunde*. The Hague: WP van Stockum, NV Zoon: 12-15.

Bartscht, K.I. & Heribert, C. (1999). Evaluation of cell activity and of methods for the cultivation of bacteria from a natural lake community. *FEMS Microbiol Ecol* 28: 249-259.

Bauld, J. (1981). Occurrence of benthic microbial mats in saline lakes. *Hydrobiologia*, 81(1): 87-111.

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

Bautista-Gallego, J., Arroyo-López, F.N., Durán-Quintana, M.C. & Garrido-Fernandez, A. (2008). Individual effects of sodium, potassium, calcium, and magnesium chloride salts on *Lactobacillus pentosus* and *Saccharomyces cerevisiae* growth. *J Food Prot* 71(7): 1412-1421.

Belachew, M. (2010). *Phytoplankton production in lake Arenguade (Hora-hado),Ethiopia*. Unpublished Masters thesis. Ethiopia: University of Addis Ababa.

Beyazova, M. & Lechevalier, M.P. (1993). Taxonomic utility of restriction endonuclease fingerprinting of large DNA fragments from *Streptomyces* strains. *Int J Syst Bacteriol* 43: 674–682.

Bowers, K.J., Mesbah, N. M. & Wiegel, J. (2009). Biodiversity of poly-extremophilic bacteria: Does combining the extremes of high salt, alkaline pH and elevated temperature approach a physico-chemical boundary for life? *Saline Systems* 5(9):1-8.

Brettar, I., Christen, R. & Höfle, M.G. (2003). *Idiomarina baltica* sp. nov., a marine bacterium with a high

optimum growth temperature isolated from surface water of the central Baltic Sea. *Int J Syst Evol Micr* 53(2): 407-413.

Buckley, D.H., Graber, J.R. & Schmidt, T.M. (1998). Phylogenetic analysis of non-thermophilic members of the kingdom Crenarchaeota and their diversity and abundance in soils. *Appl Environ Microbiol* 64: 4333–4339.

Cai, Y., Xue, Q., Chen, Z. & Zhang, R. (2009). Classification and Salt-tolerance of Actinomycetes in the Qinghai Lake Water and Lakeside Saline Soil. *J Sustain Devel* 2(1): 107-110.

Casamayor, E.O., Massana, R., Benlloch, S., Øvreås, L., Díez, B., Goddard, V.J., Gasol, J.M., Joint, I., Rodríguez-Valera, F. & Pedrós-alió, C. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ Microbiol* 4(6): 338-348.

Cavicchioli, R. (2007). General characteristics and important model organisms. Cavicchioli, R., (ed.), *Archaea: molecular and cellular biology.* Wiley-Blackwell: ASM Press: 18-40.

Chater, K.F. & Chandra, G. (2006). The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev* 30: 651-672.

Chen, M.Y., Wu, S.H., Lin, G.H., Lu, C.P., Lin, Y.T., Chang, W.C. & Tsay, S.S. (2004). *Rubrobacter taiwanensis* sp. nov., a novel thermophilic, radiation-resistant species isolated from hot springs. *Int J Syst Evol Micr* 54(1): 1849-55.

Chen, W. & Kuo, T. (1993). A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Research* 21(9):2260.

Chiba, H., Inokoshi, J., Okamoto, M., Asanuma, S., Matsuzaki K., Iwama, M., Mizumoto, K., Tanaka, H., Oheda, M., Fujita, K., Nakashima, H., Shinose, M., Takahashi, Y. & Omura, S. (2001). Actinohivin, a novel anti-HIV protein from an actinomycete that inhibits syncytium formation: Isolation, characterization, and biological activities. *Biochem Bioph Res Co* 282(2): 595-601.

Cho, J.C. & Tiedje, J.M. (2001). Bacterial species determination from DNA – DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* 67: 3677 – 3682.

Claridge, M.F., Dawah, H.A. & Wilson, M.R. (1997). Species: the units of biodiversity, the ideal species concept-and why we can't get it. Claridge, M. F., Dawah, H.A. &. Wilson, M.R. (e.d.), *Species: the units of biodiversity*. Springer: 357-376.

Classen, A.T., Boyle, S.I., Haskins, K.E., Overby, S.T. & Hart, S.C. (2003). Community-level physiological profiles of bacteria and fungi: plate type and incubation temperature influences on contrasting soils. *FEMS Microbiol Ecol* 44(3): 319-328.

Coenye, T. & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* 228(1): 45-49.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cal, J., Hippe, H. & Farrow, J.A.E. (1994). The phylogeny of the genus clostridium: Proposal of five new genera and eleven new species combinations. *Int J Syst Evol Microbiol* 44(4): 812-826.

Colwell, R.R. (1970). Polyphasic taxonomy of bacteria. Iizaka, H., Hasegawa, T., (e.d.) *Culture collections of microorganisms*. University of Tokyo Press, Tokyo: 421–436.

Conn, H.J. (1928). A type of bacteria abundant in productive soils, but apparently lacking in certain soils of low productivity. *In Agric Exp Stn Geneva Bull* 1:138.

Conville, P.S. & Witebsky, F.G. (2007). Analysis of multiple differing copies of the 16S rRNA gene in five clinical isolates and three type strains of *nocardia* species and implications for species assignment. *J Clin Microbiol* 45(4): 1146-1151.

Conville, P.S. & Witebsky, F.G. (2005). Multiple copies of the 16S rRNA gene in *Nocardia nova* isolates and implications for sequence-based identification procedures. *J Clin Microbiol*, 43(6): 2881-2885.

Deepika, L. & Kannabiran, K. (2010). Biosurfactant and heavy metal resistance activity of *Streptomyces* spp. isolated from saltpan soil. *J Pharmacol Toxicol* 1(1): 33-39.

Dees, P.M. & Ghiorse, W.C. (2001). Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA. *FEMS Microbiol Ecol* 35(1): 207-216.

Delgado, O., Quillaguamán, J., Bakhtiar, S., Mattiasson, B., Gessesse, A. & Hatti-Kaul, R. (2006). *Nesterenkonia aethiopica* sp. nov., an alkaliphilic, moderate halophile isolated from an Ethiopian soda lake. *Int J Syst Evol Micr* 56: 1229-32.

De Souza, M.L., Wackett, L.P., & Sadowsky, M.J., (1998). The *atzABC* genes encoding atrizine catabolism are located on a self-transmissible plasmid in *Pseudomonas* sp strain ADP: gene sequence, enzyme purification, and protein characterization. *J Bacteriol* 178: 4894-4900.

Donachie, S.P., Foster, J.S. & Brown, M.V., (2007). Culture clash: challenging the dogma of microbial diversity. *Int Soc Microbiol Ecol* 1:97-102.

Duineveld, B.M., Kowalchuk, G.A., Keijzer, A. & van Elsas, J.D. (2001). Analysis of bacterial communities in the rhizosphere of *chrysanthemum* via denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA as well as DNA fragments coding for 16S rRNA. *Appl Environ Microbiol* 67: 172 – 178.

Duckworth, A.W., Grant, S., Grant, W., Jones, B.E. & Meijer, D. (1998). *Dietzia natronolimnaios* sp. nov., a new member of the genus *Dietzia* isolated from an East African soda lake. *Extremophiles* 2(3): 359-366.

Duckworth, A.W., Grant, W.D., Jones, B.E. & Van Steenbergen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbiol Ecol*,19:181-191.

Dunbar, J., Ticknor, L.O., & Kuske, C.R., (2000). Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. *Appl Environ Microbiol* 66: 2943 – 2950.

Embley, T.M. & Stackebrandt, E. (1994). The molecular phylogency and systematics of the Actinomycetes. *Annu Rev Microbiol*, 48(1): 257-289.

Eppard, M., Krumbein, W.E., Koch, C., Rhiel, E., Staley, J.T. & Stackebrandt, E. (1996). Morphological, physiological, and molecular characterization of actinomycetes isolated from dry soil, rocks, and monument surfaces. *Arch Microbiol* 166(1): 12-22.

Ersts, P.J. (2011). Geographic distance matrix generator (version 1.2.3). USA: American museum of natural history, center of biodiversity and conservation. [online]. Available http://www.*biodiversityinformatics.amnh.org/open_source/gdmg/*

Ferrari, B.C. & Gillings, M.R. (2009). Cultivation of fastidious bacteria by viability staining and micromanipulation in a soil substrate membrane system. *Appl Environ Microb* 75(10): 3352-3354.

Fortina, M.G., Schumann, P., Mora, D., Parini, C., Manachini, P.L., Stackebrandt, E., Tian, C.F., Wang, E.T., Wu, L. J., Han, T.X. & Chen, W.F. (2009). *Ureibacillus* gen. nov., a new genus to accommodate *Bacillus thermosphaericus* (Andersson *et al*. 1995), emendation of *Ureibacillus thermosphaericus* and description of *Ureibacillus terrenus* sp. nov. *Rhizobium fabae* sp. nov., a bacterium that nodulate. *Int. J Syst Evol Microbiol* 59:1258.

Foti, M.J. Schumann, P., Mora, D., Parini, C., Manachini, P.L., Stackebrandt, E., Tian, C.F., Wang, E.T., Wu, L. J., Han, T.X. & Chen, W.F. (2008). Bacterial diversity and activity along a salinity gradient in soda lakes of the Kulunda Steppe (Altai, Russia). *Ambio* 12(1):133-145.

Fry, J.C. (2004). Culture-dependent microbiology. Bull, A.T., (ed), *Microbial diversity and bioprospecting*, ASM Press: 80-87.

Fujiwara, N., Masui, A. & Imanaka, T. (1993). Purification and properties of the highly thermostable alkaline protease from an alkaliphilic and thermophilic *Bacillus* sp. *J Biotechnol*,30: 245-256.

Galatenko, O.A. & Terekhova, L.P. (1990). Isolation of antibiotic-producing actinomycetes from soil samples exposed to UV light. *Antibiot Khimioter* 35: 6-8.

Galinski, E.A. (1995). Osmoadaptation in bacteria. *Adv Microb Physiol* 37: 273-328.

Gelsomino, A., Keijzer-wolters, A.C., Cacco, G. & van Elsas, J.D. (1999). Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J Microbiol Meth* 38:1-15.

Genilloud, O., González, I., Salazar, O., Martín, J., Tormo, J.R. & Vicente, F. (2011). Current approaches to exploit actinomycetes as a source of novel natural products. *J Ind Microbiol Biot* 38(3): 375-89.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., de Peer, Y.V.,Vandamme, P., Thompson, F.L. & Swings, J. (2005). Re-evaluating prokaryotic species. *Nat Rev Micro* 3(9): 733-739.

Gizaw, B. (1996). The origin of high bicarbonate and fluoride concentrations in waters of the Main Ethiopian Rift Valley, East African Rift system. *J Afr Earth Sci* 22(4): 391-402.

Goodfellow, M. & Fiedler, H.P. (2010). A guide to successful bioprospecting: Informed by actinobacterial systematics. *Anton Leeuw J Int G* 98(2): 119-42.

Goodfellow, M., Lonsdale, C., James, A. L. & MacNamara, O. C. (1987b). Rapid biochemical tests for the characterisation of streptomycetes. *FEMS Microbiol Lett* 43: 39–44.

Goodfellow, M. & O'Donnell, A.G. (1993). Roots of bacterial systematics. Goodfellow, M.& O'Donnell, A. G., (e.d.), *Handbook of New Bacterial Systematics*, London: Academic Press: 3–56.

Grant W.D. (2006) Alkaline environments and biodiversity. Gerday, C., Glansdorff, N., (e.d.), *Extremophiles: Encyclopedia of Life Support Systems*, Oxford: Eolss Publishers: 1-20.

Grant, W.D. (1990). Alkaliphiles: ecology, diversity and applications. *FEMS Microbiol Rev* 75: 255-269.

Grant, W.D. (2004). Life at low water activity. *Phil Trans R Soc Lond* 359: 1249-1267.

Green, J. & Bohannan, B.J.M. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol Evol* 21(9):501-507.

Greene, E.A. & Voordouw, G. (2003). Analysis of environmental microbial communities by reverse sample genome probing. *J Microbiol Methods* 53: 211 – 219.

Gremion, F., Chatzinotas, A., Kaufmann, K., Von Sigler, W. & Harms, H. (2004). Impacts of heavy metal contamination and phytoremediation on a microbial community during a twelve-month microcosm experiment. *FEMS Microbiol Ecol* 48:273-283.

Groth, I., Schumann, P., Rainey, F.A., Martin, K., Schuetze, B. & Augsten, K. (1997). *Demetria terragena* gen. nov., sp. nov., a new genus of actinomycetes isolated from compost soil. *Int J Syst Bacteriol* 47(4):1129-1133.

Groth, I., Schumann, P., Schutze, B., Augsten, K. & Stachebrandt, E. (2002). *Knoellia sinensis* gen. nov., sp. nov. and *Knoellia subterranea* sp. nov., two novel actinobacteria isolated from a cave. *Int J Syst Evol Micr* 52: 77-84.

Hayakawa, M., Otoguro, M., Takeuchi, T. & Yamazaki, T.Y.I. (2000). Application of a method incorporating

differential centrifugation for selective isolation of motile actinomycetes in soil and plant litter. *Anton Van Leeuw* 78:171–185.

Hasegawa, T., Takizawa, M. & Tanida, S. (1983). A rapid analysis for chemical grouping of aerobic actinomycetes. *J Gen Appl Microbiol* 29: 319–322.

Head, I.M., Saunders, J.R. & Pickup, R.W. (1998). Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb Ecol* 35(1):1-21.

Herrera, A. & Cockell, C.S. (2007). Exploring microbial diversity in volcanic environments: A review of methods in DNA extraction. *J Microbiol Meth* 70: 1 - 12.

Heuer, H., Krsek, M., Baker, P., Smalla, K. & Wellington, E.M. (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol* 63(8): 3233-3241.

Hill, G.T., Mitkowski, N.A., Aldrich-Wolfe, L., Emele, L.R., Jurkonie, D.D. & Ficke, A. (2000). Methods for assessing the composition and diversity of soil microbial communities. *Appl Soil Ecol* 15: 25-36.

Horan, A.C. & Brodsky, B.C. (1986). *Micromonospora rosaria* sp. nov., nom. rev., the Rosaramicin producer. *Int J Syst Evol Microbiol* 36(3): 478-480.

Horikoshi, K. (1999). Alkaliphiles: Some applications of their products for biotechnology. *Microbiol Mol Biol Rev* 63(4): 735-750.

Horikoshi, K. (2010). *Extremophiles Handbook*. Horikoshi, K., Antranikian, G., Bull, A.T., Robb, F., Stetter, K.O., (e.d.), Springer Verlag.

Hough, D.W. & Danson, M.J. (1999). Extremozymes. *Curr Opin Chem Biol* 3(1): 39-46.

Hubert, C., Shen, Y. & Voordouw, G. (1999). Composition of toluene- degrading microbial communities from soil at different concentrations of toluene. *Appl Environ Microbiol* 65: 3064 – 3070.

Imai, H., Suzuki, K., Morioka, M., Numasaki, Y., Kadota, S., Nagai, K., Sato, T., Iwanami, M. & Saito, T. (1987). Okilactomycin, a novel antibiotic produced by a *Streptomyces* species. *J Antibiot* XL(11): 1475-1987.

Islam, E., Dhal, P.K., Kazy, S.K. & Sar, P. (2011). Molecular analysis of bacterial communities in uranium ores and surrounding soils from Banduhurang open cast uranium mine, India: A comparative study. *J Environ Sci Heal* 46(3):271-280.

Janssen, P.H., Yates, P.S., Grinton, B.E., Taylor, P.M. & Sait, M. (2002). Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Appl Environ Microbiol* 68(5): 2391-2396.

Jiang, H., Dong, H., Zhang, G., Yu, B., Chapman, L.R. & Fields, M.W. (2006). Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern china. *Appl Environ Microbiol* 72(6): 3832-3845.

Jones, B. & Renaut, R.W. (1996). Influence of thermophilic bacteria on calcite and silica precipitation in hot springs with water temperatures above 90°C: Evidence from Kenya and New Zealand. *Can J Earth Sci* 33(1): 72-83.

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

Jones, B.E. & Grant, W.D. (1999). Microbial diversity and ecology of the soda lakes of East Africa. Bell, C. R., Brylinsky M., &. Johnson-Green, J.P., (e.d.), *Microbial Biosystems: New Frontiers*:*Proceedings of the 8th International Symposium for Microbial Ecology.* (2000). pp. 681-687.

Joynt, J., Bischoff, M., Turco, R., Konopka, A. & Nakatsu, C. (2006). Microbial community analysis of soils contaminated with lead, chromium and petroleum hydrocarbons. *Microb Ecol* 51(2):209-219.

Jurgens, G., Lindström, K. & Saano, A. (1997). Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl Environ Microbiol* 63(2): 803-805.

Kassen, R. & Rainey, P.B. (2004). The ecology and genetics of microbial diversity. *Annu Rev Microbiol* 58(1): 207-231.

Kebede, E. (1997). Response of *Spirulina platensis* (= *Arthrospira fusiformis*) from Lake Chitu, Ethiopia, to salinity stress from sodium salts. *J Appl Phycology* 9(1): 551-558.

Kekuda, T.R.P., Shobha, K.S. & Onkarappa, R. (2010). Fascinating diversity and potent biological activities of actinomycete metabolites. *J Pharm Research* 3(2):250-256.

Kevbrin, V.V., Zhilina, T.N., Rainey, F.A. & Zavarzin, G.A. (1998). *Tindallia magadii* gen. nov., sp. nov.: an alkaliphilic anaerobic ammonifier from soda Lake deposits. *Curr Microbiol* 37(1): 94-100.

Khetmalas, M.B., Egger, K.N., Massicotte, H.B., Tackaberry, L.E. & Clapperton, M.J. (2002). Bacterial diversity associated with subalpine fir (*Abies lasiocarpa*) ectomycorrhizae following wildfire and salvage-logging in central British Columbia. *Can J Microbiol* 48(1): 611-625.

Kim, B., Falconer, C., Williams, E. & Goodfellow, M. (2002). *Streptomyces thermocarboxydovorans* sp. nov. and *Streptomyces thermocarboxydus* sp. nov., thermophilic actinomycete isolated from soil two moderately thermophilic carboxydotrophic species from soil. *Int J Syst Evol Microbiol* 52: 889-894.

Kim, B., Sahin, N., Minnikin, D.E., Zakrzewska- Cerwinska, J., Mordarski, M. & Goodfellow, M. (1999). Classification of thermophilic streptomycetes, including the description of *Streptomyces thermoalcalitolerans* sp. nov. *Int J Syst Bacteriol* 49: 7-17.

Kirby, B., Le Roes-Hill, M., Cary, S.C., Burton, S.G., Tuffin, I.M. & Cowan, D.A . (2011). Actinobacterial diversity associated with Antarctic Dry Valley mineral soils. De Bruijn, F.J., (ed.), *Handbook of molecular microbial ecology II: Metagenomics in different habitats*. New Jersey: John Wiley & Sons.

Kirby, R. & Rybicki, E. P. (1986). Enzyme-linked immuno- sorbent assay (ELISA) as a means of taxonomic analysis of Streptomyces and related organisms. *J Gen Microbiol* 132: 1891–1894.

Kirk, J.L. Beaudette, L.A., Hart, M., Moutoglis, P., Klironomos, J.N., Lee, H. & Trevors, J.T. (2004). Methods of studying soil microbial diversity. *J Microbiol Meth* 58(1):169 - 188.

Koeleman, J.G.M., Stoof, J., Biesmans, D.J., Savelkoul, P.H.M. & Vandenbroucke-grauls, C.M.J.E. (1998). Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. *J Clin Microbiol* 36(9): 2522-2529.

Korn-Wendisch, F. & Kutzner, H. J. (1992). The family *Streptomycetaceae*. The Prokaryotes. Balows, A., Trusper, H.G., Dworkin, M., Harder, W. & Schleifer, K.H. (e.d.). New York, Springer: 921-995.

Krebs, C.J. (1989). Ecological Methodology. Addison–Wesley, ISBN 032102-1738, 1989.

Krsek, M. & Wellington, E.M.H. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *J Microbiol Meth* 39(1): 1-16.

Kudo, T., Matsushima, K., Itoh, T., Sasaki, J. & Suzuki, K. (1998). Description of four new species of the genus *Kineosporia*: *Kineosporia succinea* sp. nov., *Kineosporia rhizophila* sp. nov., *Kineosporia rnikuniensis* sp. nov. and *Kineosporia rharnnosa* sp. nov., isolated from plant samples, and amended description. *Int J Syst Bacteriol* 48(1):1245-1256.

Kurapova, A.I., Zenova, G. M., Orleanskii, V. K., Manucharov, A.S. & Norovsuren, Z. (2008). Mesophilic and thermotolerant actinomycetes in strongly heated soils. *Mosc Univ Soil Sci Bull* 63(3): 142-147.

Kurtböke, D. & French, J. (2007) Use of phage battery to investigate the actinofloral layers of termite gut microflora. *J Appl Microbiol* 103: 722-734.

Lacey, J. (1971). *Thermoactinomyces sacchari* sp. nov., a thermophilic actinomycete causing Bagassosis. *J Gen Microbiol* 66: 327-338.

Labeda, D. P. (1992). DNA-DNA hybridization in the systematics of *Streptomyces. Gene* 115: 249–253.

Lam, K.S. (2006). Discovery of novel metabolites from marine actinomycetes. *Curr Opin Microbiol* 9(1): 245-251.

Lechevalier, M.P. & Lechevalier, H. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. *Int J Syst Bacteriol* 20(4):435-443.

Lechevalier, M.P., De Bievre, C. & Lechevalier, H. (1977). Chemotaxonomy of aerobic Actinomycetes: Phospholipid composition. *Bioch Syst Ecol* 5(4): 249-260.

LeTurdu, C., Tiercelin, J.J., Gibert, E.,Travi, Y., Lezzar, K.E., Richert, J.P., Massault, M., Gassee, F., Bonnefille, R., Decobert, M., Gensous, B., Jeudy, V., Tamrat, E., Mohammed, M.U., Martens, K., Atnafu, B., Chernet, T., Williamson, D. & Taieb, M. (1999). The Ziway–Shala lake basin system, Main Ethiopian Rift: Influence of volcanism, tectonics, and climatic forcing on basinformation and sedimentation. *Palaeogeogr Palaeocl* 150: 135–177.

Li, H., Yang, S.Z., Mu, B.Z., Rong, Z.F. & Zhang, J. (2006). Molecular analysis of the bacterial community in a continental high-temperature and water-flooded petroleum reservoir. *FEMS Microbiol Lett*, 257(1): 92-98.

Li, W.J., Zhang, Y.G., Zhang, Y.Q., Tang, S.K., Xu, P., Xu, L.H. & Jiang, C.L. (2005). *Streptomyces sodiiphilus* sp. nov., a novel alkaliphilic actinomycete. *Int J Syst Evol Microbiol* 55(1): 1329-1333.

Litchfield, C.D. & Gillevet, P.M. (2002). Microbial diversity and complexity in hypersaline environments: A preliminary assessment. *J Ind Microbiol Biotech*, 28(1): 48-55.

Lulustyaningati, A., Aminin, N. & Warganegara, F.M. (2008). Culture-independent and culture-dependent approaches on microbial community analysis at Gedongsongo (GS-2) Hot Spring. *Int J Integr Biol* 2(3): 45-152.

Maarit-Niemi, R., Heiskanen, I., Wallenius, K. & Lindstrom, K. (2001). Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *J Microbiol Methods* 45: 155 – 165.

MacDonald, G. & MacDonald, G.M. (2003). Biogeography: space, time and life. MacDonald, G. & MacDonald, G.M., (e.d.), Wiley.

Maheshwari, D.K., Dubey, R.C. & Saravaramurthu, R. (2010). Industrial exploitation of microorganisms Maheshwari, D.K., Dubey, R.C. & Saravaramurthu, R., (e.d.), I.K.International Pvt. Limited.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M. & Tiedje, J.M. (2001). The RDP-II (Ribosomal Database Project). *Sci Tech* 29(1): 173-174.

Malkawi, H.I. & Al-omari, M.N. (2010). Culture-dependent and culture-independent approaches to study the bacterial and archaeal diversity from Jordanian hot springs. *Afr J Microbiol Res* 4(10): 923-932.

Margesin, R. & Schinner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles* 5(1):73-83.

Matsumoto, A., Kasai, H., Matsuo, Y., Omura, S., Shizuri, Y. & Takahashi, Y. (2009). *Ilumatobacter fluminis* gen. nov., sp. nov., a novel actinobacterium isolated from the sediments of an estuary. *J Gen Appl Microbiol*

55: 201–205.

Maturrano, L., Santos, F., Rossello, R. & Anton, J. (2006). Microbial diversity in Maras Salterns, a hypersaline environment in the Peruvian Andes. *Appl Environ Microbiol* 72(6): 3887-3895.

Mañes-Lázaro, R., Ferrer, S., Rodas, A.M., Urdiain, M. & Pardo, I. (2008). *Lactobacillus bobalius* sp. nov., a lactic acid bacterium isolated from Spanish Bobal grape must. *Int J Syst Evol Microbiol* 58: 2699-703.

Mayr, E. (1942). *Systematics and Origin of Species from the viewpoint of a zoologist.* Columbia University Press:216.

Mesbah, N.M. & Wiegel, J. (2008). Life at extreme limits: the anaerobic halophilic alkalithermophiles. *Anna N Y Acad Sci* 1125: 44-57.

Messaoudi, A. & Wagenlehner, F. (2010). Identification of meat-isolated *Enterobacteriaceae* by an in silico-ARDRA approach. *Emir. J. Food Agric* 22(2): 91-102.

Milford, A.D., Achenbach, L.A., Jung, D.O. & 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(11): 4715-24.

Minnikin, D. E., O'Donnell, A. G., Goodfellow, M., Alderson, G.,Athalye, M., Schaal, A. & Parlett, J. H. (1984). An integratedprocedure for the extraction of bacterial isoprenoid quinones and polar lipids. *J Microbiol Meth* 2: 233–241.

Möbius, K. (1877). Die Auster und die Austernwirthschaft. Eine Austernbank ist eine Biocönose oder Lebensgemeinde. Berlin: Verlag von Wiegandt: 72-87.

Moore, E.R.B., Arnscheidt, A., Krüger, A., Strömpl, C. & Mau, M. (1999). Simplified protocols for the preparation of genomic DNA from bacterial cultures. Akkermans, A.D.L., van Elsas, J.D., de Bruijn, F.J., (e.d.) *Molecular Microbial Ecology Manual* 1.01:3-18.

Muyzer, G., De Waal, 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 rRNA. *Appl Environ Microbiol* 59(3): 695-700.

Mwirichia, 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.

Nannipieri, P. & Smalla, K. (2006). Nucleic acids and proteins in soil, Nannipieri, P. & Smalla, K., (e.d.), *Soil Biology.* Springer.

Nemec, A., Musílek, M., Maixnerová, M., De Baere, T., van der Reijden, T.J.K., Vaneechoutte, M. & Dijkshoorn, L. (2009). *Acinetobacter beijerinckii* sp. nov. and *Acinetobacter gyllenbergii* sp. nov., haemolytic organisms isolated from humans. *Int J Syst Evol Microbiol* 59: 118-124.

Niemi, R.M., Heiskanen, I., Wallenius, K. & Lindström, K. (2001). Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *J Microbiol Meth* 45(3): 155-65.

Nieto, N.J., Ventosa, A. & Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. *Microbiol Mol Biol Rev* 62(2): 504-544.

Ninet, B., Monod, M., Emler, S., Pawlowski, J., Metral, C., Rohner, P., Auckenthaler, R. & Hirschel, B. (1996). Two different 16S rRNA genes in a mycobacterial strain. *J Clin Microbiol* 34(10):2531-2536.

Nocker, A., Burr, M. & Camper, A.K. (2007). Genotypic microbial community profiling: A critical technical review. *Microbial Ecol*,54: 276-289.

Nogi, Y., Takami, H. & Horikoshi, K. (2005). Characterization of alkaliphilic *Bacillus* strains used in industry: proposal of five novel species. *Int J Syst Evol Microbiol* 55:2309-2315.

Norovsuren, Z., Oborotov, G., Zenova, G., Aliev, R. & Zvyagintsev, D. (2007). Haloalkaliphilic actinomycetes in soils of Mongolian desert steppes. *Biol Bull* 34(4): 417-422.

Oliver, J.D. (2005). The viable but nonculturable state in bacteria. *J Microbiol* 43: 93-100.

Oren, A. (2002). Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol Ecol* 39:1-7.

Pospisilova, L. & Fasurova, N. (2009). Spectroscopic characteristics of humic acids originated in soils and lignite. *Soil and Water Res* 4:168-175.

Pikuta, E.V., Hoover, R.B., Bej, A.K., Marsic, D., Whitman, W.B., Cleland, D. & Krader, D. (2003). *Desulfonatronum thiodismutans* sp. nov., a novel alkaliphilic, sulfate-reducing bacterium capable of lithoautotrophic growth. *Int J Syst Evol Microbiol* 53(5):1327-1332.

Pozo, C., Rodelas, B. & Gonza, J. (2002). D, L-Hydantoinase activity of an *Ochrobactrum anthropi* strain. *J Appl Microbiol* 92:1028-1034.

Prauser, H. (1970). The Actinomycetes. Prauser, H., (e.d.), *The Actinomycetales*. G. Fischer:13-19.

Preston-Mafham, J., Boddy, L. & Randerson, P.F. (2002). Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - a critique. *FEMS Microbiol Ecol* 42(1):1-14.

Radzom, M., Zeeck, A., Antal, N.& Fiedler, H.P. (2006). Fogacin, a novel cyclic octaketide produced by *Streptomyces* strain Tü 6319. *J Antibiot* 59(5):315-7.

Rappe, M.S. & Giovannoni, S.J. (2003). The uncultured microbial majority. *Annu Rev Microbiol* 57:369-394.

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.

Robe, P., Nalin, R., Capellano, C., Vogel, T.M. & Simonet, P. (2003). Extraction of DNA from soil. *Eur J Soil Biol* 39:183-190.

Rodriguez-Valera, F., Ruiz-Berraquero, F. & Ramos-Cormenzana, A. (1981). Characteristics of the heterotrophic bacterial populations in hypersaline environments of different salt concentrations. *Microb Ecol* 7:235-243.

Rosselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* 25(1): 39-67.

Rosello-Mora, R. & Kampfer, P. (2004). Defining microbial diversity-the species concept for prokaryotic and eukaryotic microorganisms. Bull, A.T., (ed), *Microbial diversity and bioprospecting*, ASM Press: 40-48.

Sabet, S., Diallo, L., Lauren, H., Jung, W. & Dillon, J.G. (2009). Characterization of halophiles isolated from solar salterns in Baja California, Mexico. *Extremophiles* 13(1): 643-656.

Sahin, N., Minnikin, D.E., Zakrzewska-czerwinska, J., Mordarski, M. & Goodfellow, M. (1999). Classification of thermophilic streptomycetes , including the description of *Streptomyces thermoalcalitolerans* sp. nov. *Int J Syst Bacteriol* 49:7-17.

Sambrook, J. & Russell, D.W. (2001). Protocol 24: The inoue method for preparation and transformation of competent *E. coli*: "Ultra-Competent" cells. *Molecular Cloning: A Laboratory Manual 1*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York: 1.112-1.115.

Sanglier, J.J., Whitehead, D., Saddler, G.S., Ferguson, E.V. & Goodfellow, M. (1992). Pyrolysis mass-spectrometry as a method for the classification, identification and selection of actinomycetes. *Gene* 115: 235–242.

Saikia, R. (2008). Taxonomy of the genus *Frankia*. Saikia, R., (e.d.), *Microbial Biotechnology.* New India Publishing Agency:381-387.

Saleem, F., Ahmed S. & Jamil, A. (2008). Isolation of a xylan degrading gene from genomic DNA library of a thermophilic fungus *Chaetomium thermophile* ATCC 28076. *Pak J Bot* 40:1225-1230.

Sanchez-Porro, C., Martin, S., Mellado, E. & Ventosa, A. (2003). Diversity of moderately halophilic bacteria producing extracellular hydrolytic enzymes. *J Appl Microbiol* 94:295–300.

Sato, M., Beppu, T. & Arima, K. (1983). Studies on antibiotics produced at high alkaline pH. *Agric Biol*

Chem 47(9):2019-2027.

Sayeh, R., Birrien, J.L., Alain, K., Barbier, G., Hamdi, M. & Prieur, D. (2010). Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. *Extremophiles* 14: 501-514.

Scholz, H.C. Hubalek, Z., Nesvadbova, J., Tomaso, H., Vergnaud, G., Le Flèche, P., Whatmore, A.M., Al Dahouk, S., Krüger, M., Lodri, S. & Pfeffer, M. (2008). Isolation of *Brucella microti* from soil. *Emerg Infect Dis* 14(8):1316-1317.

Seckbach, J. (2000). Journey to diverse microbial worlds: adaptation to exotic environments. Seckbach, J., (e.d.), *Cellular origin and habitat of extreme environments*. Kluwer Academic:55-187.

Segonds, C., Paute, S. & Chabanon, G. (2003). Use of amplified ribosomal DNA restriction analysis for identification of *Ralstonia* and *Pandoraea* Species: Interest in determination of the respiratory bacterial flora in patients with Cystic Fibrosis. *J Clin Microbiol* 41(7): 3415-3418.

Selyanin, V.V., Oborotov, G.E., Zenova, G.M. & Zvyagintsev, D.G. (2005). Alkaliphilic soil actinomycetes. *Microbiology* 74(6): 729-734.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst Technol* 27:379-423.

Shirling, E.B. & Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *Int J Syst Bacteriol* 16(3):313-340.

Shirling, E.B. & Gottlieb, D. (1968). Cooperative description of type cultures of *Streptomyces* III. *Int J Syst Bacteriol* 18(4):279-392.

Singh, L.S., Mazumder, S. & Bora, T.C. (2009). Optimisation of process parameters for growth and bioactive metabolite produced by a salt-tolerant and alkaliphilic actinomycete , *Streptomyces tanashiensis* strain A2D. *Journal de Mycologie Medicale* 19:225-233.

Sjoling, S. & Cowan, D. (2003). High 16S rDNA bacterial diversity in glacial meltwater lake sediment, Bratina Island, Antarctica. *Extremophiles* 7: 275–282.

Skinner, R.H. & Cundliffe, E. (1980). Resistance to the antibiotics Viomycin and Capreomycin in the *Streptomyces* species which produce them. *J Gen Microbiol* 120(1):95-104.

Smith, J., Tow, L., Stafford, W., Cary, C. & Cowan, D. (2006). Bacterial diversity in three different antarctic cold desert mineral soils. *Microbial Ecol* 51(4): 413-421.

Solingen, P.V., Meijer, D., Kleij, Wilhelmus, A.H., Barnett, C., Bolle, R., Power, S.D. & Jones, B.E. (2001). Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake. *Extremophiles* 5:333-341.

Song, Z., Zhi, X., Li, W., Jiang, Ho., Zhang, C. & Dong, H. (2009). Actinobacterial Diversity in Hot Springs in Tengchong (China), Kamchatka (Russia), and Nevada (USA). *Geomicrobiol J* 26: 256-263.

Sorokin, D.Y., Muyzer, G., Brinkhoff, T., Kuenen, J.G. & Jetten, M.S.M. (1998). Isolation and characterization of a novel facultatively alkaliphilic *Nitrobacter* species, *N. alkalicus* sp. nov. *Arch Microbiol* 170(5): 345-352.

Sorokin, D.Y., Lysenko, A.M., Mityushina, L.L., Tourova, T.P., Jones, B.E., Rainey, F.A., Robertson, L.A. & Kuenen, G. J. (2001). *Thioalkalimicrobium aerophilum* gen. nov., sp. nov. and *Thioalkalimicrobium sibericum* sp. nov., and *Thioalkalivibrio versutus* gen. nov., sp. nov., *Thioalkalivibrio nitratis* sp. nov,. and *Thioalkalivibrio denitrificans* sp. nov., novel obligately alkaliphilic and obligately chemolithoautotrophic sulfur-oxidizing bacteria from soda lakes. *Int J Syst Evol Microbiol* 51:565-580.

Sorokin, D.Y., Jones, B.E. & Kuenen, J.G. (2000). An obligate methylotrophic, methane-oxidizing *Methylomicrobium* species from a highly alkaline environment. *Extremophiles* 4(3):145-55.

Sorokin, D.Y. & Kuenen, J.G. (2005a). Chemolithotrophic haloalkaliphiles from soda lakes. *FEMS Microbiol Ecol* 52(3):287-95.

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

Sorokin, D.Y., Zhilina, T.N., Lysenko, A.N., Tourova, T.P. & Spiridonova, E.M. (2006). Metabolic versatility of haloalkaliphilic bacteria from soda lakes belonging to the *Alkalispirillum–Alkalilimnicola* group. *Extremophiles* 10:213–220.

Spring, S., Elshahed, M.S., Youssef, N.H., Luo, Q., Najar, F.Z., Roe, B.A., Sisk, T.M., Bu, S.I., Hinrichs, K. & Krumholz, L.R. (2007). Phylogenetic and metabolic diversity of Planctomycetes from anaerobic, sulfide- and sulfur-rich Zodletone spring, Oklahoma. *Appl Environ Microbiol* 73(15): 4707-4716.

Stach, J.E.M., Bathe, S., Clapp, J.P. & Burns, R.G. (2001). PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol Eco* 36:139-151.

Stach, J.E.M. Maldonado, L., Ward, A.C., Goodfellow, M. & Bull, A.T. (2003). New primers for the class Actinobacteria: application to marine and terrestrial environments. *Environ Microbiol* 5(10): 828-841.

Stackebrandt, E. & Schumann, P. (2006). Introduction to the Taxonomy of Actinobacteria. *Prokaryotes* 3: 297-321.

Stackebrandt, E. & Goebel, B.M., (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44(4): 846-849.

Stakenborg, T., Vicca, J., Butaye, P., Maes, D., De Baere, T., Verhelst, R., Peeters, J., de Kruif, A., Haesebrouck, F. & Vaneechoutte, M. (2005). Evaluation of amplified rDNA restriction analysis (ARDRA) for the identification of Mycoplasma species. *Infect Dis* 5: 46-56.

Streit, W.R. & Schmitz, R.A. (2004). Metagenomics – the key to the uncultured microbes. *Curr Opin Microbiol* 7: 492-498.

Subramaniyan, S. & Prema, P. (2002). Biotechnology of microbial xylanases: Enzymology, molecular biology and application. *Critical Rev Biotech* 22(1):33-46.

Svensson, B.H., Dubourguier, H.C., Prensier, G. & Zehnder, A.J.B. (1992). *Clostridium quinii* sp. nov., a new saccharolytic anaerobic bacterium isolated from granular sludge. *Arch Microbiol* 157: 97-103.

Taber, W.A. (1960). Evidence for the existence of acid-sensitive actinomycetes in soil. *Can J Microbiol* 6:503–514.

Takeuchi, M., Hamana, K. & Hiraishi, A. (2001). Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium, Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic. *Int J Syst Evol Microbiol* 51: 1405-1417.

Tan, I., Suwanto, A., Suhartono, M.T. & Purwadaria, T. (2001). Characterization of xylanase from a xylanolytic thermophilic bacterium isolated from Gunung Pancar hot spring, West Java. *Biotropia*. 17:1-8.

Tang, S., Tian, X., Zhi, X., Cai, M., Wu, J., Yang, L.L., Xu, L. & Li, W. (2008). *Haloactinospora alba* gen. nov., sp. nov., a halophilic filamentous actinomycete of the family *Nocardiopsaceae*. *Int J Syst Evol Microbiol* 58: 2075-2080.

Tiedje, J.M., Asuming-Brempong, S., Nusslein, K., Marsh, T.L. & Flynn, S.J. (1999). Opening the black box of soil microbial diversity. *Appl Soil Ecol* 13: 109-122.

Torsvik, V., Daae, F.L., Sandaa, R. & Øvreas, L. (1998). Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotech* 64:53 - 62.

Torsvik, V, Goksyr, J. & Daae, F.L. (1990a). High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56(3): 782-787.

Torsvik, V., Salte, K., Srheim, R. & Goksyr, J. (1990b). Comparison of phenotypic diversity and DNA heterogeneity in population of soil bacteria. *Appl Environ Microbiol* 56(3): 776-781.

Le Turdu, C., Tiercelin, J.J., Gibert, E., Travi, Y., Lezzar, K.E., Richert, J.P., Massault, M., Gasse, F., Bonnefille, R., Decobert, M., Gensous, B., Jeudy, V., Tamrat, E., Mohammed, M., Umer Martens, K., Atnafu, B., Chernet, T., Williamson, D. & Taieb, M. (1999). The Ziway-Shala lake basin system, Main Ethiopian Rift: influence of volcanism, tectonics, and climatic forcing on basin formation and sedimentation. *Palaeogeogr Palaeocl* 150:135-177.

Ulukanli, Z. & Digrak, M. (2002). Alkaliphilic micro-organisms and habitats. *Turk J Biol* 26:181-191.

Vaneechoutte, M., Beenhouwer, H.D.E., Claeys, G., Verschraegen, G., De Rouck, A., Paepe, N., Elaichouni, A. & Portaels, F.O. (2000). Identification of *Mycobacterium* Species by Using Amplified Ribosomal DNA Restriction Analysis. *J Clin Microbiol* 31(8): 2061-2065.

Van Elsas, J.D., Mäntynen, V. & Wolters, A.C. (1996). Soil DNA extraction and assessment of the fate of *Mycobacterium chlorophenolicum* strain PCP-1 in different soils by 16S ribosomal RNA gene sequence based most-probable-number PCR and immunofluorescence. *Biol Fertil Soil* 24: 188-195.

Venil, C.K.K. & Lakshmanaperumalsamy, P. (2009). An insightful overview on microbial pigment, Prodigiosin. *Electron J Biol* 5(3): 49-61.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. & Van Sinderen, D. (2007). Genomics of Actinobacteria: Tracing the evolutionary history of an ancient phylum .*Microbiol Mol Biol Rev* 71(3):495-548.

Wang, Y, Zhang, Z. & Ruan, J. (1996). A proposal to transfer *Microbispora bispora* (Lechevalier 1965) to a new genus, *Themobispora* gen. nov., as *Thermobispora bispora* comb. nov. *Int J Syst Bacteriol* 46(4): 933-938.

Wang, Y.U.E. & Zhang, Z. (1997). The Actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. *J Bacteriol* 179(10): 3270-3276.

Wani, A.A., Surakasi, V.P., Siddharth, J., Raghavan, R.G., Patole, M.S., Dilip, R. & 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.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, G.E., Stackebrandt, E., Starr, M.P. & Truper, H.G. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Bacteriol* 37(4): 463-464.

Weisburg, W.G. Barns, S.M., Pelletier, D.A. & Lane, D.J. (1991). 16S Ribosomal DNA amplification for Phylogenetic study. *J Bacteriol* 173(2): 697-703.

Wellington, E.M.H. & Williams, S.T. (1981). Host ranges of phages isolated to *Streptomyces* and other genera. *Zentbl Bakteriol Hyg I Abt Suppl* 11: 93–98.

Whalen, J.K. & Sampedro, L. (2010). Soil ecology and management. Bloem, J., Hopkins, D.W., Benedetti, A., (e.d.), *Microbial methods for assessing soil quality*. CABI: 27-50.

Williams, W.D. (1981). Inland salt lakes: An introduction. *Hydrobiologia* 81: 1-14.

Wintzingerode, F., GÃbel, U.B. & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21(3): 213-229.

Woese, C.R. (2004). A New Biology for a New Century. *Microbiol Mol Biol Rev* 68(2): 173-186.

Woese, C.R. (1987). Bacterial Evolution. *Microbiol Rev* 51(2): 221-271.

Yap, W. H. & Wang, Y. (1999). Molecular cloning and comparative sequence analyses of rRNA operons in *Streptomyces nodosus* ATCC 14899. *Gene* 232:77-85.

Zhang, L. & Xu, Z. (2008). Assessing bacterial diversity in soil. *J Soil Sediment* 8:379-388.

Zhang, Z., Wang, Y. & Ruan, J. (1998). Reclassification of *Thermomonospora* and *Microtetraspora. Int J Syst Bacteriol* 48:411-422.

Zhilina, T.N., Zavarzin, G.A., Rainey, F.A., Pikuta, E.N., Osipov, G.A. & Kostrikina, N A. (1997). *Desulfonatronovibrio hydrogenovorans* gen. nov., sp. nov., an Alkaliphilic, Sulfate-Reducing Bacterium. *Int J Syst Bacteriol* 47(1): 144-149.

Zhilina, T.N., Kostrikina, N.A. & Lysenko, A.M. (1996). *Spirochaeta alkalica* sp. nov., *Spirochaeta aficana* sp. nov., *and Spirochaeta asiatica* sp. nov., Alkaliphilic Anaerobes from the Continental Soda Lakes in Central Asia and the East African Rift. *Genome* 46(1):305-312.

Zhilina, T.N. & Zavarzin, G.A. (1994). Alkaliphilic anaerobic community at pH 10. *Curr Microbiol* 29(2): 109-112.

Zhilina, T.N., Zavarzin, G.A., Detkova, E.N. & Rainey, F.A. (1996). *Natroniella acetigena*, gen. nov. sp. nov., an Extremely Haloalkaliphilic, Homoacetic Bacterium: A New Member of *Haloanaerobiales*. *Curr Microbiol* 32(6): 320-326.

Zhou, J., Bruns, M.A. & Tiedje, J.M. (1996). DNA recovery from soil of diverse composition. *Appl Environ Microbiol* 62: 316-322.

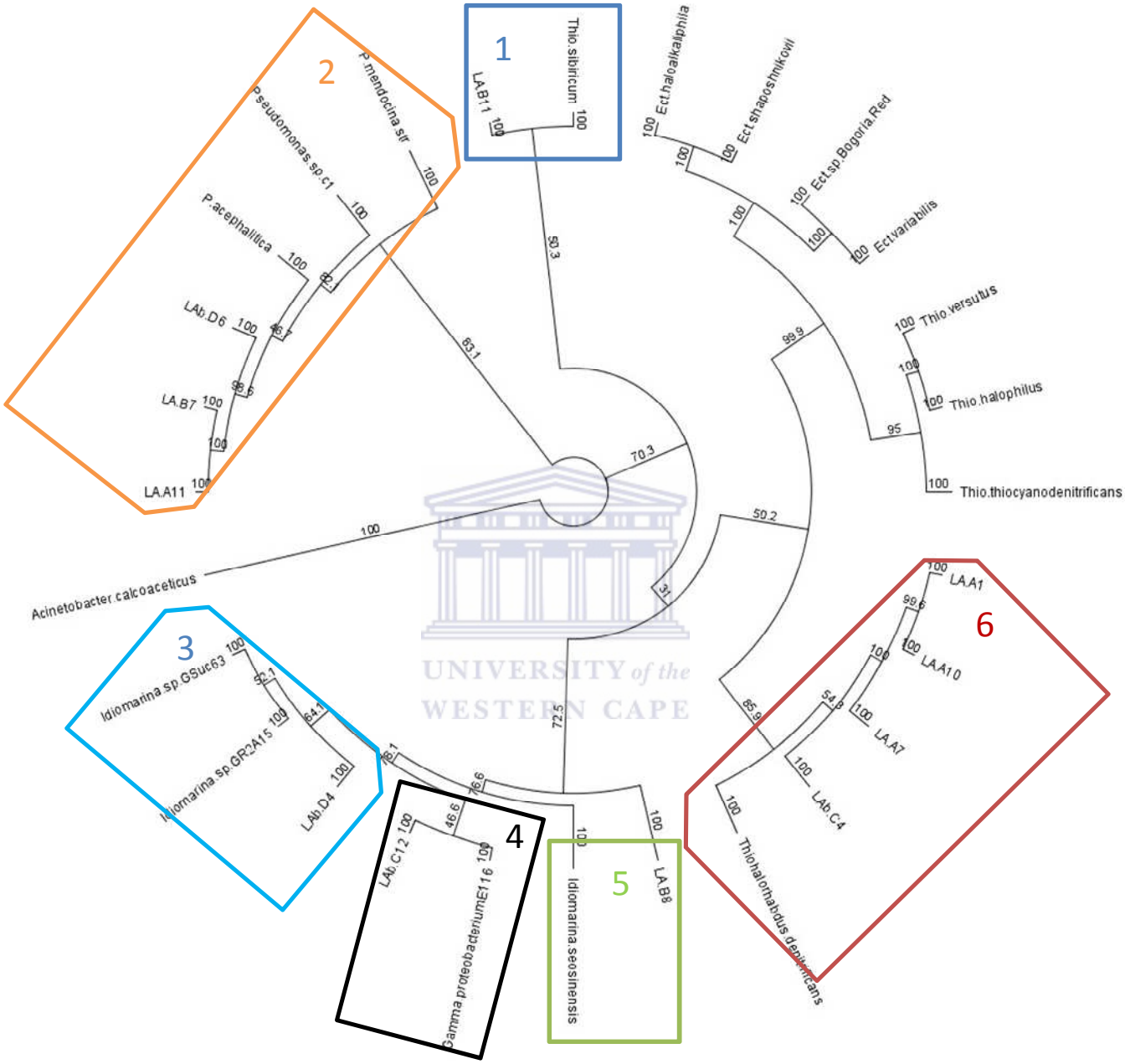
Zinabu, G.M., Kebede-Westhead, E. & Desta, Z. (2002). Long-term changes in chemical features of waters of seven Ethiopian rift-valley lakes. *Hydrobiologia* (477):81-91.

Zuckerkandl, E. & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. *J Theor Biol* 8: 97-166.

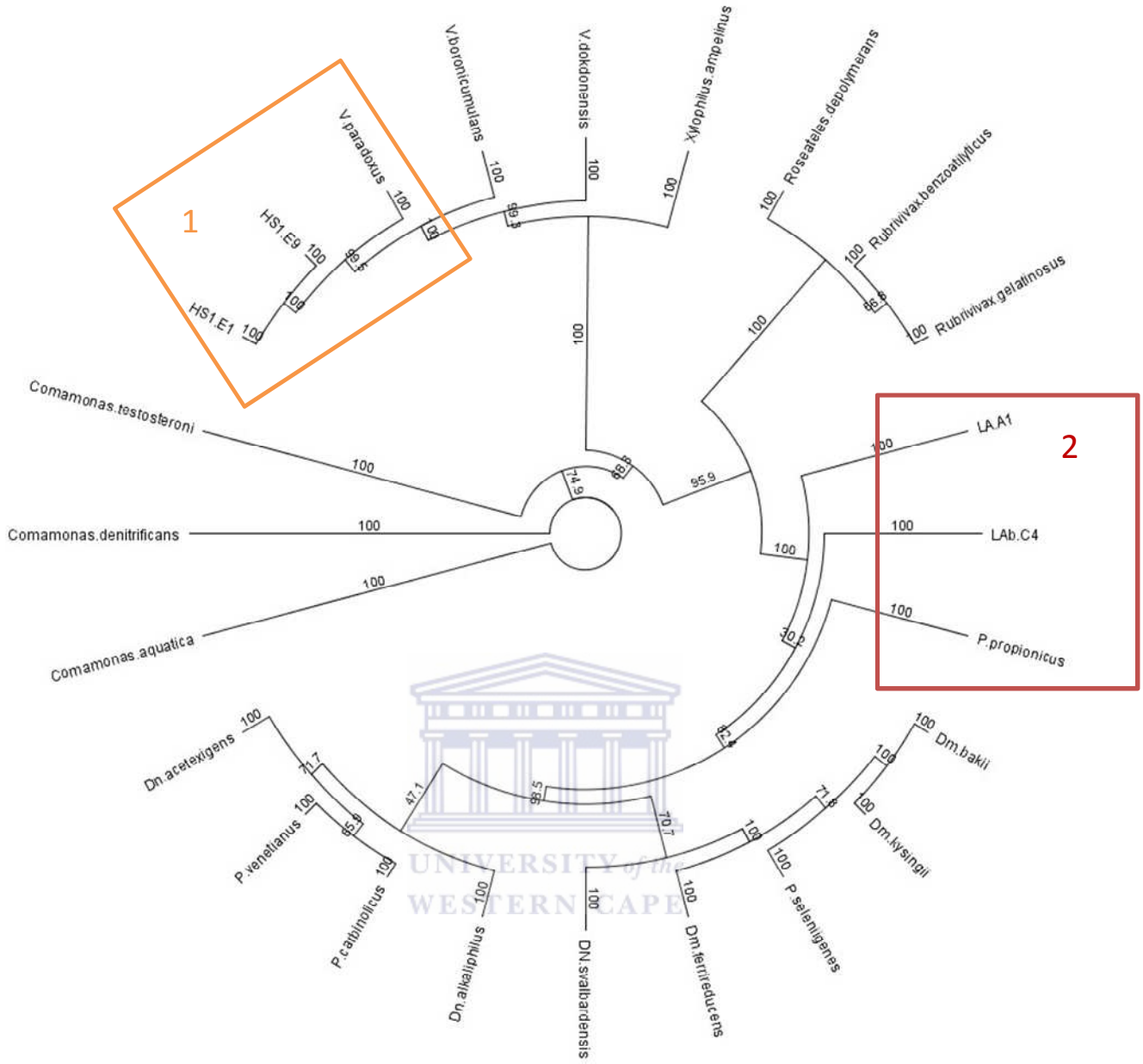
Øvreås, L., Forney, I., Daae, F.I. & Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* 63(9): 3367-73.

Øvreås, L. (2000). Population and community level approaches for analysing microbial diversity in natural environments. *Ecol Lett* (3): 236-251.

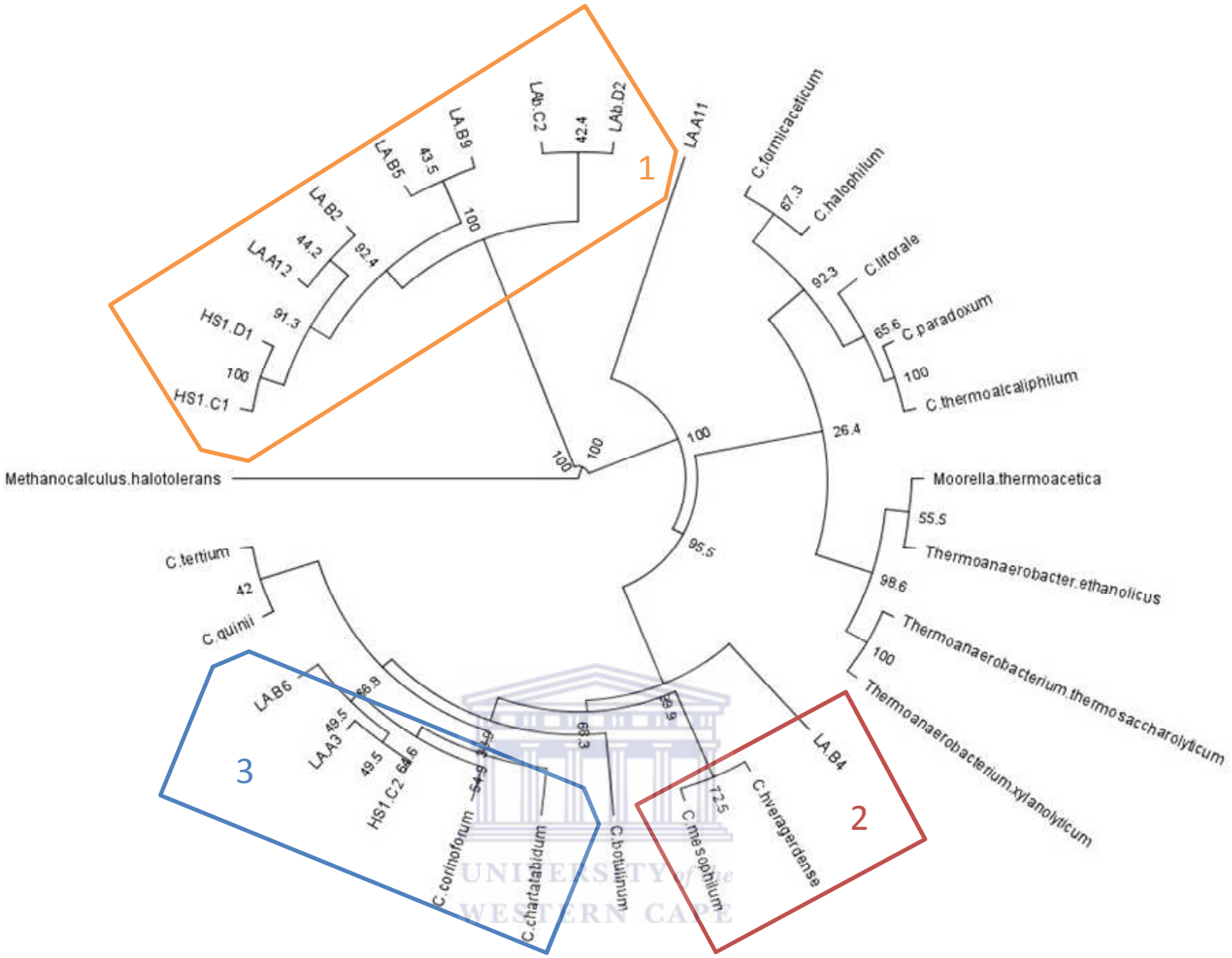
SUPPLEMENTARY MATERIAL



Supplementary Figure 53. A Neighbour-Joining tree depicting the 16S rRNA eubacterial library sequences related to Gamma-Proteobacteria.



Supplementary Figure 54. A Neighbour-Joining tree depicting the 16S rRNA eubacterial library sequences related to Beta- & Delta-Proteobacteria



Supplementary Figure 55. A Neighbour-Joining tree depicting the 16S rRNA eubacterial library sequences related to Clostridia.

Supplementary Table 1

	ERV7		ERV8	
Spore chain morphology	***Rectiflexibiles***		***Spirales***	
Spore-surface				
Production of diffusible pigments	+		-	
Degradation of:	**+TE**	**-TE**	**+TE**	**-TE**
Casein	+	+	+	+
Adenine	-	-	-	-
Hypoxanthine	+	-	+	-
Gelatin	+	+	+	+
Cellulose	-	-	-	-
Starch	+	+	+	+
L-Tyrosine	-	-	-	-
Xylan	++	++	++	++
Tween 80	+	+	+	+
Urea	-	-	-	-
Hippurate	+	-	+	-
Pectin	-	-	-	-
Proteolytic activity	-	-	-	-
Inhibitory compounds				
Phenol	-	-	-	-
Crystal violet	+	-	+	-
Temp range	**28-40**		**28-40**	
NaCl range	**0-10**		**0-10**	
pH4	-		-	
pH range	**7.0- $CaCO_3$ saturation**		**7.0-12.0**	
Antibiotic resistance	**+TE**	**-TE**	**+TE**	**-TE**
Neomycin (50ug/ul)	-	-	-	-
Penicillin G (10 I.U.)	-	-	-	-
Streptomycin (100ug/ul)	+	+	+	+
Ampicillin (100 µg/µl)	+	+	+	+
Chloramphenicol (34 µg/µl)	+	+	+	+
Erythromycin (20µg/µl)	-	-	-	-
Rifampicin (50µg/µl)	-	-	-	-
Lincomycin (100 µg/µl)	-	-	-	-
Gentamicin (100 µg/µl)	-	-	-	-
Tetracycline (10& 50 µg/µl)	-	-	-	-
Kanamycin sulfate (30 µg/µl)	-	-	-	-

+ Weak ++ Good +++ Substantial – None