Metagenomic analysis and characterization of microbial diversity from hydrothermal samples of El Tatio geyser field, Chile

Ncebakazi Galada

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Biotechnology

University of the Western Cape

Supervisor: Prof D. A. Cowan

February 2012

Declaration

I declare that "*Metagenomic analysis and characterization of microbial diversity from hydrothermal samples of El Tatio geyser field, Chile"* 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.

…….

Ms Ncebakazi Galada February 2012

PhD candidate

Abstract

The El Tatio geyser field (ETGF) is the largest known geothermal field in Chile, forming part of a wide spectrum of extremophilic habitats in the country. The ETGF is NaCl rich, with high concentrations of toxic elements such as Li, As and Cs, which are contributed mainly by volcanic activities in the region. Most previous studies in the area have focused on the geology and geochemistry for mining purposes, as well as on the search for geothermal resources for power generation. Very little is currently known about the composition of the microbial communities of the ETGF, which makes the study reported here of particular novelty.

A metagenomic approach, involving the amplification of 16S rRNA gene phylogenetic markers from metagenomic DNA was used to investigate seven different sites within the geyser field. The sample sites were characterized by high temperatures (80-85 °C) and a range of pH values (6.3-8). Various molecular methods, including clone library construction and PCR-DGGE analyses were used to target a wide range of microbial populations within the ETGF sites. Multivariate analysis was also applied to assess differences in the microbial diversity from different sites and to correlate microbial diversity with environmental conditions. Culture-dependent screening of novel nanoarchaeal species was also undertaken. These were coupled with PCR and other detection methods such as fluorescent *in situ* hybridization (FISH) to trace the presence of nanoarchaeal signals from enriched cultures.

The results have shown that the ETGF encompasses a limited microbial diversity represented by only 30 dominant phylotypes, and most likely due to the toxic chemical content of the geyser field. The microbial representatives identified were assigned to OTUs from archaeal, nanoarchaeal and bacterial taxonomic groups. The dominant microbial taxa included members of the Proteobacteria, Firmicutes, Aquificae, Actinobacteria, Euryarchaeota (Halobacteriales, Archaeoglobales), Crenarchaeota (Thermoproteales, Desulfurococcales), together with uncultured representatives of the bacteria, archaea and nanoarchaeota. Notably, representatives of mesophilic, thermophilic and hyperthermophilic taxonomic groups were all detected in ETGF samples. This is attributed to various factors such as temperature gradients and dispersal mechanisms (e.g. natural forces such as rain and volcanic activities). Principal

component analysis (PCA) showed significant differences ($P < 0.05$) in the microbial diversity of the ETGF samples, with principal components (based on the sequenced species from both 16S rRNA clone libraries and PCR-DGGE profiles) explaining up to 62.7% of variance. Furthermore, CCA showed that the differences in phylogenetic diversity were most influenced by temperature and salinity. This was also confirmed by the sequencing results, which showed that hyperthermophilic and haloarchaeal taxa were dominant in the ETGF sites. However, conductivity and pH were also found to contribute to variations in the microbial diversity of the experimental samples, with TDS (total dissolved solids) being a less influential factor. Attempts to generate nanoarchaeal-host co-cultures, and to recover sufficient nanoarchaeal genomic DNA for fosmid and/or large insert cloning for comparative genome analysis, were unsuccessful.

This study is the first to employ metagenomic approaches to analyse the microbial diversity of sites in the ETGF, and has expanded our knowledge of microbiota present in this geyser field.

Dedication

This thesis is dedicated to all the men and women who stood by me through the course of this project, and have tirelessly been my pillars of strength. May God Almighty continue to richly and abundantly bless you always.

Acknowledgements

This work was carried out at the Institute for Microbial Biotechnology and Metagenomics (IMBM), Department of Biotechnology, University of the Western Cape. Electron microscopy and culturing experiments were conducted at Bioscience Foundation laboratory, Santiago, Chile. Flow cytometry was conducted at the University of Cape Town. The Financial assistance was provided by National research Foundation (NRF), The Department of Agriculture, Forestry and Fisheries (DAFF), and the Department of Science and Technology.

With great pleasure I humbly express my sincere gratitude to:

- The Almighty God for giving me a chance and strength to take up this project. Through it all, I have learned to trust in Him and depend upon His Word!
- My supervisor, Professor Donald Cowan for his support, guidance and patience through the course of this project.
- My co-supervisor, Dr. Ana Casanueva for her support through the course of her involvement in this project.
- My host Dr Jenny Blamey at Bioscience Foundation, Santiago, Chile for all the guidance and wisdom and the research team for the warmth and reception rendered. Special thanks go to Liza for all the company, the outings and the shopings ☺ I found a true friend and a sister in you.
- To everyone who has been directly and indirectly involved in this project, you know yourselves, every effort you have made counts and is highly appreciated. Dr Marla Tuffin, Dr Heide Goodman, Dr Mark Taylor, Dr Jean Baptiste, Dr Tsepo Tsekoa, Dr Walter Sanyika, Dr Meshaick Ndou, Ms Nontembeko Dudeni-Tlhone your support and competence was greatly useful and is highly appreciated.
- The research team at IMBM, for the company, support, friendship and the cakes ☺ we shared. Special thanks go to Dr Joseph aka Jeberez Lako and Thulani Makhalanyane aka Toolz, for being there always when needed the most, even when not needed☺. Thanks for the good and the bad times we had, the chats, laughter and the LECTURES! It was all for the good course. I love you, and appreciate you dearly
- A special thanks to my parents Nozolile and Enoch Galada, to the rest of my family and friends, Baba Shelt, Noxy, Tabita, Ta Xi, Shakes, Boyboy, Lulu, Nhonho, Babes, Luba, Gino, Mbali, Ntsako, Kuhle and my little Sino. Thank you for being the pillar of strength and support always, may God richly and abundantly bless you, nizizithandwa zentliziyo yam.
- I can never forget my spiritual family, Rapha Fellowship Centre (RFC) and Rapha Campus Ministry (RCM). Special gratitude goes to my spiritual parents, Bishop Fino and Lady Bishop Nosandi Dlamini for their covering, love and encouragement. To my leaders, my loveolies and my ntshunshuz, Ta Kgaps, sis Jacks, Spakes, Noni, Stu, Anza, Vuyo, Pearls, I don't even know what to say to you, you've been amazing, thanks for all that you've been and made me to become, uBawo anenze ngeNceba.
- Lastly but definitely not least, to my colleagues at NACI, thank you for the encouragement, the storm is over now, no more long shifts! Special gratitude goes to my BOSS soon to be Dr Thiambi aka Mr T the honorable Chief of NACI, your understanding and support is highly appreciated, ndo levhuwa nga manda!

Table of abbreviations

Abbreviation	Definition
AMMI	additive main effects and multiplicative interaction
ANOVA	analysis of variance
ARDRA	amplified ribosomal DNA restriction analysis
APS	ammonium persulphate
As	arsenic
ATP	adenosine triphosphate
Bp	base pairs
BSA	bovine serum albumin
°C	degrees Celsius
Ca	calcium
$CaCl_2$	calcium chloride
CCA	canonical correspondence analysis
Cfu	colony forming units
Cm	centimeter
Cs	caesium
CTAB	cetyl trimethyl ammonium bromide
DAPI	4',6-diamidino-2-phenylindole
df	degree of freedom
DGGE	denaturing gradient gel electrophoresis
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide

Table of abbreviations (continued)

Abbreviation	Definition
ETGF	El Tatio geyser field
FISH	fluorescence *in situ* hybridization
FSC	forward scatter
GLM	general linear models
g/L	grams per litre
Gyr	giga years
H’	Shannon diversity index
HCl	hydrochloric acid
Hrs	hours
IPA	iso-propyl alcohol
IPTG	isopropyl β-D-thiogalactosidase
KCl	potassium chloride
Km	kilometer
km^2	square kilometer
l	litre
LB	Luria Bertani medium
LB-amp	Luria Bertani medium containing ampicillin
LBA	long branch attraction
Li	lithium
m	metre
M	molar
MDA	multiple-strand displacement amplification
MDS	multi-dimensional scaling
mg	milligram

Table of abbreviations (continued)

Abbreviation	Definition
LGT	lateral gene transfer
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
min	minute
ml	milliliter
mm	millimetre
MS	mean of squares
mS/cm	milliSiemens per centimetre
µg	microgram
µl	microlitre
µm	micrometre
µM	micromolar
mM	millimolar
$MnCl_2$	manganese chloride
NaCl	sodium chloride
NaH_2PO_4	sodium phosphate
ng	nanogram
$(NH_4)_2SO_4$	ammonium sulphate
OTU	operational taxonomic unit
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
ppm	parts per million
RCF	relative centrifugal force

Table of abbreviations (continued)

Abbreviation	Definition
Redox	reduction–oxidation
RDA	redundancy analysis
rDNA	ribosomal deoxyribonucleic acid
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
rRNA	ribosomal oxyribonucleic acid
s	second
SAS	statistical analysis system
SD	standard deviation
SDS	sodium dodecyl sulphate
SOM	self-organizing maps
sp.	species
SS	sum of squares
SSC	side scatter
SSU	small subunit
TAE	tris-acetic acid EDTA
TBE	tris-borate EDTA
TDS	total dissolved solids
TE	tris EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TMAC	tetramethylammonium chloride
Tris	tris (hydroxymethyl)-aminomethane
UV	ultraviolet
UPGMA	unweighted pair group method with arithmetic mean
v/v	volume per volume
WGA	whole genome amplification
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase
YNP	Yellowstone National Park

Table of Contents

List of Tables

List of Figures

LIST OF APPENDICES

Chapter 1

Literature Review

1.1 Biodiversity

The term "biological diversity" can be traced back to the 1950s (Gerbilskii & Petrunkevitch, 1955). Biological diversity, which is also used interchangeably with "biodiversity", has been divided into three components, namely, (i) genetic diversity, (ii) species diversity, and (iii) ecological diversity (diversity of community) (Magurran, 2004). It is defined as the variation among living, or biological, organisms from all possible environments including terrestrial, marine and other aquatic systems, and the ecological complexes from which they originate (Heywood, 1995). Magurran (2004) has defined biodiversity as "the variety and abundance of species in a defined unit of study". Diversity may be classified on a number of different scales (Magurran, 1988; Southwood, 1978), such as:

- α-diversity, which is the diversity of species present in an environment.
- β-diversity, which deals with diversity variations from one habitat to the next, and is also viewed as a measure of the extent and rate of change in species richness and distribution along environmental gradients.
- γ-diversity, which is the species richness along a range of habitats in a geographical area.

For the purposes of this study, biodiversity will be referred to as microbial diversity, in the context of α-diversity.

1.2 Factors affecting microbial diversity

Molecular techniques have become an important tool in metagenomics and in the study of microbial diversity, giving insight into a variety of genotypes and species from different environments (Kassen & Rainey, 2004). Microbial communities are highly diverse; for example, a cubic centimetre of soil can harbour 10^9 cells representing about 7 000 different taxa (Kassen & Rainey, 2004). Species diversity may also be influenced to a great extent by changes in environmental conditions such as temperature, pH, depth (Ward & Fraser, 2005), different soil types, moisture content (Buckley & Schmidt, 2001; Schloter *et al.*, 2000) and geographical location (Haubold & Rainey, 1996; Whitaker *et al.*, 2003). Gene dispersal and recombination also play a significant role in microbial diversity. Further investigation of the operational mechanisms involved and their consequences are still required (Cohan, 2002).

The extent of microbial diversity may also be determined by the environmental structure, and be controlled by such factors as parasites, migration and interaction. Just like any other community or population, microorganisms live in association and interaction with one another. Such interaction can be defined as antagonistic or synergistic. Further investigations of microbial habitat complexities and microbial interactions are therefore necessary. With constantly changing environments, one can predict that these create platforms for novel mutations to evolve, in response to changing ecological or environmental conditions.

Consequently, more complex interactions leading to more complex evolutionary lineages could be generated. Therefore, the challenge to the researchers lies in finding the means to incorporate multidisciplinary fields of study that permit investigation of microbial evolution holistically, with the focal points directed at microbial interactions and changes in environmental gradients (Kassen & Rainey, 2004). There are various other factors affecting microbial diversity, which will be discussed briefly in the following sections. These include (i) natural selection, (ii) fitness, (iii) adaptive evolution, and (iv) mutations.

1.2.1 Natural selection

Natural selection, a mechanism of evolutionary change, was first described by Charles Darwin as the ability of the individuals of a population to adapt better to their environment and reproduce (Darwin, 1859). It can also lead to loss of diversity, whereas maintenance of diversity can be achieved through genotype-by-genotype interaction.

1.2.2 Fitness

Fitness is a measure of the ability of one genotype to reproduce relative to other genotypes. Changes in fitness can be influenced by interactions among and between genotypes and their environment. It is also worth noting that molecular variation may not always reflect accurately the quantity of genetic variation in fitness, but may indicate of what the community is composed of. This may be due to a variety of reasons such as primer selection, primer specificity, and the culturing or metagenomic approach used (Kassen & Rainey, 2004).

1.2.3 Adaptive evolution

Adaptive evolution or adaptive radiation is defined as a period in which a single lineage rapidly diversifies to generate multiple niche-specialist types (Kassen & Rainey, 2004). It is also referred to as branching evolution in which the populations become isolated from each other by adapting to different ecological niches and consequently become separate species (Kassen & Rainey, 2004). During adaptive radiation, diversification requires access of the lineage to the niche. Phenotypic modification then occurs, such as an increase in enzyme expression, something that may be limiting due to adaptation to the new environment. The increased enzyme expression enhances growth and consequently a rise in the population (Kassen & Rainey, 2004). This then allows subsequent divergence through modification and refinement of the existing genome.

One of the mechanisms that occur during invasion of the novel niche is gene duplication. For genes involved in nutrient acquisition, gene duplication is favoured under selective growth. When duplication has occurred, one gene copy accumulates mutations while the other maintains its original state. The lineage experiencing gene duplication may therefore take advantage of new and unexploited resources. It is also possible for one of the duplicated gene copies to be lost, or for mutations to occur in both copies and thus destroy complementary functions through a process known as sub-functionalization (Lynch & Force, 2000).

Changes in levels of enzyme expression can also be responsible for the evolution of some mutants, whereas alterations in regulatory genes may serve as a means for introducing microorganisms to novel environments (Treves *et al*., 1998). In these new environments, adaptation is crucial for the survival of the microorganisms, and also depends on the introduction of new mutations.

1.2.4 Mutations

Mutations serve as the ultimate source of genetic variation. In order to understand microbial diversity, it is crucial to first understand the genetic mechanisms underlying the origin of diversity and the effect these mechanisms have within a given ecological context (Kassen & Rainey, 2004). A variety of mechanisms are responsible for the introduction of mutations, which may be either random, directed, deleterious or beneficial (Wright, 2000; Maynard, 1998). Such mechanisms may include replication errors, environmental conditions, cellular events, repair processes, starvation or stress, DNA destabilizing events, etc. (Wright, 2000).

According to neo-Darwinian definition, random or undirected mutations are those not related to metabolic function of the genes and are undirected by specific selective environmental conditions (Wright, 2000). Examples of conditions that result to undirected mutations may include UV radiation, horizontal gene transfer (HGT) and other mobile genetic elements.

The harmful or deleterious mutations are known to decrease the fitness of the organism while beneficial mutations will increase fitness and promote desirable traits. Deleterious mutations

never achieve high frequencies in populations, whereas the beneficial mutations (which are often of lower frequency) enhance adaptive evolution. Despite their rarity, beneficial mutations cause long term adaptations and substantial variation in fitness (Desai & Fisher, 2007). In the absence of new mutations the variance disappears. Thus new mutations are crutial to maintain variation of species or organisms through which natural selection can act.

Thererfore, in order to understand the ecology and diversity of hyperthermophiles, it is important to recognize the influence of the factors affecting diversity as discussed above. The following sections take account of the tools and techniques involved in understanding the dynamics of the ecology and diversity of hyperthermophiles.

1.3 Ecology and diversity of hyperthemophiles

All life forms are divided into three different domains, namely, the Eukarya, Bacteria and Archaea (Figure 1.1) (Woese & Fox, 1977). These three domains and their respective subdivisions have been defined and categorised according to their 16S rRNA and 18S rRNA gene sequences (Woese & Fox, 1977). The quest to understand the composition of these domains has resulted in investigations of microbial life forms in almost every ecological niche accessible, including those considered "extreme" environments from an anthropocentric perspective (Robertson *et al*., 2005). Ecological niches from the tropics to the Arctic and Antarctic, underground mines, deserts, the Dead Sea, hot springs and hydrothermal vents have

all been investigated. For the purposes of this study, only high temperature environments will be reviewed.

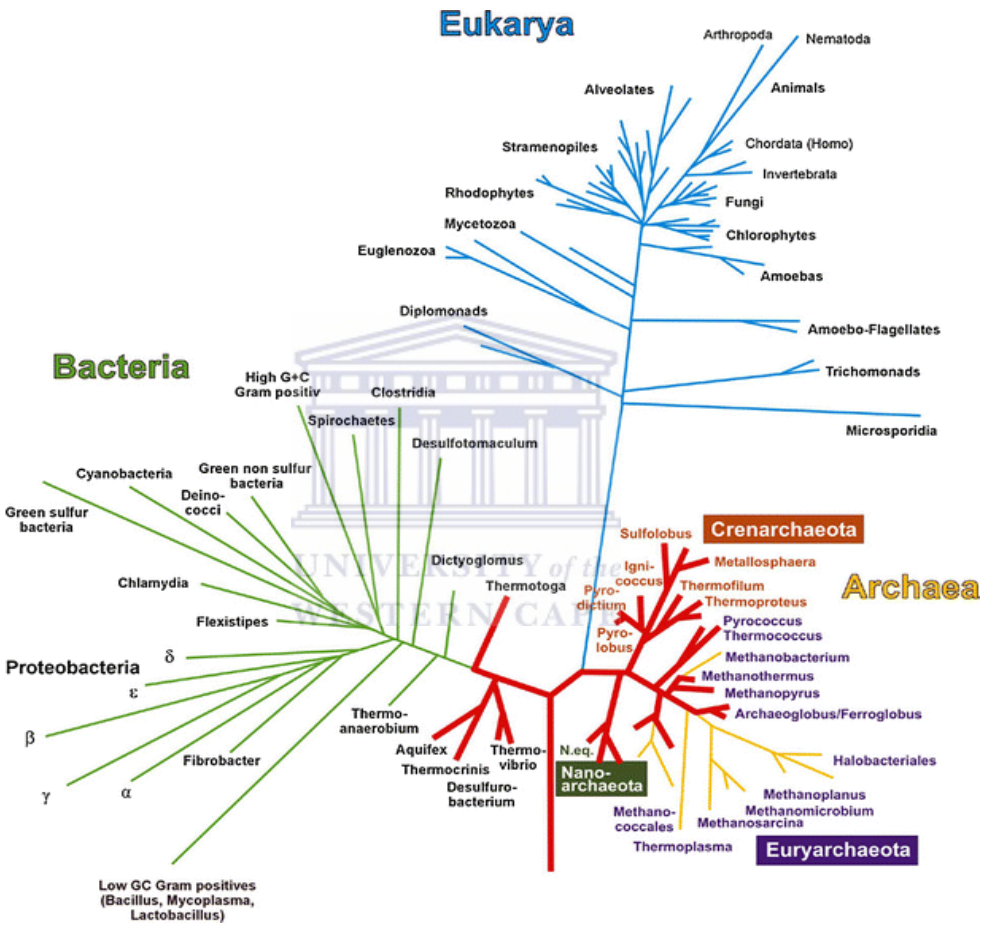

Figure 1.1: A universal phylogenetic tree based on rRNA gene sequences. The lines highlighted in red represent hyperthermophiles (Stetter, 2006b).

From the construction of the universal phylogenetic tree (Figure 1.1), hyperthermophiles are located at a basal position, forming a phylogenetic group distinct from non-hyperthermal subdivisions and thereby indicating a primitive life form (Huber & Stetter, 1998).

1.3.1 The last universal common/cellular ancestor (LUCA)

It is commonly believed that all life forms evolved from a common ancestor. However, the nature of this ancestor remains unknown. Various opinions, theories and hypotheses have been advanced that address this issue, but currently nothing concrete can be affirmed (Ouzounis *et al.*, 2006). It is for this reason that the search for the root of the universal tree is a crucial concern. Subsequently, the last universal common/cellular ancestor (LUCA) has been proposed as the most probable root for the universal tree. LUCA is a hypothetical life form that was presumably a progenitor of the three life domains, i.e. Archaea, Bacteria and Eukarya (Ouzounis *et al.*, 2006). The controversy regarding the nature of LUCA is based on various factors such as its mesophily or (hyper) thermophily. It is not clear whether LUCA had a simple or a complex genome, whether it was prokaryotic or eukaryotic, nor is its monophyly or polyphyly understood (Ouzounis *et al.*, 2006).

Hypotheses that LUCA might have been a hyperthermophile have been disputed, because of the instability of RNA at high temperatures. It is for this reason that a mesophilic or moderately thermophilic nature is favoured for this primordial ancestor (Forterre, 2001). Alternative views have also suggested that LUCA might have been a complex protoeukaryote with an RNA

genome. Such a protoeukaryote is postulated to have been adapted to a broad range of moderate temperatures, being genetically redundant, morphologically and metabolically diverse (Glansdorff *et al.*, 2008). Post the acquisition of an endosymbiosis mechanism, LUCA might have assumed a complex life form embodied as a mesophilic eukaryote, which could have resulted in adaptation to an aerobic atmosphere (Glansdorff *et al.*, 2008). Reductive evolution is also believed to have taken its course, generating a complex and heterogeneous entity more related to prokaryotic domains. Due to these ideologies there has been increasing artifactual topologies, as a result new improved methods for analysis and elimination of these artifacts are required. These would be instrumental in the identification of LUCA and so are potentially important in restructuring the universal tree of life.

1.3.2 High temperature environments

The early life forms on earth are estimated to have been present between 3.5 and 3.9 Gyrs ago, when the planet's surface is thought to have been much hotter than it is today (Stetter, 2006a). After heavy meteorite bombardment, microbes might have been scattered across the planets and moons of the early Solar System. Owing to the high temperature of the earth at the time, only "super-heat-loving" microbes similar to hyperthermophiles would have survived (Stetter, 1992).

High temperature environments have been studied for decades in the search for new and primitive life forms in order to gain insight into the history and evolution of life. These locations include Yellowstone National Park (YNP), with its pristine thermal areas (Barns *et al*., 1994), Rotorua and Tokaanu in New Zealand (Niederberger *et al*., 2006), Iceland (Takacs *et al*., 2001), Japan (Nakagawa & Fukui, 2003), Italy (Kvist *et al*., 2005) and Chile (Casanueva *et al*., 2008). These are some of the major hydrothermal fields around the world and have been the source of a plethora of novel archaea, bacteria and uncharacterized microbial communities.

1.3.2.1 El Tatio geyser field, Chile

The El Tatio geyser field (ETGF) is the largest known geothermal site in Chile, and is located within the Andes Mountains, in Region II of Antofagasta, northern Chile at 4300 m above mean sea level, making it the highest major geothermal field in the world (Landrum *et al.*, 2009; Tassi *et al.*, 2005) (Figure 1.2). It encompasses a combination of thermophilic sites made up of 80 fumaroles, 62 hot springs, 40 geysers, five mud volcanoes and extensive sinter terraces distributed over a 30 km^2 area.

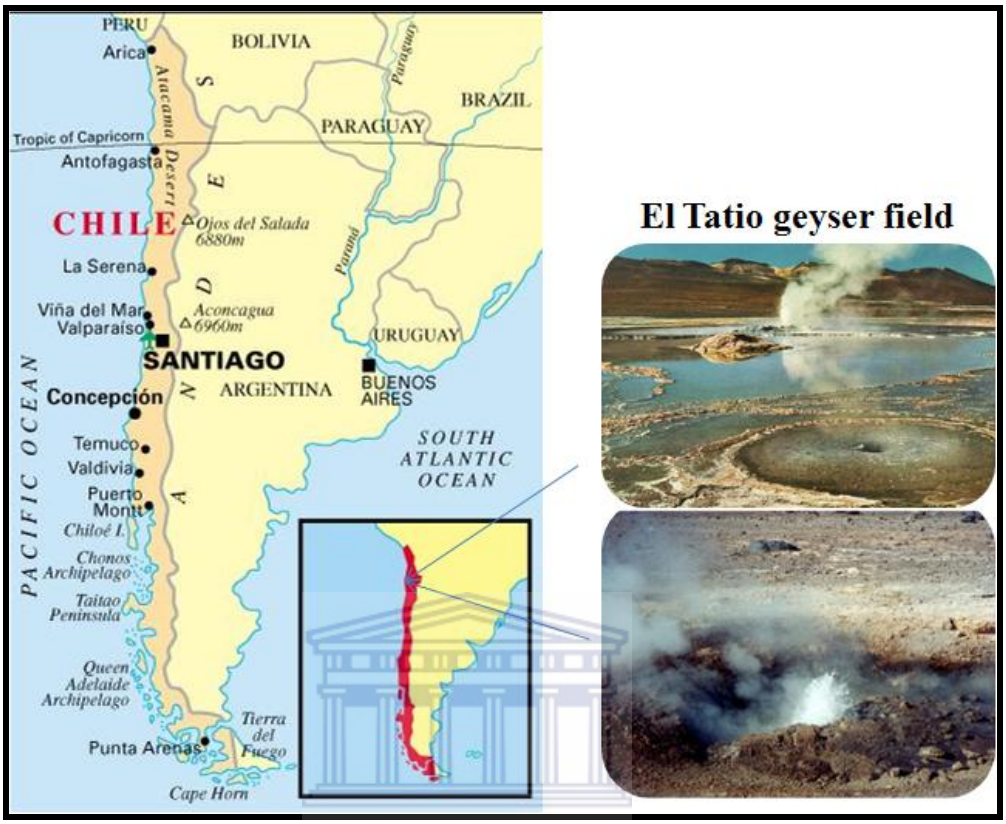

Figure 1.2: Location of El Tatio geyser field in Chile, South America (courtesy of Dr Jenny Blamey).

The ETGF experiences seasonal rainfall, which occurs between December and March, followed by an eight-months-long dry season (Fernandez-Turiela *et al.*, 2005). Evaporation of hot waters results in high concentrations of dissolved solutes such as NaCl, generating supersaturated waters. At their boiling point (86 °C at an altitude of 4300 m), the NaCl-rich waters are ejected in the form of geysers, hot and boiling springs, fumaroles, mud pots, and areas of sinters (Figure 1.3) (Jones & Renaut, 1997).

Figure 1.3: The El Tatio geothermal field (courtesy of Dr Jenny Blamey).

Microbial mats that accumulate in this field have been investigated previously and selective members of microbial communities have been characterised by 16S rRNA gene analysis (Engel, 2008). However, no full determination of the microbial diversity of the ETGF has been reported. In view of lack of such data, the geyser field has become a site of particular interest to investigate and explore the broader microbial communities in the region. The study reported here is based on the metagenomic analysis and characterisation of archaeal, nanoarchaeal and bacterial and 16S rRNA genes using a variety of molecular techniques such as PCR, ARDRA,

FISH, and DGGE. The presence of novel nanoarchaeal species based on culture-dependent methods has also been investigated.

1.3.3 Classification of hyperthermophiles

Hyperthermophiles are classified as those microorganisms that are adapted to grow optimally at temperatures ranging from 70-121 °C (Kashefi & Lovely, 2003). They exhibit different cell morphologies and sizes, which include irregular cocci, lobed or wedge-shaped irregular discs with ultra-flat areas, and regular rods, some with spheres protruding at the ends ("golf clubs"). Their natural biotopes include water-containing volcanic areas such as terrestrial solfataric fields and hot springs, submarine hydrothermal systems, sea mounts and abyssal hot vents (Huber & Stetter, 1991). Some of these thermal environments are man-made, and include smouldering coal refuse piles, geothermal power plants and household hot water geysers (Marsh & Norris, 1985; Fuchs *et al.*, 1995).

A variety of environmental factors are important in maintaining favourable growth conditions, including temperature, pH and salinity. For example, most terrestrial hyperthermophiles require low salinity, whereas those of marine biotopes are adapted to the higher salt concentration of sea water (Stetter, 1999). In some biotopes such as solfataric fields and hot springs, there may be a combination of low salinity and pH values ranging from 0.5 to 8.5. The marine biotopes may have high salinity and relatively narrow pH range (pH 5-8.5).

The upper limit of life has been estimated to lie probably between 140 °C and 150 °C (Cowan, 2004; Stetter, 1998). Although the failure to extend the upper temperature limit well above 120 °C over the past decade argues for a reduced upper limit, evidence for microbial life forms at 250 °C has been reported (Baross & Deming, 1983), even though this has not been accepted by many researchers (Bernhardt *et al*., 1984; Trent *et al*., 1984; White, 1984). Identification of microorganisms that survive beyond the accepted upper temperature limits could therefore be an advanced step in microbial ecology research. Furthermore, this would also enable the development of novel hypotheses of evolution in hot environments (Kashefi, 2004).

1.3.4 The diversity of Archaea

The Archaea consist of two well-described phyla, the Crenarchaeota and Euryarchaeota. Three additional phyla have been proposed, the Korarchaeota (Barns *et al*., 1994), the Nanoarchaeota (Huber *et al.*, 2002) and the Thaumarchaeota (Brochier-Armanet *et al.*, 2008).

1.3.4.1 The Crenarchaeota

The phylum Crenarchaeota represents one of the deep-branching prokaryotic lineages (Perevalova *et al.*, 2008). It is known to compose of phenotypically homogeneous members and sulphur-dependent hyperthermophiles. These hyperthermophiles are widely distributed among terrestrial and marine hot springs, including active volcanic regions (Stetter, 1986). As a result, substantial numbers of crenarchaeal organisms have been reported from geothermal

areas such as Yellowstone National Park, Iceland, Japan, Italy and Thailand (Barns *et al.*, 1996).

Crenarchaeota are composed of four of the hyperthermophilic orders: Sulfolobales, Thermoproteales, Desulphurococcales and 'Caldisphaerales' (Garrity & Holt, 2001; Garrity *et al.*, 2005). Mesophilic crenarchaeotes belonging to the 'marine group-1' have also been reported. These have been shown to be abundant in marine environments, and were initially described as Cenarchaeales or Nitrosopumilales. They have recently been proposed as a separate phylum, the 'Thaumarchaeota' (see Section 1.3.4.5) (Brochier-Armanet *et al.*, 2008; Könneke *et al.*, 2005). The distribution of members of this group is not only limited to low and moderate temperature environments but also to thermophilic settings (de la Torre *et al.*, 2008; Hatzenpichler *et al.*, 2008). In addition, non-thermophilic members of the Crenarchaeota have been reported from soils (Ochsenreiter *et al.*, 2003), freshwater and marine systems (MacGregor *et al.*, 1997), deep-subsurface mines (Inagaki *et al.*, 2001) and deep-subsurface palaeosols (Chandler *et al.*, 1998), terrestrial hot springs (Barns *et al.*, 1996; Kvist *et al.*, 2005), marine hydrothermal vents (Takai & Horikoshi, 1999) and in the oceans (Karner *et al.*, 2001). Various unclassified crenarchaeal phylotypes from mixed environmental samples are also listed in the nucleotide database http://www.ncbi.nlm.nih.gov.

1.3.4.2 The Euryarchaeota

The Euryarchaeota phylum is metabolically and physiologically diverse, and characterised as five functionally distinct groups: the methanogens, extreme halophiles, sulphate reducers, cell

wall-less archaea, and hyperthermophilic sulphur metabolisers (Brochier *et al.*, 2005; Gao & Gupta, 2007; Gribaldo & Brochier-Armanet, 2006).

The phylum is categorised as 11 orders, namely: the Archaeoglobales, Halobacteriales, Methanobacteriales, Methanococcales, Methanomicrobia (Methanomicrobiales, Methanocellales, Methanosarcinales, and unclassified Methanomicrobiales), Methanopyrales, Thermococcales, Thermoplasmatales, and more than 50 genera (http://www.ncbi.nlm.nih.gov). This diverse nature has enabled members of the phylum to be well adapted to different habitats such as deep-sea vents, hot springs, salt lakes and anoxic locations (Pommerville, 2010). For example, euryarchaeal members related to the methanogens, Thermoproteales, Thermococcales and Archaeoglobales have been detected from an active deep-sea chimney of the Monthra vent field, at Juan de Fuca Ridge (Schrenk *et al.*, 2003)

The methanogens are composed of 5 orders: the Methanobacteriales, Methanococcales, Methanopyrales, Methanomicrobiales and Methanosarcinales. These are further subdivided into two main groups, namely, Class I (Methanobacteriales, Methanococcales and Methanopyrales) and Class II (Methanomicrobiales and Methanosarcinales (Bapteste *et al.*, 2005; Fox *et al.*, 1977).

1.3.4.3 Korarchaeota

A number of small subunit (SSU) rRNA gene sequences of organisms from hydrothermal environments form deep-branching phylogenetic lineages that lie outside the crenarchaeote and

euryarchaeote groups. These include members of the Korarchaeota (Barns *et al.*, 1994), the Ancient Archaeal Group (Takai & Horikoshi, 1999), the Marine Hydrothermal Vent Group (Inagaki *et al.*, 2003) and the Nanoarchaeota (Waters *et al.*, 2003).

The phylum Korarchaeota was assigned on the basis of non-culturable species that did not cluster with either the Crenarchaeota or Euryarchaeota (Barns *et al.*, 1996); its members have been detected in several geothermal environments (Auchtung *et al.*, 2006; Barns *et al.*, 1994; Reysenbach *et al.*, 2000; Takai & Horikoshi, 1999). According to the NCBI database (http://www.ncbi.nlm.nih.gov), the Korarchaeota comprise one culturable isolate "*Candidatus Korarchaeum cryptofilum"*, one uncharacterised species, Korarchaeote SRI-306, which belongs to the *Candidatus Korarchaeum* genus, and four phylotypes from environmental samples.

Recently, the distribution and abundance of representatives of the Korarchaeota from 19 hot springs in Hveragerdi and Krysuvik, Iceland, and in Kamchatka, Russia, was investigated within the temperature range 70-97 °C and pH range 2.5-6.5. The analysis of the 16S rRNA gene sequences showed that 12 of the hot springs contained Korarchaeota sequences with > 93.5% similarity to "*Candidatus Korarchaeum cryptofilum*". The overall analysis revealed that the Korarchaeota represented a minor fraction of the microbial community in these hot springs, and up to 7% of an archaeal group (Reigstad *et al.*, 2010).

1.3.4.4 Nanoarchaeota

Nanoarchaeota are one of the most deeply branching phyla reported to date. *Nanoarchaeum equitans* is the only known cultured species representing this phylum (Huber *et al.*, 2002), and was assigned to a new phylum based on its unique features such as small cell size (400 nm in diameter) and a highly reduced genome (490 885 bp). *N. equitans* is a hyperthermophilic, obligate symbiont that grows only in co-culture with a "normal-sized" crenarchaeote, *Ignicoccus hospitalis* (Huber *et al.*, 2003; Jahn *et al.*, 2008). However, there is difficulty in culturing this host-parasite, or host-symbiont, co-culture. This is because, on average, only one out of 50 attempts has produced a stable *N. equitans-I. hospitalis* co-culture (Paper *et al.*, 2007). A specialised form of association that cannot be assigned to mutualism, commensalism, or parasitism has been proposed (Jahn *et al.*, 2008). Two other nanoarchaeal phylotypes have been reported, CU-1 and OP-9, based on environmental rRNA sequences (Hohn *et al.*, 2002). Studies on the nanoarchaeotes have revealed a variation of 16S rRNA gene sequences from diverse geothermal biotopes around the globe (Casanueva *et al.*, 2008; Hohn *et al.*, 2002; McCliment *et al.*, 2006), including mesophilic hypersaline environments (Casanueva *et al.*, 2008). The corresponding 16S rRNA sequences are available from the GenBank database (http://www.ncbi.nlm.nih.gov).

These reports indicate that the diversity of nanoarchaeal 16S rRNA gene sequences retrieved from environmental samples may represent different strains or species which are not necessarily hyperthermophilic. Should it be possible to culture these potential novel species, their genomes could be sequenced and characterized. This action would generate data that

would aid in resolving the phylogenetic placement of the Nanoarchaeota, which remains controversial.

1.3.4.4.1 Genome analysis of *N. equitans*

About 95% of the *N. equitans* genome is predicted to encode proteins or stable RNA, with 552 gene-coding DNA sequences of 827 bp average length (Waters *et al.*, 2003). The genome also lacks conserved operons and has a large number of split genes (Makarova & Koonin, 2005; Di Giulio, 2008). *N. equitans* has little non-coding DNA, few pseudogenes (Waters *et al.*, 2003) and lacks genes for lipid, cofactor and amino-acid biosyntheses. It therefore possesses limited biosynthetic and catabolic capacity (Waters *et al.*, 2003). Compared with bacterial parasites and symbionts, it has been argued that the absence of operons and pseudogenes in nanoarchaeal genomes does not prove an ancestral origin, as these features are known to be present in other prokaryotic genomes, though they may differ from organism to organism (Makarova & Koonin, 2005).

Studies based on small subunit rRNA gene and concatenated ribosomal proteins suggested that *N. equitans* was a fast-evolving euryarchaeal lineage related to the Thermococcales rather than a novel phylum (Brochier *et al.*, 2005). It has also been hypothesised that the genome of *N. equitans* is derived rather than primitive (Brochier *et al.*, 2005; Makarova & Koonin, 2005). Adding to the controversy, further research findings have suggested that this lineage may be a result of genome reduction that was accompanied by rearrangement of the gene order (Di Giulio, 2006). The phylogenetic placement of *N. equitans* as a distinct branch from the

Euryarchaeota and Crenarchaeota has become a controversial issue which does not rule out the possibility of long branch attraction (LBA). LBA occurs when two or more phylogenetically distant species evolve significantly faster than the rest of the taxa (as shown by the long branches), thereby creating artifactual, closely related phylogenies (Lartillot *et al.*, 2007). Furthermore, these long branches may be drawn or attracted to phylogenetically distant outgroups, thereby generating artificial basal lineages (Philippe & Laurent, 1998). Due to a number of uncertainties over nanoarchaeal phylogenetic placement, there is need for extended research on nanoarchaea. As it has been experienced previously, the progress in nanoarchaeal research may be hampered by culturing nanoarchaeal isolates, for which only a 2% success rate in co-culturing attempts has been reported (Paper *et al.*, 2007). More genetic markers such as ribosomal proteins or other functional genes, are required to resolve the phylogenetic placement of this phylum and to improve the understanding of its phylogenetic relationships and evolutionary traits (Brochier-Armanet *et al.*, 2008).

1.3.4.5 Thaumarchaeota

A new phylum of archaea has recently been proposed (Brochier-Armanet *et al.*, 2008). This phylum, termed the Thaumarchaeota (from the Greek word ‘thaumas’ meaning ‘wonder’), consists of mesophilic Crenarchaeota (Brochier-Armanet *et al.*, 2008). The Thaumarchaeota is believed to be one of the ancient archaeal lineages, based on the basal placement of one of its representatives. The phylogenetic arrangement of the Thaumarchaeota places *N. equitans* under

the Euryarchaeota, closely related to Methanopyrales and as a sister group to the hyperthermophilic Crenarchaeota (Brochier-Armanet *et al*., 2008).

1.3.4.6 Microbial diversity in deep marine subsurface habitats

Investigations of microbial diversity in deep marine subsurface habitats have contributed greatly in understanding the ecology of microorganisms in hydrothermal systems (Teske & Sørensen, 2008). Deep marine sediments form one of the most extensive habitats which harbour prokaryotic and eukaryotic life forms, covering more than two-thirds of the earth's surface (Teske & Sørensen, 2008). These marine environments are metabolically active and mostly dominated by archaeal cells.

Depending on their depth, deep subsurface sediments can be characterized by being organic-rich or organic-poor, in which microbial density and activity decrease logarithmically with depth due to the low concentration of organic carbon (D'Hondt *et al.*, 2002, 2004). In these subsurface sediments, archaeal cells are likely to be dominant over bacterial cells, and probably evolved faster to adapt to the energy limitations and extreme conditions (Valentine, 2002). Sulphate occurs throughout the deep marine subsurface sediments as the most highly concentrated electron acceptor. Sulphate reduction, among other anaerobic metabolic activities such as methanogenesis, is one of the most prominent in these settings (D'Hondt *et al*., 2002; Jørgensen, 1982).

Investigation of microbial communities in deep subsurface sediments has increased markedly over the past decade. This research has been motivated by such factors as (i) discoveries of novel lineages, which are mostly unculturable, (ii) the diversity of metabolic activities displayed by these organisms, and (iii) extremophilic physiologies.

The main archaeal phylogenetic lineages that thrive in deep marine sediments include: (i) the Marine Benthic Group B (MBG-B), (ii) the Ancient Archaeal Group (AAG), (iii) the Marine Hydrothermal Vent Group (MHVG), (iv) the Marine Group I (MG-I) Archaea, (v) the South African Goldmine Euryarchaeotal Group (SAGMEG), (vi) the Marine Benthic Groups A and D (MBG-A, MBG-D), (vii) the Miscellaneous Crenarchaeotic Group (MCG), (viii) the Terrestrial Miscellaneous Euryarchaeotal Group (TMEG), (ix) and the Deep Sea Hydrothermal Vent Euryarchaeotal Group 6 (DHVE-6) (Teske & Sørensen, 2008), which will no be discussed in detail.

1.3.4.7 Recent novel microbial species identified from geothermal biotopes

Novel microbial species continue to be discovered from a variety of geothermal biotopes around the globe (Table 1.1). These include the recently proposed *Desulfurococcus kamchatkensis* sp. nov., which was assigned to a novel species based on detailed phylogenetic analysis. *D. kamchatkensis* is an obligate anaerobic chemo-organoheterotroph which uses elemental sulphur as a growth stimulant and an electron acceptor. It grows optimally at 85 °C and pH 6.5 and possesses the unique characteristic of metabolising proteins, including native α-keratin as the sole energy and carbon source (Kublanov *et al.*, 2009b).

A novel species proposed as *Archaeoglobus infectus* sp. nov. was identified from a deep-sea hydrothermal field at Suiyo Seamount in the Izu-Bonin Arc, western Pacific Ocean. *A. infectus* is a strictly anaerobic chemolithoautotroph, utilizing sulphur compounds for respiration, and acetate as a carbon source for growth, suggesting an incomplete CoA pathway for CO_2 fixation (Mori *et al*., 2008).

Table 1.1: Recent novel isolates of the archaeal domain

Isolate/clone	Source environment	Temperature (°C)	Phylum	Reference
Candidatus Korarchaeum cryptofilum'	Obsidian Pool, YNP	85	Korarchaeota	Elkins *et al*., 2008
Desulfurococcus kamchatkensis	Uzon caldera, Kamchatka Peninsula, Russia	85	Crenarchaeota	Kublanov *et al*., 2009a
Nanoarchaeaum equitans	Hydrothermal system at the Kolbeinsey Ridge, north of Iceland	90	Nanoarchaeota	Huber *et al*., 2002
Ignicoccus hospitalis	Kolbeinsey Ridge, North of Iceland	90	Crenarchaeota	Paper *et al*., 2007
Thermococcus celericrescens	Suiyo Seamount, Izu-Bonin Arc, western Pacific Ocean	80	Euryarchaeota	Kuwabara *et al*., 2007
Archaeoglobus infectus sp. nov.	Suiyo Seamount, Izu-Bonin Arc, western Pacific Ocean	70	Euryarchaeota	Mori *et al*., 2008

1.3.5 Bacterial diversity

The Bacteria domain is currently made up of 27 phyla, comprising: 1, Actinobacteria (Gram positive, with high GC content); 2, Aquificae; 3, Bacteroidetes; 4, Chlorobi; 5, Chlamydiae; 6, Lentisphaerae; 7, Verrucomicrobia; 8, Chloroflexi (green non-sulphur bacteria); 9, Chrysiogenetes; 10, Cynobacteria (blue-green algae); 11, Deferribacteres; 12, Deinococcus-Thermus (containing radio-resistant micrococci); 13, Dictyoglomi; 14, Elusimicrobia; 15, Fibrobacteres; 16, Acidobacteria; 17, Firmicutes (Gram positive); 18, Fusobacteria; 19, Gemmatimonadetes; 20, Nitrospirae; 21, Planctomycetes; 22, Proteobacteria (purple bacteria and relatives); 23, Spirochaetes; 24, Synergistetes; 25, Tenericutes; 26, Thermodesulfobacteria; and 27, Thermotoga (http://www.ncbi.nlm.nih.gov).

A wide range of thermophilic and mesophilic bacterial phyla have been identified in hydrothermal pools. These include: Acidobacteria, Bacteriodetes, Nitrospirae, Proteobacteria, Deinococcus-thermus, Plantomycetes, Verrucomicrobia, Actinobacteria, Chloroflexi, and Cyanobacteria (Kanokratana *et al.*, 2004).

Hyperthermophilic bacteria from terrestrial and marine geothermal areas are complex, consisting of a diverse range of chemolithoautotrophic and heterotrophic bacteria with optimum growth temperatures at 80 °C and above (Stetter *et al*., 1989). *Aquifex* and *Thermotoga* are the most hyperthermophilic bacteria known and form part of the deepest phylogenetic lineages within the bacterial domain (Winker & Woese, 1991). The genome sequences of *T. maritima* and *A. aeolicus* revealed 16% and 24% coding sequence similarity to archaeal genes, respectively (Aravind *et al*., 1998; Nelson *et al*., 1999). They also contain

reverse gyrase, a universal component of hyperthermophiles that clusters with other archaeal-like genes (Forterre *et al.*, 2000). These archaeal-like genes are thought to have been a consequence of massive gene transfer between archaeal and bacterial hyperthermophiles through lateral or horizontal gene transfer from Archaea (Forterre *et al.*, 2000*;* Aravind *et al*., 1998; Nelson *et al*., 1999). Evidence from 16S rRNA gene analysis reveals that these genera form part of the deepest-branching lineages, which justifies the proportion of archaeal-like genes in these groups (Kyrpides & Olsen, 1999; Logsdon & Faguy, 1999).

Aquificales are hyperthermophilic bacteria with a temperature limit of 95 °C (Huber & Stetter, Encyclopedia of Life Sciences: *Thermotogales*, http://www.els.net). These Gram-negative, rod-shaped bacteria display optimal growth mainly at neutral pH. Aquificales are chemolithoautotrophs with H_2, thiosulphate or sulphur as electron donors and O_2 as an electron acceptor (Huber *et al.*, 1992; Huber *et al.*, 1998; Stöhr *et al.*, 2001), where H_2S and/or H_2SO_4 are produced in the presence of sulphur (Eder & Huber, 2002). Nitrate is used as an alternative electron acceptor (Huber *et al.*, 1992; Suzuki *et al.*, 2001).

Thermotogales comprise a group of rod-shaped, anaerobic, Gram-negative, non-spore forming thermophilic bacteria. *T. maritima* was the first member to be identified, followed by a series of other discoveries (Huber *et al*., 1986; Belkin *et al*., 1986; Jannasch *et al*., 1988; Huber & Stetter, 1992). As a result, more than 20 different members in more than five different genera are known and have been isolated in diverse biotopes including freshwater and marine hot springs, blacksmokers, and hot oil wells (Van Ooteghem *et al*., 2001). Thermotogales metabolise carbohydrates with H_2 yields that may reach up to 4 mol per mol of glucose

(Takahata *et al.*, 2001; van Niel *et al.*, 2002). Acetic acid and CO_2 are also produced as fermentation products (Huber *et al.*, 1986). Some members have been shown to grow at temperatures between 47-88 °C, although optimal growth is mainly at 80 °C (Stetter *et al.*, 1989).

In addition to *Aquifex* and *Thermotoga*, novel bacterial species from hydrothermal systems continue to be identified around the globe (Table 1.2). For example, a novel bacterial phylum Caldiserica was proposed recently, represented by *Caldisericum exile* gen. nov., sp. nov. *C. exile* was identified from a hot spring at Otari, Nagano Prefecture, Japan, and could not be related to any of the known phyla (Mori *et al.*, 2009). In another study, an anaerobic thermophilic endospore-forming bacterial strain was isolated from a thermal mud at 60-80 °C, located in the Euganean hot springs, Abano Terme, Padova, Italy. The isolate grew optimally at 65 °C and pH 7.2 (Poli *et al.*, 2009). Based on 16S rRNA gene, phenotypic and chemotaxonomic analysis, the isolate was characterized as the type strain belonging to the newly proposed species "*Anoxybacillus thermarum* sp.nov." (Poli *et al.*, 2009).

Recent discoveries from biodiversity studies of hot springs sites at Garhwal in the Indian Himalaya have also revealed a wide range of diverse taxa thriving in these biotopes. These isolates could survive in the temperature range of 40-90 °C and at a pH of 4-11. However, their optimal growth was between 65 °C and 70 °C, and in the pH range 6-8. These organisms were identified as including 11 strains similar to *Geobacillus stereothermophilus*, *G. kaustophilus*, and uncharacterized *Geobacillus* species (Sharma *et al.*, 2009).

Actinobacteria were also found to be widely distributed in hot springs at Tengchong Frog Mouth, China; at Robb Flag hot spring, Kamchatka, Russia; and at the Nevada boiling spring, USA, in a temperature range of 50-81 °C and at pH 6.8-7.9 (Song *et al*., 2009).

Microbial communities in the hydrothermal systems of Vulcano Island, Italy, have also been investigated and found to contain a significant number of metabolically active microorganisms. These include: Thermococcales, *Thermotoga*, *Thermosipho* sp. and *Bacillus* species. Thermophilic members able to oxidize organic acids with O_2, nitrate or sulphate were also identified. Oxic thermophiles such as Aquificales and *Thermus* sp. were also present (Rusch *et al*., 2005).

Caldinitratiruptor microaerophilus gen. nov., sp. nov. is one of the novel bacterial species proposed (Fardeau *et al.*, 2010), belonging to a newly erected genus (Table 1.2). *C. microaerophilus* gen. nov., sp. nov. is a facultative microaerophilic nitrate reducing bacterium isolated from a thermal spring in France and grows optimally at 65 °C and at pH 7. This species was found to be closely related to *Symbiobacterium thermophilum* and *S. toebii*. Based on its unique phylogenetic and physiological characteristics, the bacterium was assigned to a novel genus and species (Fardeau *et al*., 2010).

A study conducted on a terrestrial hot spring in the south-western foothills of the Rincon volcano in Costa Rica has identified an anaerobic chemolithotrophic isolate that uses $S_2O_3^{2-}$ or H_2 as an electron donor and O_2 as an electron acceptor. The 16S rRNA gene analysis revealed the isolate to be closely related to *Thermocrinis ruber*. However, its phylogenetic and

physiological characteristics have caused the isolate to be proposed as a novel species termed *Thermocrinis minervae* sp. nov. (Caldwell *et al*., 2010).

Additional novel bacterial species identified from hydrothermal systems include members belonging to the OP5 candidate phylum. The OP5 phylum was originally proposed on the basis of 16S rRNA gene sequences retrieved from environmental samples collected at YNP (Hugenholtz *et al*., 1998). The inferred phylogenetic relationships revealed representatives of the OP5 phylum to be affiliated independently as a phylum-level lineage, confirming candidate phylum OP5 to be quite different and clearly separate from other known bacterial phyla. These include Firmicutes, Proteobacteria, and Chloroflexi. Analyses of other environmental samples have indicated the presence of OP5 representatives based on 16S rRNA gene analysis (Inagaki *et al*., 2006).

Evidence of a bacterial community isolated at a 306 °C 'black smoker' and cultured at 250 °C has also been reported (Baross & Deming, 1983). However, there have been numerous assertions that this work was not reproducible (reviewed by Prieur *et al.*, 1995). As a result, none of the currently inferred phylogenies surviving at these temperatures can be considered genuine or accurate, and thus no firm phylogenetic conclusions can be drawn from the claims based on this material (Lopez *et al.*, 1999).

Table 1.2: Recent novel isolates of the bacterial domain

Isolate/clone	**Source environment**	**Temperature (°C)**	**Phylum**	**Reference**
Anoxybacillus thermarum sp. nov.	Euganean hot springs, Abano Terme, Italy	65	Firmicutes	Poli *et al*., 2009
Caldisericum exile gen. nov., sp. nov	Otari, Japan	65	Caldiserica phyl. nov. (OP5)	Mori *et al*., 2008
Caldinitratiruptor microaerophilus gen. nov., sp. nov.	Chaudes-Aigues, France	65	Firmicutes	Fardeau *et al*., 2010
Thermocrinis minervae sp. nov.	Rincón, Costa Rica	75	Aquificae	Caldwell *et al*., 2010
Ammonifex thiophilus sp. nov.	Kamchatka, Russia	75	Firmicutes	Miroshnichenko *et al*., 2008

1.4 Investigation of microbial diversity using molecular analysis

Reliable classification of microbial isolates as species and strains is often required in order to analyse and monitor microbial diversity. The emerging use of molecular biology in microbial ecology has revealed that less than 1% of the microorganisms present in known environments are culturable (Amann *et al.*, 1995). The metagenomic-based or culture-independent molecular techniques, such as PCR, cloning, FISH, ARDRA and DGGE (which will be discussed later in this chapter), allow access to a greater portion of unculturable microorganisms than conventional cultivation methods (Robe *et al.*, 2003).

Current molecular techniques rely essentially on the detection and analysis of the small subunit rRNA genes. These molecular identification methods can be grouped into two categories: molecular probes (e.g. FISH) and DNA fragment analysis (e.g. ARDRA). The former group requires prior knowledge of the targeted microbial population. Conversely, the other group does not necessitate prior knowledge of microbial community structure, but requires DNA extraction from the sample and PCR amplification of the template DNA. Irrespective of the approach used, either molecular probes or DNA fingerprinting, these techniques typically detect only the dominant species of microbial communities. Subdominant species, which may represent less than 1% of an ecosystem, may remain undetected, and these may be as important for ecosystem adaptability as their dominant counterparts.

1.4.1 Metagenomics

Microbiology has been based traditionally on the physiological characterisation of microorganisms. Genetics and Molecular Biology were introduced into the field of Microbiology from the 1940s (Dupré & O'Malley, 2007). In the 1990s single-gene studies became a gateway for the investigation of whole genomes. The central role of microbial knowledge in Molecular Biology was intensified mostly with the rapid advancement of sequencing data and genomic studies (Dupré & O'Malley, 2007). Genomic studies of culturable microbial isolates have provided invaluable information and gave insight into comparative evolutionary biology. However, this has generated limited knowledge due to the presence of numerous unculturable microorganisms, which are thought to represent approximately 99 % of all prokaryotes. Subsequently, this challenge has led to the introduction of metagenomic studies (Amann *et al.*, 1995; Nelson, 2003; Ward & Fraser, 2005).

The term 'metagenome' was first defined by Handelsman and colleagues as the collective genome of the total microbiota of a particular environment (Handelsman *et al.*, 1998). Metagenomics, synonymous also with the terms environmental genomics, community genomics, ecogenomics or microbial population genomics, has been defined as the comprehensive study of nucleotide sequences, structure, regulation and function from environmental samples, or a genome-based analysis of entire biological communities from environmental samples (Dupré & O'Malley, 2007). Metagenomics allows access to a large diversity of microorganisms, both culturable and unculturable. This approach takes advantage of both molecular biology and genetics in order to access genetic material of diverse microorganisms from the environment, and extends beyond the sequencing of microbial

isolates to sequencing metagenomic DNA. It expands the scope of data collection to the understanding of biodiversity and evolutionary relationships (Pace, 1997; Xu, 2006). In essence, the primary focus of metagenomics has been on biodiversity, and is based on high-throughput DNA sequencing of clone libraries and environmental samples. Metagenomics gives insight into the vast, previously unknown microbial diversity. The data generated allows the erection of hypotheses on the role of microbial communities in evolution and ecology (Hamilton, 2006). Metagenomics enables understanding of the ecophysiological characteristics of communities, such as their adaptation to environmental conditions, which results in metabolic and morphological changes (DeLong *et al.*, 2006).

Previous genomic studies provided insight into the identification of potential key metabolic pathways and modelling of complex biochemical pathways from relatively comprehensive data sets (Schleper *et al.*, 2005). Metagenomics has also enabled the reconstruction of metabolic pathways from genome sequences by assigning functional roles to different taxa, thereby revealing unanticipated functions and evolutionary mechanisms (Allen & Banfield, 2005; Tringe *et al.*, 2005).

A diversity of sampling environments has been assessed to date. These include but not limited to ocean waters at different temperatures and depths (Béjà *et al.*, 2000; DeLong *et al.*, 2006; Grzymski *et al.*, 2006; Venter *et al.*, 2004), marine sediments (Hallam *et al.*, 2004), air (Holden, 2005) and agricultural soils (Rondon *et al.*, 2000). One of the complex ocean communities analysed thus far is that of the Sargasso Sea (Venter *et al.*, 2004), which has been described as a 'megagenome' due to its large scale sequencing capacity that generated over one billion bp and about 1.2 million genes were reported (Handelsman, 2004).

Metagenomics have not only resulted to great insights for microbial ecology, but have also generated information that can be used in many applications such as drug design, identifying novel genes, protein expression, determining molecular structure and function, predicting metabolic pathways relevant for energy, bioremediation, and numerous other biotechnological applications (DeLong *et al.*, 1999).

1.4.1.1 Metagenome library construction

A variety of vector and host systems have been used successfully to construct metagenome libraries. These include small (plasmid, phage) and large (based mainly on bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), cosmid and fosmid systems) insert libraries. These are screened for enzymatic activity or for specific gene sequences and phylogenetic markers (Riesenfeld *et al.*, 2004). Another more comprehensive approach involves random shotgun sequencing of small insert libraries, whereby individual genomes can be reconstructed from the sequence data (Tyson *et al.*, 2004).

1.4.1.2 Screening the metagenomic library

Advances in sequencing technology have greatly expanded metagenomic approaches, allowing access to single genes, operons, metabolic pathways, individual genomes and communities (Riesenfeld *et al.*, 2004). Different screening methods used include: (i) phenotype screening; (ii) screening for specific phylogenetic markers such as 16S rRNA gene sequences; and (iii) identification of the phylogenetic origin of the screened DNA. The entire genome can then be sequenced and genes of interest identified (Riesenfeld *et al.*, 2004). Specific genes can then be analysed for metabolic activity in order to gain insight into the

roles of community members and the functioning of the microbial community involved (Tringe & Rubin, 2005; Tringe *et al.*, 2005).

1.4.2 Polymerase chain reaction (PCR)

The polymerase chain reaction is a technique invented by Kary B. Mullis in 1983 and later published by Saiki *et al*. (1985). The very first application of PCR for the survey of mixed microbial communities in ocean water led to the discovery of ubiquitous and abundant groups of novel microorganisms (Giovannoni *et al.*, 1990).

PCR is a means by which specific DNA sequences are amplified in a series of repeated cycles. It has been used extensively to amplify gene fragments from pure and mixed microbial communities. Genes derived from amplification products should contain defined fragments that can be cloned and sequenced from known or newly designed priming sites. For PCR to be successful the targeted gene sequence must be complementary to the primer sequence, therefore prior knowledge of the targeted sequence is required to design the primer set to be used. Primers are short synthetic oligonucleotides that hybridize to the complementary DNA sequence. The forward and reverse primers are coupled together and used to amplify a specific DNA sequence in a PCR reaction. It is important that these primers anneal exclusively to the target DNA in order to enhance specificity (Lexa *et al.*, 2001).

A typical PCR reaction includes oligonucleotide primers that anneal to a denatured double-stranded DNA fragment. DNA polymerase then uses the 3′ end of the DNA fragment to extend the newly synthesized strand. The forward and the reverse primers anneal to the 3′- and the 5′- ends of the denatured DNA strand, respectively, such that the product of one

primer becomes the template of another. This process is repeated over a series of cycles until the DNA template has increased exponentially as desired (Saiki *et al.*, 1985). Only microorganisms with the matching primer sequence are able to be amplified and detected, and those that do not have the matching sequence are easily omitted and not represented in the community profile. Effects of single or a few mismatches can be reduced by optimizing the annealing temperature, and by use of degenerate primers (Ishii & Fukui, 2001). The more mismatches, the lower the amplification efficiency (Teske & Sørensen, 2008). A review of bacterial and archaeal primer specificity showed that it would be impossible to obtain a 100% match for a 16S rRNA gene segment among all the prokaryotic taxonomic groups, hence the use of degenerate primers, which play an important role in amplifying broader taxonomic coverage compared to selective primers (Baker *et al.*, 2003). Furthermore, a combination of appropriate DNA extraction methods and the use of specific or degenerate primer sets enable selective amplification of target DNA. These can also be used in conjunction with a series of other molecular tools such as DGGE, ARDRA and sequencing to strengthen the analysis.

1.4.2.1 PCR advantages and drawbacks

PCR has gained its popularity as a molecular identification method based on its rapidness, sensitivity and robustness (Strachan & Read, 1999). Its rapidity stems from its ability to simultaneously amplify multiple gene copies or loci. PCR's sensitivity can be both an advantage and a disadvantage, allowing even very small amounts of DNA to be amplified (Li *et al.*, 1988). Consequently, small amounts of contaminating DNA can also be amplified and thereby generating biased results (Farrelly *et al.*, 1995; Suzuki & Giovannoni, 1996). PCR is also suitable for the identification of specific DNA genes, as it permits selective amplification

of specific target DNA within a pool of heterogeneous DNA such as genomic or metagenomic DNA (Strachan & Read, 1999).

The main disadvantages of PCR include requirements for target sequence information, limiting amounts of product information and DNA replication infidelity (Strachan & Read, 1999). Prior knowledge of the target sequence information is required for the design of suitable primers targeting specific taxonomic groups. Poor primer design can also lead to problems such as non-specificity and poor annealing of primers. Primer annealing efficiency and specificity greatly influence PCR amplification, such that non-specific binding reduces amplification efficiency (Brunk *et al.,* 1996). The choice of primers used is therefore crucial in order to reduce bias. Furthermore, annealing temperature, the number of amplification cycles and the inhibition of polymerase by humic compounds may also contribute to PCR bias. Part of the bias introduced during PCR is due to selective priming, where certain sequences may be discriminated for, or against (Amann *et al.*, 1995). The bias may also be a result of the so-called 'universal' primers, which are allegedly able to amplify almost all known phylogenetic taxa (Potvin & Lovejoy, 2009; Zheng *et al.*, 1996). The more genotypes identified, the more biased the 'universal' primers become and so fail to amplify the DNA of some of the newly discovered taxonomic groups. This implies that the amplifiable sequences may not be a true representation of what exists in the ecosystem, and more microbial representatives may be overlooked and therefore total diversity underestimated due to primer bias. It is thus important to continuously update the existing 'universal' primers in order to incorporate novel sequences. This would reduce amplification bias and improve the representation of the diversity. An example of this is the recently discovered *N. equitans*, which failed to amplify with the archaeal 'universal' primers (Huber *et al.*, 2002).

Consequently, new archaeal primers were designed that could amplify all four archaeal phyla, namely, the Crenarchaeota, Euryarchaeota, Korarchaeota and Nanoarchaeota (Baker *et al.*, 2003).

Regardless of the bias from PCR, this technique remains a crucial tool for the rapid amplification of DNA fragments and for studying microbial ecology. The primers used should be continuously assessed and proper optimization of PCR conditions should be applied in order to obtain optimum results (Dieffenbach & Dveksler, 1995).

1.4.3 DNA cloning systems

Numerous efficient *in vivo* cloning systems have been developed. These allow separation of small and large DNA fragments by cloning into plasmids, lambda phage, cosmid, BAC and YAC systems (O'Connor *et al.*, 1989; Zhang & Wu, 2001). These systems accommodate a variety of insert sizes, depending on the vector system used. For plasmid vectors, approximately 0.5-2 kb inserts can be cloned (small insert libraries) with the upper limit of ~10 kb. Cosmid and fosmid systems can accommodate 35-45 kb inserts, whereas BAC can accept 80-120 kb inserts, with the maximum of ~200 kb. YAC inserts can be 200-800 kb, with a maximum of ~1.5 Mb (Frangeul *et al.*, 1999). These large insert libraries are ideal for studies of genome organization and are more efficient for structural and functional genomic studies.

1.4.4 Amplified ribosomal DNA restriction analysis (ARDRA)

ARDRA is briefly defined as 'restriction analysis of the rRNA gene', also referred to as PCR-RFLP analysis of the rRNA gene. It is a molecular identification tool based on restriction endonuclease digestion of amplified 16S rRNA gene fragments (Schramm & Amann, 1999). ARDRA combines molecular techniques such as DNA extraction, PCR, cloning and restriction digestion of amplified fragments, so enabling the rapid screening of isolates or phylotypes for sequence polymorphisms. The fragments generated are separated according to their sizes using gel electrophoresis; the variance of the patterns is then analysed. Depending on the position of the restriction sites, DNA bands of different numbers and sizes appear and form patterns which can be used rapidly to screen large numbers of isolates. These patterns can be compared with those obtained from a reference database. Alternatively, the DNA of the respective patterns can be sequenced in order to assign isolates to their taxonomic groups or clusters (Stackebrandt & Lang, 1998).

Although the 16S rRNA sequences of species may be almost identical, ARDRA is able to emphasize the few differences present without the need for extensive 16S rRNA sequence analysis (Stakenborg *et al.*, 2005). It should be noted, however, that some base differences may not alter the ARDRA pattern, unless the mutation changes the restriction site. In addition, ARDRA gives little or no information about the type of microorganisms present in the sample, but can be used to assess genotypic changes, including genetic similarities, the genetic structure of phylogenetic or functional groups, and the impact of environmental disturbance on microbial communities (Ranjard *et al.*, 2000).

For ARDRA to be effective, the choice of restriction enzyme is crucial in order to obtain optimum resolution. A typical analysis of the restriction digest of isolates or clones is often

performed on agarose gels at the relatively high concentration of 2.5-3% (w/v) for improved resolution. Resolution can also be improved by using polyacrylamide gels. Owing to the limited size of the 16S rRNA gene (≤1 500 bp), the selection of restriction enzyme should preferably be limited to frequent base cutters such as tetrameric enzymes, in order to generate 5 or 6 fragments on average; this allows manual analysis of ARDRA patterns. However, use of appropriate software such as GelCompar (Applied Maths) and AlphaEase FC imaging analysis (Alpha Innotech) is more efficient and minimizes any error and bias that may be incurred by manual analysis (Vaneechoutte & Heyndrickx, 2001). Selection of restriction enzymes based on computer simulation can also be applied (Heyndrickx *et al.*, 1996).

ARDRA has gained various applications including the characterization of the microbial community structure in different environments (Gich *et al.*, 2000). For example, a variety of ARDRA applications have been reported, these include characterization of more than 300 clones from the metagenomic libraries of YNP hot springs (Hugenholtz *et al.*, 1998). Thereafter, 122 representatives of unique ARDRA patterns were sequenced.

Nevertheless, ARDRA remains an effective technique for identifying phylogenetic groups in highly diverse microbial communities. Furthermore, appropriate isolates can be selected for subsequent physiological and phylogenetic investigations. The development of automated equipment and analysis software would also be advantageous for the effective processing of ARDRA data.

1.4.5 DGGE

DGGE was initially introduced in 1993 and has become a useful genetic fingerprinting technique, which provides genetic diversity profiles within a microbial consortium (Muyzer *et al.*, 1993; Muyzer & Smalla, 1998).

DGGE separates DNA fragments of the same length but with different sequences, based on the reduced migration of partial-melted ds-DNA fragments held together by G-C clamped primers. The G-C clamped primers are made up of a sequence of guanines (G) and cytosines (C) added to the 5´- end of one of the primers. This G-C clamp, which varies from 30-50 nucleotides, prevents the DNA from completely denaturing when subjected to electrophoresis in polyacrylamide gels containing a linear gradient of denaturant (urea and formamide), and thus acts as a high-melting domain (Muyzer *et al.*, 1997). When the stretches of base pairs reach the melting temperature (T_m), the helical structure of the DNA changes to a partially melted form and consequently stop the migration. Optimising the denaturing gradient and duration of electrophoresis is thus crucial for the optimum separation of DNA fragments, thereby giving improved resolution (Muyzer & Smalla, 1998).

1.4.5.1 DGGE applications

DGGE can be used for profiling complex microbial communities, allowing comparison of microbial profiles from different environments (Muyzer *et al.*, 1993; Muyzer *et al.*, 1996). It facilitates the rapid estimation of microbial diversity, where gene fragments (16S rRNA gene in this case) can be separated according to their different sequences, in order to assess the heterogeneity of the analysed genes. This method therefore serves as an ideal tool to monitor changes in microbial community composition.

DGGE is conducted by first extracting genomic or metagenomic DNA from the samples, followed by PCR amplification of the gene(s) of interest (Saiki *et al.*, 1988). The PCR products are then subjected to DGGE, in which a banding pattern is created with the number of DNA bands theoretically corresponding to the number of predominant species in the microbial consortium. In order to obtain more detailed information on the different community members, the DGGE profiles can be blotted onto membranes and hybridized with radioactively labelled oligonucleotide probes (Amann *et al.*, 1992). Alternatively, the DGGE bands can be excised from the gel, reamplified and purified for sequencing (Muyzer & de Waal, 1994).

A variety of environmental samples have been analysed using DGGE (Muyzer *et al.*, 1995). As a result, this technique has become a useful tool for monitoring community behaviour in response to environmental changes. DGGE can also be combined with other techniques, such as the use of microsensors for more reproducible and informative results (Santegoeds *et al.*, 1996). It has also been used to compare the efficacy and reproducibility of different DNA extraction methods (Heuer & Smalla, 1997). It is possible to determine the redundancy and abundance of clone libraries using DGGE, including PCR and cloning biases (Keohavong & Thilly, 1989; Kowalchuk *et al.*, 1997). Knowing the specificity of the primer target regions and the use of a variety of denaturant conditions are crucial in order to achieve optimum results. Furthermore, *in silico* DGGE patterns can be predicted with known sequences from reference databases and comparative analysis conducted to verify the DGGE results.

1.4.5.2 Limitations of DGGE

Various limitations of DGGE analysis are the result of factors such as sample handling errors, PCR and DNA extraction bias (Rochelle *et al.*, 1994). DNA extraction can be a limiting factor due to insufficient lysis of microorganisms in the sample. The use of reliable and reproducible DNA extraction methods is therefore crucial, so that intact nucleic acids are extracted. Removal of PCR inhibitors such as humic acids and ex-polysaccharides is also required in order to generate efficient DGGE results (Alm & Stahl, 1996). Moreover, PCR-generated artefacts such as chimeras, mutations and heteroduplexes can be formed which, when subjected to DGGE, form artificial separation patterns, resulting in false representation of the microbial diversity being investigated (Qiu *et al.*, 2001). These artefacts cause bias due to sequence infidelity and overestimation of the actual community constituents (Myers *et al.*, 1989; Wang & Wang, 1997).

1.4.6 Fluorescence *in situ* hybridization (FISH)

Fluorescence *in situ* hybridization (FISH) is one of the molecular detection tools used in this study. FISH allows simultaneous visualization, identification, enumeration and localization of individual microbial cells through the detection of specific DNA sequences in intact cells. The detection is by fluorescently labelled probes that hybridize specifically to the complementary target sequences within the cells of interest. The principal steps of FISH involve alignment of DNA sequences, identification of conserved regions to design the probes (chemically synthesized single-stranded, short [usually 15-25 nucleotides in length] DNA molecules), synthesis and labelling of complementary nucleic acid probes, and finally the experimental evaluation and optimization of the probe specificity and assay sensitivity.

The procedure includes the following steps: (i) fixation of the specimen; (ii) preparation of the sample, possibly including specimen pre-treatment steps; (iii) hybridization with respective probes for detecting the target sequences; (iv) washing stages to remove unbound probes; and (v) mounting and visualization (Moter & Göbel, 2000).

When choosing which probes to use, it is necessary to consider specificity, sensitivity and ease of cell penetration. The probes are fluorescently labelled specifically to stain different microbial members, with a single fluorescent dye molecule attached to the 5'- end. Commonly used dyes are fluorescein derivatives (excitation wavelength, 490 nm; emission wavelength, 520 nm) such as fluorescein-isothiocyanate (FITC), 5-(-6-)carboxyfluorescein-*N*-hydroxysuccimide-ester (FluoX); rhodamine derivatives (excitation wavelength, 550 nm; emission wavelength, 575 nm), tetramethyl-rhodamine-isothiocyanate (TRITC), and Texas red (excitation wavelength, 578 nm; emission wavelength, 600 nm). Cyanine dyes like Cy3 and Cy5 have been reported to be superior to the classic dyes because they provide significantly brighter staining and are stable to photobleaching (Moter and Göbel, 2000). Aromatic diamidines like 4',6-diamidine-2-phenylindole dihydrochloride (DAPI) bind non-intercalatively to DNA with great affinity and can be used for blue fluorescent counterstaining (Zimmer & Wähnert, 1986).

1.4.6.1 Advantages and disadvantages of FISH

FISH has been widely used for the identification of microorganisms from a variety of sources. Culture-based methods do not reflect the exact composition of mixed microbial communities (Wagner *et al.*, 1993). In comparison, FISH has the advantage of detecting even "non-culturable" microorganisms and can therefore help in understanding complex microbial

communities. It does not require prior extraction of nucleic acids, isolation and enrichment of the microorganisms. Furthermore, FISH permits the rapid, simple and accurate detection of microorganisms.

The pitfalls associated with FISH include permeability associated problems. Penetration of the probe into the cell is a vital step for FISH to be successful. However, cell permeability does not guarantee hybridization of the target sequence (O'Donnell & Head, 1997). Poor accessibility of the target sequence is also possible if the target is in contact with ribosomal proteins. Thus, use of sequences other than those of rRNA genes for phylogenetic identification could improve the efficiency and permeability of the probes (Schönhuber *et al.*, 2001). Furthermore, the sensitivity of FISH increases with the increasing number of active metabolizing cells; and a weak signal is obtained if low cell concentrations or starving cells are used (Amann *et al.*, 1995).

1.4.6.2 Applications of FISH

FISH has been used extensively in the field of microbial ecology. It has been useful in the investigation of microbial communities from natural environments such as aquatic habitats, soils and on root surfaces (Moter & Göbel, 2000). FISH has also been used to demonstrate the relationships between microbial symbionts (Huber *et al.*, 2002). Furthermore, it has found application in the analysis of environments such as bioreactors and treatment plants (Amann *et al.*, 1996).

Like DGGE, the use of FISH can be improved by combination with other techniques, such as microsensors, in order to give additional information on the functional state of the

microorganisms of interest. FISH combined with microautoradiography has enabled the visualization of metabolic activities (Lee *et al.*, 1999). Such techniques have facilitated the identification and analysis of changes in population structure and growth of biofilms, and over time have made it possible to attribute specific metabolic activities to *in situ* microbial populations (Schramm & Amann, 1999).

It is widely accepted that molecular detection methods can overcome most of the limitations of conventional culturing techniques. However, any method used is subject to some degree of bias, and thus FISH is not an exception. It is therefore important to use a combination of detection methods in order to reduce this bias and improve experimental accuracy. In the study reported here, FISH was used with a culture-dependent approach in combination with PCR to trace the presence of Nanoarchaea from the El Tatio geyser field.

1.4.7 The 16S rRNA gene as a phylogenetic marker

The use of gene sequences as molecular markers was first proposed by Zuckerkand and Pauling (1965). The 16S rRNA gene has since became the principal target for this purpose and was first used to define the three domains of life (Woese & Fox, 1977; Woese, 1987). It is now used extensively as a universal signature sequence or phylogenetic marker, for the assessment of microbial diversity in varied source environments (Barns *et al.*, 1996). Its popularity as a molecular marker is based on its universal distribution among prokaryotes, slow evolutionary rate, low susceptibility to lateral or horizontal gene transfer and its large gene size (1.5 kb) (Case *et al.*, 2007). The conserved regions of the 16S rRNA gene serve as primer and probe target sites, and the hypervariable regions are used to infer relationships among phylogenetically distant microorganisms (Case *et al.*, 2007).

Due to a number of factors, such as 16S rRNA gene copy number, DNA extraction and PCR bias, microbial diversity can be inaccurately represented and consequently introduce bias. The 16S rRNA gene copy number can range from 1 to 15 copies per cell, with an average of 4.2 copies which may not be identical; i.e. heterogeneous gene copies (Case *et al.*, 2007). Consequently, the 16S rRNA gene heterogeneity may lead to erroneous taxonomic conclusions, because one cell with multiple 16S rRNA gene copies may falsely indicate the presence of multiple species (Rainey *et al.*, 1996). This has led to the reliability of the 16S rRNA gene as a phylogenetic marker being questioned, especially where closely related taxa are investigated. Consequently, this has labeled the 16S rRNA gene as capable of generation poor taxonomic resolutions (Naum *et al*., 2008). To overcome this problem, single copy genes, such as protein coding genes, have been used as molecular markers. These have been reported to enable accurate community representation and better phylogenetic resolution. For example, the *rpoB* gene has been shown to provide better phylogenetic resolution than the 16S rRNA gene (Case *et al*., 2007). Further examples include the *amoA* gene, which encodes the ammonia monooxygenase enzyme involved in ammonia oxidation (Junier *et al*., 2010), the *hao* gene (hydroxylamine oxidoreductase), the *hzo* gene (hydrazine oxidoreductase) (Schmid *et al*., 2008) and the *accB* gene (accase biotin carboxylase) (Auguet *et al*., 2008). Despite the availability of alternative phylogenetic markers, the 16S rRNA gene continues to be a useful phylogenetic marker in microbial ecology studies, and its use may be complemented by alternative phylogenetic markers rather than being phased out.

1.4.8 Sequencing and phylogenetic analysis

The basic principle of DNA sequencing was established in the 1970s. In the mid-1990s, automated DNA sequencing methods were developed, such as the Sanger method. The Sanger method, developed by Frederick Sanger and co-workers in 1975, is based on chain termination method using dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators (Sanger & Coulson, 1975). It makes use of ssDNA templates, DNA primers, DNA polymerase and radioactive or fluorescently labelled nucleotides, and a modified dideoxynucleotide that terminates DNA strand elongation (Sanger & Coulson, 1975). Before the Sanger method was introduced, other sequencing methods such as the wandering-spot analysis (Gilbert & Maxam, 1973) and chemical-based sequencing methods were used (Maxam & Gilbert, 1977). However, the Sanger method proved more efficient and rapidly replaced these other techniques.

The developments in sequencing technology have generated large amounts of sequencing data, which have resulted in the development of phylogenetic tools such as sequence alignment, the basic local alignment search tool (BLAST) and phylogenetic tree construction (Pop & Salzberg, 2008). The end result of using these tools is to reveal evolutionary relationships of the phylotypes or sequenced species by determining the order of the nucleotides within a gene or the arrangement of genes within a genome. Furthermore, DNA sequencing is instrumental in assigning function to genes and determining their relatedness from one organism to another.

Microbial phylogenetic trees are based on sequences of subunits of macromolecules such as 16S rRNA, which is the basis of the study reported here. The 16S rRNA gene has been used here to illustrate an overview of the diversity of ETGF and the relatedness of the microbial

representatives thriving in this field. It can be sequenced directly or as cloned fragments using standard sequencing techniques as referred to above. Sequencing is followed by aligning the unknown sequence(s) with known sequences retrieved from a reference database such as the Ribosome Database Project (http://rdp.cme.msu.edu/) (Cole *et al.*, 2003) or the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov). Using sequence analysis programs such as Bioedit and MEGA 4 (Tamura *et al.*, 2007), multiple sequence alignment can be achieved with Clustal W (Thompson *et al*., 1994), where sequences can be compared with each other, and the variable and conserved regions identified. The evolutionary relatedness of the aligned sequences can then be determined by construction of a phylogenetic tree, which is based on a variety of analytical methods, such as distance matrix (e.g. Unweighted Pair Group Method with Arithmetic mean (UPGMA) and neighbor-joining method (Saitou & Nei (1987)) and maximum parsimony (Eck & Dayhoff, 1966). The distance matrix determines the evolutionary distance of the compared sequences based on the nucleic acid differences, whereas maximum parsimony relies on the simplest phylogenetic tree that can explain species relatedness. This means that the tree with the fewest assumed mutational events, referred to as the most parsimonious, will then be considered (Staley *et al*., 2007).

Despite the extensive progress made in metagenomics and phylogenetic studies, limitations and challenges continue to exist. However, irrespective of these challenges, the evolving high-throughput sequencing technologies seem to be a promising means to overcome most of the challenges introduced by conventional and clone-based identification methods. Moreover, the advances made in high-throughput sequencing continue to generate much useful

knowledge and have already made noteworthy impact on metagenomics and microbial ecology in large.

1.4.8.1 The impact of high-throughput sequencing on phylogenetics

The advances in metagenomics and phylogenetics have opened up a platform for high-throughput sequencing approaches, such as Roche's 454, Illumina's Solexa and Life Technologies' ABI SOLiD (Sequencing by Oligonucleotide Ligation and Detection) systems. Such methods offer a more rapid and inexpensive sequencing approach than the conventional Sanger approach (Warren *et al*., 2007; Dolan & Denver 2008). The Roche's 454 system is based on the emulsified PCR reaction of DNA fragments attached to individual beads distributed over a 1.6-million-well substrate (PicoTiter plate^{TM}) (Benaglio & Rivolta, 2010). Life Technologies' ABI SOLiD system is more similar to Roche's 454, and also makes use of emulsified beads. However, it differs by the use of its ligase-dependent sequencing, which is based on multiple cycles of hybridization and ligation (Benaglio & Rivolta, 2010). Illumina's Solexa system makes use of bridge amplification, in which amplification is conducted on a glass surface covering the flow cell, using the "reversible terminator" method (Bentley *et al*., 2008). Roche's 454 system prides itself of the relatively large fragment reads, which can range up to 500 nucleotides compared to Illumina's and the SOLiD system's shorter reads of 35-100 and 50-75 nucleotides, respectively (Benaglio & Rivolta, 2010). Furthermore, the SOLiD system has the advantage of internal error correction, based on independent base reading, thereby enabling higher accuracy than the other systems (McKernan *et al*, 2009).

High-throughput sequencing has gained considerable success across the scientific research spectrum. This includes applications in genotyping, single nucleotide polymorphism (SNP)

detection, transcriptome profiling and microbial identification (Petrosino *et al*., 2009; Walter *et al*., 2009). Furthermore, high-throughput sequencing has provided evidence that it can yield much in-depth data coverage compared to conventional sequencing and metagenomic screening (Lazarevic *et al*., 2009). As a result of this robust data generation, the feasibility of data analysis can be achieved only through the development of advanced computational software. Such software needs to be tailor-made to address the specific challenges raised by particular research studies (Schloss, 2009). High-throughput sequencing also offers elimination of pre-cloning procedures, which overcome the obstacles posed by metagenomics (Edwards *et al*., 2006).

Major challenges associated with high-throuput sequencing systems lie with the bioinformatics and computational infrastructure required to process the massive data outputs. Reliable assembling and annotation tools are needed to overcome some of these challenges (Heidelberg *et al*., 2010). The yield of short reads, which can range from 35 to 100 nucleotides, is also one of the challenges limiting high-throughput sequencing. Due to the short read length, hypervariable regions of specific genes such as the rRNA gene have been targeted particularly for microbial identification (Petrosino *et al*., 2009). Other potential pitfalls of this technology includes the increased downstream processing of contings from small fragments, scaffolding and database annotation due to chimeras and putative genes (Lazarevic *et al*., 2009).

It is envisaged that future developments of next generation high-throughput sequencing will improve on longer sequence read lengths, data capturing and annotation software (Huse *et al*., 2007). It could potentially revolutionise our understanding of metagenomics further, allowing

direct and cost-effective sequencing of complex samples, accompanied by refined sequence analysis tools (QingYun &YuHe, 2011).

1.5 Aims and objectives of this study

Understanding the community structure of a given niche can help to explain how microorganisms contribute in shaping the community, and how they interact with each other. A limiting factor in understanding community structures from hyperthermophilic environments has been culturing the hyperthermophiles. To overcome this obstacle, a systematic metagenomic approach, based on cloning of 16S rRNA genes, was used to examine and characterise the composition of microbial communities from hydrothermal sites of the El Tatio geyser field in Chile. Culture-independent and culture-based molecular tools, such as PCR, ARDRA, DGGE, and FISH, were the basis of this study, with the following specific objectives:

- To assess the microbial diversity using a systematic metagenomic approach with a variety of bacterial, archaeal and nanoarchaeal primer set combinations.

- To determine the variability of community structures among the different El Tatio sites using DGGE analysis.

- To estimate and analyse the extent of microbial diversity from both clone libraries and DGGE profiles.

- To investigate the presence of Nanoarchaea in hydrothermal sites using PCR and FISH analysis.

Chapter 2

Materials and Methods

2.1 Sampling

Hydrothermal sediments and soil samples from various El Tatio geyser field (ETGF) sites were collected using a metal sampler at temperatures between 80 °C to 85 °C, and in the pH range 5-8 (Table 2.1). Subsequently, the samples were stored at 4 °C and transported to the laboratory, where they were sub-sampled for immediate use and prolonged storage at -80 °C.

Table 2.1: Environmental conditions of ETGF samples at 4 300 m altitude

Site	Temperature (°C)	Conductivity (mS/cm)	TDS (ppm)	pH	Salinity (g/L)
TAT1-1	83	46.1	960	6.5	4
TAT2-2	80	36.8	989	6.3	2.3
TAT6	85	22.8	1137	6.5	2.1
TAT6-2	85	25.6	1345	8	1.9
TAT9-1	85	44.6	1448	6.5	2.4
TAT9-2	85	38.6	1235	6.4	N/A
TAT10-1	85	58.8	1668.4	7	1.9

N/A: not available

2.2 DNA extraction

Metagenomic DNA from 0.5-1.0 g of the samples was extracted using a modified Zhou method (Stach *et al.*, 2001). Each sample was weighed aseptically into a sterile 2 ml Eppendorf microcentrifuge tube, to which 0.675 ml of freshly prepared extraction buffer [0.1% cetyltrimethylammonium bromide (CTAB); 0.45 ml sterile water; 100 mM Tris-Cl at pH 8; 100 mM NaH_2PO_4; 100 mM ethylenediaminetetraacetic (EDTA); 1.5 M NaCl and 2.5 µl Protease K (20 mg/ml)] was added. The tubes were shaken horizontally at 37 °C for 30 min

at 225 rpm, and 75 µl of 20% sodium dodecyl sulphate (SDS) was added to each tube. The samples were incubated at 65 °C for 2 hrs with gentle inversion every 20 min, followed by centrifugation at 3 000 ×g for 10 min. The supernatants were then transferred to sterile Eppendorf microcentrifuge tubes and equal volumes of phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) were added to each sample, mixed gently and centrifuged at 16 000 ×g for 10 min. The supernatants were transferred to new sterile tubes, and equal volumes of chloroform were added and the samples re-centrifuged at 16 000 ×g for 10 min. The chloroform wash was repeated when the supernatant was deeply coloured, and the supernatants were then transferred to a sterile Eppendorf microcentrifuge tube and 0.6 volumes of iso-propyl alcohol (IPA) added. Samples were incubated overnight at room temperature and centrifuged at 10 000 ×g for 10 min. The DNA pellets were washed in 70% ethanol and re-centrifuged at 10 000 ×g for 10 min. The supernatants were then discarded and the DNA pellets air-dried in a laminar-flow cabinet. The dry DNA pellets were resuspended in 40 µl sterile H_2O and quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, USA).

2.3 Amplification of archaeal, nanoarchaeal and bacterial 16S rRNA genes

The polymerase chain reaction (PCR) was conducted on metagenomic DNA samples using archaeal, nanoarchaeal and bacterial targeting primer sets as listed in Table 2.2. Template DNA (10-100 ng) was amplified using 0.5 µM primers with 200 µM deoxynucleoside triphosphates (dNTPs), 1 mg/ml bovine serum albumin (BSA), 3.5 mM $MgCl_2$ and 1 U of Taq DNA polymerase (Fermentas) in a total reaction volume of 25 µl. This was used as a

standard reaction, unless otherwise specified. Amplification parameters were as described in Table 2.2.

2.4 Electrophoresis conditions

PCR products were electrophoresed against a DNA marker using 1% agarose gels stained with 0.5 µg/ml ethidium bromide. Electrophoresis was performed in 0.5 × TAE (40 mM TRIS base (w/v), 0.2 mM glacial acetic acid (w/v), 10 mM EDTA (w/v), pH 8.0) buffer at 100 V, and visualized under ultraviolet (UV) light using the Alphaimager 3400 transilluminator imaging system (AlphaInnotech Corporation, San Leandro, CA). All electrophoresis procedures were conducted in this manner, unless otherwise specified.

Table 2.2: 16S rRNA gene and vector primer sequences used in PCR amplification

Primer set	Sequence (5' to 3')	Target	Specificity	PCR cycling parameters	Reference
A571Fb UA1204R	GCY TAA AGS RIC CGT AGC TTM GGG GCA TRC IKA CCT	16S rRNA	"Universal" archaeal	94 °C for 4 min; 30 cycles at 94 °C for 1 min, 56 °C for 30 s and 72 °C for 1 min; 72 °C for 10 min.	(Baker *et al.*, 2003)
A3Fa Ab927R	TCCGGTTGATCCYGCCGG CCCGCCAATTCCTTTAAGTTTC	16S rRNA	Archaea except Korarchaea and Nanoarchaea	94 °C for 4 min; 25 cycles at 94 °C for 45 s, 55 °C for 45 s and 72 °C for 1 min; then 10 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min; 72 °C for 20 min.	(Baker *et al.*, 2003) (Jurgens *et al.*, 1997)
E9F U1510R	GAGTTTGATCCTGGCTCAG GGTTACCTTGTTACGACTT	16S rRNA	"Universal" for bacteria	94 °C for 4 min; 30 cycles at 94 °C for 1 min, 55 °C for 30 s and 72 °C for 1 min; 72 °C for 10 min.	(Hansen *et al.*, 1998) (Reysenbach and Pace, 1995)
N989R	GGTTTCCGGTGTCAGTTC	16S rRNA	Nanoarchaea (coupled with A571Fb)	94 °C for 4 min; 30 cycles at 94 °C for 1 min, 55/56 °C for 30 s and 72 °C for 1 min; 72 °C for 10 min.	(Casanueva *et al.*, 2008)
M13F M13R	GTA AAC GAC GGC CAG T CAC ACA GGA AAC AGC TAT GAC CAT	M13 vector sequence		94 °C for 4 min; 30 cycles at 94 °C for 1 min, 50 °C for 30 s and 72 °C for 1 min; 72 °C for 10 min.	(Yanisch-Perron *et al.*, 1985)

Primer numbering corresponds to *E. coli* 16S rRNA sequence numbering.
M13 primer numbering corresponds to pGEM®-T Easy vector.
Degenerate bases represented by R = A or G; S = C or G; Y = C or T; K = G or T.

2.5 Denaturing gradient gel electrophoresis (DGGE)

PCR-DGGE was conducted with first-round and second-round GC-clamped DGGE primers as listed in Table 2.3. All DGGE separations were performed in 9% (w/v) polyacrylamide gels (acrylamide:N,N'-methylene bisacrylamide, 37.5:1 (w/w)] in 1 × TAE buffer [40 mM Tris base (w/v), 0.2 mM glacial acetic acid (w/v), 10 M EDTA (w/v)]. Denaturing gradients were made up with 0% urea-formamide (Fluka) [40% acrylamide:N,N' bis-acrylamide (37.5:1) and 1 × TAE (40 mM Tris-HCl, 10 mM glacial acetic acid, 1 mM EDTA, pH 8.0)] and 100% Urea-formamide (Fluka) [7 M urea and 40% (v/v) deionized formamide denaturant stock solutions] using the Bio-Rad Gradient-former (Bio-Rad, Hercules, USA). Gel polymerization was initiated with 5% (w/v) ammonium persulphate (APS) and catalysed with 0.02% (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED). The denaturant gradient range was 30-70% and 35-65% for bacterial and archaeal profiles, respectively. The electrophoresis was conducted for 16 hrs at 100 V, at a constant temperature of 60 °C using a DCode DGGE System (Bio-Rad, USA). The gels were stained using 0.5 μg/ml EtBr in 1 × TAE for 15 min and destained in 1 × TAE for 20 min.

DGGE profile images were captured using the Alphaimager 3400 Imaging System UV transilluminator (AlphaInnotech Corporation, San Leandro, CA) and analysed using GelCompar® II, version 5.0 (Applied Maths). DNA fragments to be sequenced were excised from the gel using a sterile blade, placed in sterilized vials and rinsed four times with 20 μl of sterile water. The gel slices were resuspended in 20 μl of sterile water and the DNA was allowed to diffuse out at 4 °C overnight. Two μl of the eluate was used as template in the PCR reaction,

with non-GC-clamped primers. PCR conditions were as described in Table 2.3. Following amplification, the PCR products were gel purified using the illustra™ GFX™ PCR DNA, gel band purification kit (GE Healthcare) and QIAquick gel extraction kit (QIAGEN). The PCR products were then sequenced at the sequencing facilities of the University of Stellenbosch and of the University of Cape Town.

2.6 Construction of 16S rRNA gene libraries

Separate clone libraries were constructed for both archaeal and nanoarchaeal amplicons. The first archaeal library was constructed with the A571Fb/UA1204R primer set on TAT2-2 and TAT6-2 samples. TAT6 and TAT6-2 samples were used for the second archaeal library using the A3Fa/Ab927R primer set. The nanoarchaeal library was constructed with the TAT9-2 sample using the A571Fb/N989R primer set.

Amplicons from PCR reactions were pooled and purified using the illustra™ GFX™ PCR DNA and gel band purification kit (GE Healthcare), and the QIAquick gel extraction kit (QIAGEN). Ligation reactions with the pGEM®-T Easy vector (Promega) were conducted using approximately 10-100 ng DNA in a 3:1 insert DNA to plasmid vector molar ratio, with 1× ligation buffer and 1 U (Weiss units) T4-DNA ligase. The recombinant vector was then used to transform electrocompetent *E. coli* DH5$_\alpha$ cells. Transformed cells were plated onto Luria-Bertani (LB) agar plates supplemented with ampicillin (100 µg/ml), IPTG (20 µg/ml), and X-Gal (20

µg/ml). Using blue/white colony selection, white colonies were randomly selected and streaked onto fresh LB/ampicillin plates.

Table 2.3: 16S rRNA gene primer sets used for DGGE PCR amplification

Primer Set	Sequence (5' to 3')	Target	Specificity	PCR cycling parameters	Reference
341F-GC[a] 534R	CCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	16S rRNA	Universal for prokaryotes	94 °C for 5 min; 20 cycles at 94 °C for 45 s, 65 °C touchdown to 55 °C for 30 s and 72 °C for 1 min; 72 °C for 20 min.	(Muyzer *et al.*, 1993)
A340F-GC[a] A533R	CCCTACGGGGYGCASCAG TTACCGCGGCKGCTG	16S rRNA	Universal Archaea used for DGGE with A533R	94 °C for 4 min; 30 cycles at 94 °C for 30 s, 67 °C for 30 s and 72 °C for 1.5 min; 72 °C for 20 min	(Øvreås *et al.*, 1997)

Primer numbering corresponds to *E. coli* 16S rRNA sequence numbering.
[a]GC clamp added to primer 5'end: CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG.
Degenerate bases represented by R = A or G; S = C or G; Y = C or T; K = G or T.

2.7 Amplified ribosomal DNA restriction analysis (ARDRA)

Transformants were screened using colony PCR with M13 vector primers (Table 2.2). Positive amplicons that contained the correct insert were subjected to ARDRA, using *Rsa*I (Fermentas) and *Hae*II (Fermentas) restriction enzymes. The restriction digest reactions consisted of 2 U enzyme, 1.5 µl buffer (1 × final concentration) and 5 µl of the PCR product. Each restriction digest mixture was adjusted to 15 µl with sterile distilled water and incubated overnight at 37 °C. The digested fragments were electrophoresed in 2.5% agarose electrophoresis gels.

2.8 Sequencing and analysis

Clones revealing unique ARDRA patterns were grown overnight on LB/ampicillin agar plates and incubated at 37 °C. The clones were re-inoculated on 5 ml LB/ampicillin broth and grown overnight at 37 °C in a shaking incubator (225 rpm). The recombinant plasmids were then extracted and purified using the Zyppy™ plasmid miniprep kit (Zymo Research). Representatives of distinct pattern groups were sequenced with M13 vector primers using the ABI3730XL and ABI3700 sequencers (Applied Biosystems) with dye terminator v3.1 cycle sequencing kits (University of Cape Town sequencing facility) and the 3730xl and 3130xl Genetic Analyzers (Applied Biosystems) (University of Stellenbosch sequencing facilities). The sequences were compared with the GenBank sequences using BLASTn (http://www.ncbi.nlm.nih.gov). Subsequently, sequences were analysed and aligned with closest relatives from their taxonomic sub-divisions using the Bioedit (Clustal W) multiple alignment program. The presence of chimeras was checked using CHIMERA-CHECK on the ribosomal database project

(RDP) (Cole *et al.*, 2003). A neighbor-joining tree was constructed with 1000 bootstrap replicates using MEGA 4 analysis (Tamura *et al*., 2007).

2.9 Diversity analysis using multivariate statistical tools

Multivariate statistical tools, which include diversity indices, species richness and coverage estimators, MDS, PCA and CCA were used to analyse the microbial diversity further. Using Shannon's and Simpson's diversity indices, species diversity indices were calculated based on sequence data and the presence or absence of DNA bands from DGGE fingerprints.

In order to analyse the quality of the clone libraries and to determine whether they were screened sufficiently, estimates of species richness were calculated on a web-based model (http://www.aslo.org/lomethods/free/2004/0114a.html) using S_{Chao1} and S_{ACE} coupled with C_{ACE} and Good's coverage estimators. Coverage estimators indicate the proportion of species in the sample which are represented in the library (Chao, 1984; 1987). MDS plots, which cluster samples based on the similarity and relatedness of the identified bands, were analysed using GelCompar® II, version 5.0 (Applied Maths). PCA, CCA and the diversity indices were calculated using the MVSP version 3.13r (Kovach, 2004). The PCA and CCA of the 16S rRNA gene clone libraries and the PCR-DGGE fingerprints were conducted based on sequenced species and the presence (1) or absence (0) of the PCR-DGGE DNA bands. The analysis of variance (ANOVA) and the correlation tests were performed using the PROC GLM procedure of SAS version 9.1 (SAS Institute, Cary, NC, 2003) and Microsoft Office

Excel® 2007 statistical tools. Using significance or threshold level $\alpha = 0.05$, differences between values at $P < 0.05$ were considered significant.

2.10 Enriched El Tatio samples

2.10.1 Enrichment conditions

Four samples (TAT2-2, TAT6-2, TAT9-1 and TAT10-1) were inoculated into Master media (MM), containing the following components per litre: 5 g maltose; 5 g tryptone; 5 g yeast extract; 1.2 g NH_4Cl; 14 g NaCl; 2 g $MgSO_4$; 3.5 g $MgCl_2$; 0.35 g KCl; 0.3 g $CaCl_2$; 0.3 g KH_2PO_4; 0.006 g NaBr; 0.015 g H_3BO_3; 0.02 g $SrCl_2$; 0.006 g sodium citrate; 0.02 g KI; 10 ml traces solution (0.22 g $FeCl_3$; 0.1 g $MnSO_4$; 0.06 g $Na_2WO_4.2H_2O$; 0.04 g $NiCl_2$; 0.02 g $CoSO_4$; 0.02 g $ZnSO_4$; 0.02 g $CuSO_4$; 0.02 g Na_2MoO_4). The media was adjusted to pH 6.0, 6.3, 6.5, 7 and pH 7.5 with NaOH and HCl to mimic environmental conditions. Enrichments were grown in 120 ml serum bottles containing 20 ml of media, and incubated anaerobically at 80 °C. Anaerobic conditions were maintained using an anaerobic gassing manifold system connected to a pressurized nitrogen gas cylinder and vacuum pump. Samples were taken every 24 hrs for a period of 2 days, and were observed under the microscope to track growth. Serial dilutions up to 10^{-7} were conducted every 2 days and 2-4 ml volumes of each culture were pooled for DNA extraction and microscopic examinations.

2.10.2 DNA extraction of enriched samples

Samples of 2-4 ml of cell cultures were centrifuged up to three times in 2 ml Eppendorf microcentrifuge tubes at 10 000 rpm for 20 min, depending on the pellet formed. Genomic DNA was extracted using a

modified Zhou method (Section 2.1.2). Due to the resistance of some archaeal cells to lysis, other methods such as bead beating, sonication and freeze-thawing were also applied. Cell lysis was monitored microscopically before phenol extraction was applied, in order to ensure sufficient cell lysis.

2.10.3 Amplification, cloning and sequencing of nanoarchaeal 16S rRNA genes from enriched cultures

A series of optimisation trials was conducted according to the Taguchi method (Cobb & Clarkson, 1994) using A571F/N989R primers. The optimisation procedure was based on the equation $E = 2k + 1$**,** where E is the number of experiments required and k the number of parameters to be tested (such as primers, dNTPs and Mg^{2+} concentrations). In this case, primers, dNTPs, DNA and Mg^{2+} concentrations were considered as the main test parameters (Table 2.4). With these four test parameters, then $E = 2(4) + 1$ (Table 2.5).

Table 2.4: PCR reaction conditions used to set the Taguchi protocols

Reagents	**Reaction A**	**Reaction B**	**Reaction C**
F & R primers (5 mM)	0.25 µM	0.5 µM	1 µM
10× buffer	0.5×	1×	1.5×
dNTPs (2 mM)	100 µM	200 µM	250 µM
Mg^{2+} (25 mM)	1 mM	1.5 mM	2 mM
BSA (10 mg/ml)	0.5 mg/ml	1 mg/ml	1.5 mg/ml
TMAC (100 mM)	5 mM	10 mM	15 mM
Betaine (5 M)	0.25 M	0.5 M	0.75 M
Taq 1 U/ µl	0.5	1	2
DNA	15 ng	10 ng	20 ng

Table 2.5: Taguchi protocols, A, B and C represent the reagent concentrations as shown in Table 2.4

Reaction	A571Fb/N989R (5 mM)	10x buffer	dNTPs (2 mM)	Mg^{2+} (25 mM)	BSA (10mg/ml)	TMACl (100 mM)	Betaine (5 M)	Taq 1 U	[DNA]
1	A	A	A	A	A	A	A	A	A
2	A	B	B	A	B	C	A	C	B
3	A	C	C	B	B	B	A	B	B
4	B	A	B	C	A	C	C	B	C
5	B	C	B	A	A	C	B	B	C
6	B	A	C	A	B	B	A	A	A
7	C	B	A	B	B	A	B	A	B
8	C	B	A	B	C	A	B	B	C
9	C	A	B	C	C	C	B	A	C
10	C	C	C	C	A	B	C	C	B

Positive amplicons were cloned as previously described (Section 2.4), and clones were selected randomly and sequenced using genYtec sequencing services (Genética y Tecnología, Chile).

2.10.4 Fluorescence *in situ* hybridization (FISH)

FISH was applied to enriched, mixed cultures using archaeal- and nanoarchaeal-specific probes. The archaeal-specific probe, U1115 (5'-YAA CGA GCG AGA YCC-3'), and the antisense probe, U1115R (5'-CGR TCT CGC TCG TTR-3'), were used as positive and negative controls, respectively. A rhodamine labelled nanoarchaeal-specific probe, N1113 (5'-GGA AAC GAG CGC GAC CCG-3') was used to visualize nanoarchaeal cells. *Sulfolobus solfataricus* was also used as a positive control for Archaea and *E. coli* served as a negative control. Enriched, mixed cultures were fixed in 4% PBS buffered parafomaldehyde and incubated at 4 °C overnight. Sterile, ethanol-pre-cleaned slides were

coated with 0.1% gelatin, air-dried and covered with γ-methacryloxypropyl-trimethoxysilane (Sigma). The slides were dried and spotted with 3 µl of fixed cells and allowed to dry. This was followed by successive dehydration in 50%, 80% and 98% (v/v) ethanol for 3 min each, and air-driying. On each spot, 10 µl of hybridization solution [0.9 NaCl, 20 mM Tris-HCl (pH 7.2), 0.01 SDS and 50 ng DNA probe] was added and incubated in a moisture chamber at 45 °C. The cells were then washed with 1 × SET [150 mM NaCl, 20 mM Tris-HCl (pH 8) and 1 mM EDTA]. In order to reduce stringency, a series of washing solutions (5 × SET; 3 × SET and 1 × SET) were used, followed by staining with 10 µg/ml 4',6-diamidino-2-phenylindole (DAPI) for 5 min at room temperature. The cells were immersed in washing solution [180 mM NaCl, 20 mM Tris-HCl (pH 7.2), 5 mM EDTA and 0.01 % SDS] for 10 min at room temperature, rinsed in sterile water and air-dried. One drop of Gel Mount™ aqueous mounting medium (Sigma) was added to preserve the probes, which were then covered with clean sterile cover slips and sealed with Clarion™ (Sigma). The cells were visualized using an Axioplan 2 imaging fluorescence microscope (Zeiss).

2.10.5 Scanning electron microscopy (SEM) and transmission electron microscopy (SEM)

Scanning and transmission electron microscopy were used to visualise the presence of Nanoarchaea in the enriched samples.

2.10.5.1 SEM

Samples were fixed in 1% glutaraldehyde overnight and filtered using a Millipore GS membrane (0.22 µm). Each sample was added drop by drop to the filter, air-dried and then coated with gold for contrast. Samples were observed in a JEOL-JSM-25 SII microscope operated at an accelerating voltage of 30 kV. Images were obtained at different magnifications using the electron microscope services of the Pontifical Catholic University of Chile.

2.10.5.2 TEM

Samples were fixed with 3% glutaraldehyde in 0.134 M cacodylate buffer (pH 7.2) for 24 hrs, then washed with the same buffer overnight at 4 ºC and post-fixed in 1% aqueous osmium tetroxide for 2 hrs. The samples were then stained with 1% aqueous uranyl acetate for 90 min and then dehydrated with acetone in increasing concentrations. Sections were cut with an ultramicrotome and observed using a JEOL-JSM-25 SII microscope.

2.11 Cell sorting using flow cytometry

Cell sorting was performed on a BD FACSVantage™ SE cell sorter (BD Biosciences) at the University of Cape Town's flow cytometry facility. Forward and side light scatter (FSC and SSC) were used to measure and separate the cells based on size and granularity, respectively. Cells within the range of 0.4-200 µm were separated and collected on 96-well plates in fractions of 1, 5, 10 and 20 cells per fraction. The cell fractions were stored at 4 °C for further processing. Genomic DNA from FACS sorted cells was extracted according to Raghunathan *et al*. (2005).

2.12 Whole genome amplification (WGA)

Whole genome amplification of DNA from enriched cell cultures and cells sorted by flow cytometry was conducted in order to generate sufficient template DNA for sequencing. Different protocols were followed in order to obtain optimum results.

2.12.1 WGA using the REPLI-g® kit

Whole genome amplification was performed using the REPLI-g® whole genome amplification kit according to the manufacturer's instructions.

2.12.2 WGA according to Lage and Blanco

In order to overcome the background formation from the REPLI-g® kit, a modified protocol was developed according to Lage *et al*. (2003). The amplification reaction contained 3.3-4 µM of random RNA heptamers as primers, 25-100 ng DNA template and 1 × buffer Y^+ (Fermentas) supplemented with 0.12% Tween 20. The 10 µl reaction mixture was incubated at 94 °C for 4 min, cooled down to room temperature and incubated on ice for 10 min. In a final reaction volume of 30 µl, 0.4 mM dNTPs and 0.1 U/µl Φ29 DNA polymerase were added. The reaction mixture was incubated overnight at 32 °C.

WGA modification according to Blanco *et al*. (1989) contained 50 mM Tris-HCl (pH 7.5); 10 mM $MgCl_2$; 1 mM 1,4-dithiothreitol (DTT); 1 mM spermidine; 20 mM $(NH_4)_2SO_4$; 0.4 mM dNTPs; 50 µM primer; 50 ng DNA and 0.1 U/µl Φ29 DNA polymerase in a total volume of 25 µl. The reagents were

mixed together and incubated at 32 °C for 16 hrs. The reaction was stopped by incubating at 65 °C for 10 min.

wgaDNA was purified by size selection on 1% low-melting-point (LMP) agarose gels according to a fosmid library construction kit (Epicentre). High molecular weight wgaDNA was excised, purified and amplified with nanoarchaeal targeting primers, A571F/N989R, as described previously (Table 2.2).

Chapter 3

Phylogenetic characterization of microbial diversity using 16S rRNA gene amplification and DGGE analysis

3.1 Introduction

Molecular microbial ecology has advanced substantially over the past two decades, allowing the adoption of molecular tools which have revolutionized microbial ecology by allowing access to previously unknown microbial communities (Xu, 2006). These molecular tools include the use of phylogenetic markers such as the 16S rRNA gene in order to understand phylogenetic relationships (Riesenfeld *et al.*, 2004). As a result, studies of molecular phylogenetics have generated large amounts of microbial evolution data, which have shown microorganisms to be the most diverse life forms due to their ubiquitous nature. This ubiquitous nature of microorganisms is attributed to their small size, easy dispersal, the ability to replicate rapidly, their ability to metabolize a broad range of nutrients and to colonize diverse biotopes, including those at extreme temperatures, pH, redox potential, salinity and humidity (Kumar *et al.*, 2004). Consequently, many unexpected evolutionary lineages distributed across the three primary domains, the Archaea, Bacteria and Eukarya, have been discovered, which could not have been explored by conventional culturing methods (Pace, 1997).

The inability to reproduce the natural environment in the laboratory has limited access to numerous unculturable microbial communities. Culture-based methods are time consuming and have limitations compared to culture-independent methods. These include lack of efficient culturing techniques, limited

knowledge and understanding of the true state of microbial diversity, and the physicochemical parameters of the source environment (Alain & Querellou, 2009). To overcome these limitations, various molecular techniques such as, FISH, ARDRA and DGGE have been used for the comprehensive characterization of numerous microbial communities in various environments including those with extreme conditions (Elkins *et al.*, 2008); Kublanov *et al.*, 2009a).

3.2 Aims and objectives

The aims and objectives of this chapter were to understand and characterize the microbial diversity of the ETGF using bacterial, archaeal, and nanoarchaeal primer sets for 16S rRNA gene amplification. Several different primer combinations were used systematically to target a broad range of microbiota and to improve the analysis of microbial diversity and distribution.

3.3 Assessment of microbial community using bacterial 16S rRNA gene DGGE profiling

The V3 region of bacterial 16S rRNA genes was successfully amplified from the ETGF samples, using 'universal' and PCR-DGGE bacterial primers (see Section 2.5). The first round and the second round amplification generated 1.5 kb and 194 bp amplicons, respectively. The resulting DGGE profiles were analysed using GelCompar® II, version 5.0 (Applied Maths) and the extent of bacterial diversity and distribution determined. PCR-DGGE profiles of the TAT1-1, TAT2-2, TAT6-2 and TAT10-1 samples showed moderate to low diversity of the dominant species with considerable heterogeneity between the different samples (Figures 3.1 & 3.3). Only bands from the gel shown in Figure 3.3 were sequenced.

The hierarchical cluster analysis of the El Tatio samples for bacterial community profiles indicated that the duplicate and triplicate samples clustered together as expected (Figure 3.1). The TAT9-1 and TAT10-1 samples shared most of the bands, though some bands were more dominant in one sample than the other. The TAT2-2 and TAT6-2 samples also clustered together.

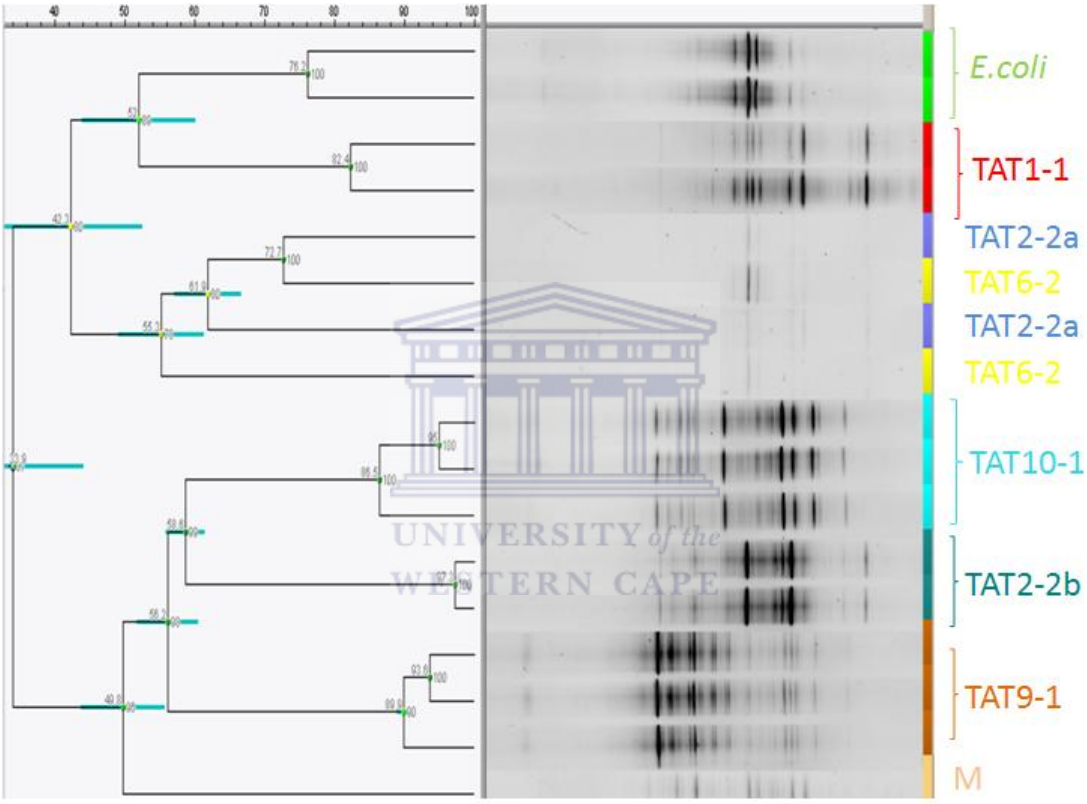

Figure 3.1: DGGE profiles and similarity dendogram based on bacterial community profiles. The standard deviation (0-9.9%), together with the cophenetic correlation values (numbers inside the branches), expresses the reliability and consistency of the cluster. The similarity values are the numbers outside the branches, and ranged between 33.93% and 97.3%.

A non-hierarchical MDS approach was used to interpret community structure and to simplify the relatedness of the samples (Figure 3.2). The reproducibility of the duplicate and triplicate samples was

also observed on both the dendogram and the MDS plot. The distance between two points reflects the level of similarity; the smaller the distance, the more related are the points.

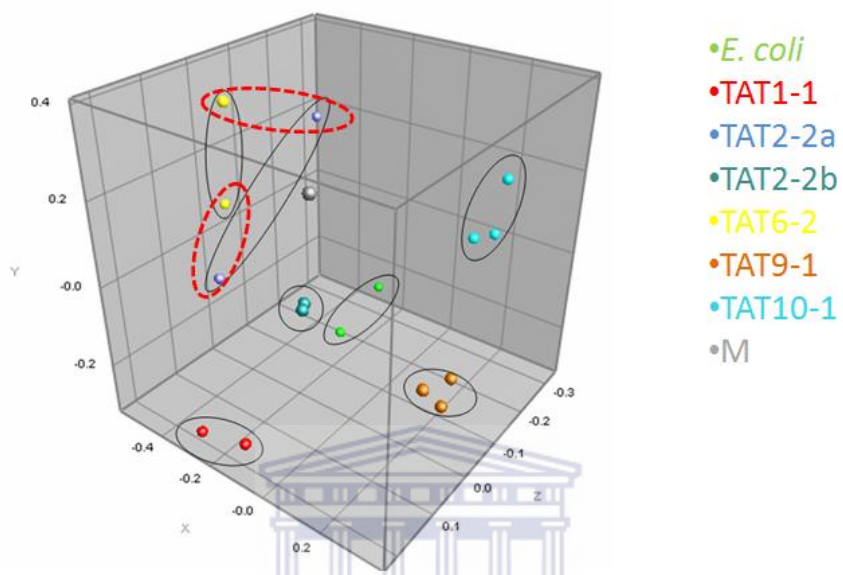

Figure 3.2: Coordinate space resulting from non-hierarchical MDS analysis of a bacterial DGGE dendogram similarity matrix. The MDS plot is based on the superimposed dendogram clusters as shown in Figure 3.1. The legend indicates the samples corresponding to the colour-coded clusters; M denotes the marker.

With the exception of TAT2-2a and TAT6-2, the MDS plot was able to confirm clusters based on the duplicate and triplicate samples. The closest related partners of TAT2-2a and TAT6-2 are demarcated by red dashes. Although some of the *E. coli* bands were visible across the different samples, TAT2-2b seemed to be the closest to the *E. coli* cluster. The MDS plot was able to confirm the community clusters as shown by the superimposed dendogram. No significant multi-sample clusters were formed, indicating the uniqueness of the community structure within the individual samples. Even though the samples were from the same geothermal field, the variation in the physicochemical parameters was evident in the relatedness of the samples.

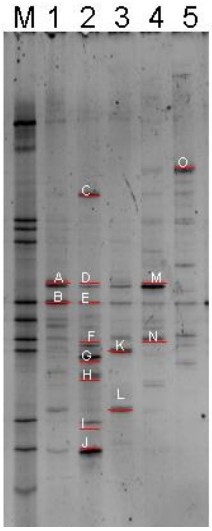

Figure 3.3: DGGE analysis of the bacterial 16S rRNA gene from the ETGF samples. Lane M: DNA molecular marker; lane 1: TAT2-2a; lane 2: TAT6-2; lane 3: TAT1-1; lane 4: TAT2-2b; and lane 5: TAT10-1. The bands labelled A-O were excised from the gel and sequenced (see Table 3.1).

A total of 107 bands were identified with GelCompar, with 28, 20, 20, 21 and 18 bands from TAT2-2a, TAT6-2, TAT1-1, TAT2-2b and TAT10-1, respectively (Figure 3.3). The main bands were excised, reamplified and sequenced. Comparisons between the different samples revealed significant differences even for the samples that originated from the same source, which were, TAT2-2a and TAT2-2b (lanes 1 and 4) (the same sample with metagenomic DNA extracted differently). Bands shared across the samples were also observed; these include bands A & D and B & E as shown by the same migration distance (Figure 3.3). Using blast sequence alignment (BLASTn suite) to compare band sequences with the same migration distance, band A (uncultured *Shewanella* sp.) showed 88% similarity to band D (uncultured bacterium), whereas band B (*Solibacillus silvestris*) had 84% similarity to band E (*Maricaulis* sp.). Based on the sequence affiliation, the same migration distance of these bands could not necessarily be attributed to sequence similarity but more to the G-C content of the sequences. Although

the sequences differed at phylum level, the short length of the sequences contributed significantly to the extent of the similarities observed. Nonetheless, the DGGE profiles were useful in portraying the differences in community structure of the samples analysed, and set a good basis for further analysis.

A summary of the phylogenetic affiliations of the sequenced amplicons is shown in Table 3.1. Not all the bands could be sequenced due to low band intensities, which may have been caused by amplification biases and/or a lower representation of these species in the template DNA (Ahn *et al*., 2009).

Table 3.1: Representatives of sequenced amplicons from 341F/534R DGGE profiles

Sample/Band	Closest match	Accession #	Similarity (%)	Phylum
TAT1-1 (**K**)	Uncultured bacterium	AJ517861	95	NR
TAT1-1 (**L**)	*Micromonospora* sp.	DQ658933	99	Actinobacteria
TAT2-2 (**A**)	Uncultured *Shewanella* sp.	EU919217	93	Proteobacteria
TAT2-2 (**B**)	*Solibacillus silvestris*	FN423774	92	Firmicutes
TAT2-2 (**M**)	*Delftia* sp.	EF692532	98	Proteobacteria
TAT2-2 (**N**)	Aquatic bacterium	AF143454	81	Proteobacteria
TAT6-2 (**D**)	Uncultured bacterium	EU539856	71	NR
TAT6-2 (**E**)	*Maricaulis* sp.	AJ623286	90	Proteobacteria
TAT6-2 (**F**)	*Brevibacillus*	FJ592179	98	Firmicutes
TAT6-2 (**G**)	*Brevibacillus* sp.	AB210951	98	Firmicutes
TAT6-2 (**I**)	Uncultured *Bacillus* sp.	EF514755	89	Firmicutes
TAT6-2 (**J**)	Uncultured *Shewanella* sp.	EU919217	94	Proteobacteria
TAT10-1 (**O**)	Uncultured *Pseudomonas* sp.	EU919222	96	Proteobacteria

NR: no rank.

Phylogenetic analysis of these sequences showed the presence of the three phylum divisions, the Proteobacteria, Actinobacteria and Firmicutes, with members belonging to different bacterial species of these taxonomic groups.

Proteobacteria

The phylum Proteobacteria was represented by *Pseudomonas*, aquatic bacterium, *Delftia*, *Maricaulis* and *Shewanella* phylotypes. Proteobacteria, as the main bacterial phylum, has been known previously to comprise only mesophilic members (Aminin *et al.*, 2008b). However, more members of this phylum have been detected from geothermal environments, including hot springs (Hugenholtz *et al.*, 1998). Typical examples are the geothermal sites in Tibet that were investigated for microbial diversity. The samples were characterized by temperature range between 60 °C and 65 °C and at pH 7.0-7.4. The bacterial representatives identified included members from α, β, and γ Proteobacteria groups (Lau *et al.*, 2009). A high frequency of proteobacterial phylotypes was identified in all samples. However, the thermophilic nature and the significance of proteobacterial distribution in thermophilic mats could not be verified (Lau *et al.*, 2009). A study conducted in sulphur-rich, highly acidic Indonesian hot springs at 70-90 °C also confirmed the presence of Proteobacteria and Bacilli, among other bacterial groups identified (Baker *et al.*, 2001). In these Indonesian hot springs, the Proteobacteria sequences were not only identified from metagenomic DNA, but also from enrichment cultures, confirming the existence of these taxa in the source samples (Baker *et al.*, 2001).

Pseudomonas

The *Pseudomonas* genus belongs to the γ-Proteobacteria. Members of this genus are able to survive in a wide range of environmental conditions and have also been detected from thermal environments such as hot springs (Aditiawati *et al.*, 2009). The thermal adaptation of *Pseudomonas* species has also been reported (Shi & Xia, 2003).

Delftia

Delftia species belong to the β-Proteobacteria and have also been detected in geothermal environments. These include the Gedongsongo hot springs of Indonesia, which are characterized by temperatures ranging from 50 °C to 86 °C and at pH 1.5-5.2 (Aminin *et al.*, 2008a). Firmicutes and other members of β and γ-Proteobacteria were also identified there.

Maricaulis

Maricaulis phylotypes are ubiquitous in water samples, and believed to play an important role in carbon cycling (Abraham *et al.*, 1999). *Maricaulis* species grow best at 20-80 g/L NaCl and in the temperature range 30-40 °C (Abraham *et al.*, 1999; Stahl *et al.*, 1992).

Shewanella

The genus *Shewanella* belongs to the γ-Proteobacteria and comprises Gram-negative facultative anaerobic bacteria (MacDonell & Colwell, 1985). Its members belong mainly to psychrophilic and mesophilic communities of freshwater and marine origin. *Shewanella* species utilize a broad spectrum of electron acceptors, such as fumarate, trimethylamine, nitrogen oxides, dimethyl sulphoxide, nitrate, nitrite, thiosulphate, and sulphite, including metal oxides (Ghosh *et al.*, 2003). The geochemistry of the ETGF displays a wide range of hydrothermal mineral deposits that support a variety of microbial

metabolic activities, which may be suitable for *Shewanella*'s broad range use of electron acceptors (Landrum *et al.*, 2009). *Shewanella* are known to be iron reducers, and their presence in the ETGF is more likely associated with iron reduction. *Shewanella*-like thermophiles have also been identified previously from hot springs and enrichment cultures grown at 66 °C (Ghosh *et al.*, 2003).

Micromonospora

The *Micromonospora* phylotypes belong to the Actinobacteria phylum, in the order of Actinomycetales. It consists of a variety of strains some of which are potential producers of bioactive compounds including antibiotics and natural rubber degraders (Zhao *et al*., 2004). Thermotolerant members of *Micromonospora* lineage have been identified from geothermal deposits such as volcanic ash near the Kamchatka hot springs (Kurapova *et al*., 2008). Diverse Actinobacterial representatives were also detected from hot springs with temperatures as high as 81 °C, suggesting capability to adapt in hyperthermal environments (Song *et al*., 2009). *Micromonospora* also comprise acidophilic and halotolerant representatives, some of which have been detected in acidic and saline soils (Zenova *et al*., 2011).

Firmicutes

Like Proteobacteria, Firmicutes also occupy a wide range of habitats and have been identified in a variety of geothermal locations (Lau *et al.*, 2009; Ogg & Patel, 2009). Firmicutes consist of three classes, the Bacilli, Clostridia and Erysipelotrichi (Mollicutes). The lineage is morphologically and physiologically diverse, with endospore- and non-endospore-forming taxa (Sneath *et al*., 1986). Heat- and dessication-resistant endospore formation is found mainly in the Bacilli and Clostridia groups,

which have high G-C content and capable of surviving in hyperthermal environments if not thermotolerant (Traag *et al.*, 2010).

Brevibacillus

Firmicutes members belonging to *Brevibacillus* have been identified in geothermal soils and other thermal environments. These include microbial mats that grow over various temperature gradients in hot springs; they were also successfully isolated in culture (Portillo *et al.*, 2009; Sridevi & Prabhune, 2009; Suzuki *et al.*, 2009). Some of the *Brevibacillus* species have been shown to display enzymatic activity at temperatures as high as 100 °C (Liang *et al*., 2009).

Bacillus

The genus *Bacillus* is made up of aerobic and facultative anaerobic Gram-positive, endospore-forming rods (Claus & Berkeley, 1986). Bacilli are able to grow both heterotrophically and autotrophically, and include thermophilic, mesophilic and psychrophilic members (Nazina *et al.*, 2001). Thermophilic Bacilli have been detected from thermal environments, including YNP (Nold *et al.*, 1996). They have also been identified from sulphur-rich and highly acidic Indonesian hot springs (Baker *et al.*, 2001).

Even though there have been reports of the presence of non-thermophilic microorganisms in hyperthermal environments, it has not been confirmed whether non-thermophilic microbial species were indigenous in these samples or imported. As a result, this study cannot rule out the possibility of contamination, whether imported by natural causes such as rain, the flow of water from one spring to another, or by anthropogenic activity.

A hyperthermal environment such as the ETGF would be expected to habour hyperthermophilic bacterial phylotypes such as Thermotoga and Aquificales. However, no hyperthermophilic bacterial phylotypes have been detected there, only those associated with mesophily and moderate thermophily. This may suggests that the non-hyperthermal bacterial members detected at the ETGF sites are allochthonous species introduced by contamination. Furthermore, the species may have metabolic preference of the chemical compounds and thermal gradients found at ETGF, which may serve as energy, electron donor and/or acceptor.

As mentioned previously, the dispersal of these non-hyperthermal microorganisms into the hot springs may be driven by natural forces such as wind, rain, ground-water flow, etc. (Hugenholtz *et al.*, 1998; Portillo & Gonzalez, 2009). Thermal gradients at the hot springs also provide a suitable environment for the mesophiles and moderate thermophiles to survive, and so adapt to the thermal environments. Although only a small fraction of the DGGE bands were sequenced, the absence of hyperthermophilic bacterial species was unexpected. Perhaps specific screening for hyperthermophilic bacteria such as the Thermotoga and Aquificales would confirm the absence of this group in the ETGF.

3.4 Assessment of microbial community using archaeal 16S rRNA gene DGGE profiling

The V3 region of the 16S rRNA gene was successfully amplified with archaeal-specific 340F/533R DGGE primers. Figure 3.4 shows the DGGE profiles of the 16S rRNA gene amplicons from sites TAT6a, TAT6b (TAT6 from A3Fa/Ab927R amplicons) and TAT6-2 amplified with the archaeal-specific 340F/533R primer set. The dominant bands (indicated by red lines) were excised for subsequent

sequencing and analysis (Table 3.2). All three sites revealed unique archaeal profiles. Sample TAT6b displayed more diversity than TAT6a and TAT6-2, as shown by the band patterns.

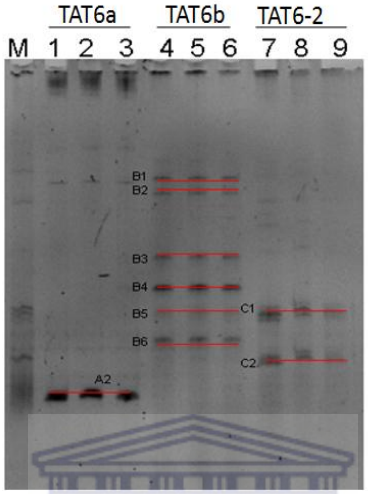

Figure 3.4: DGGE profiles of the archaeal 16S rRNA gene amplicons amplified with the 340F/533R primer set. Lane M: DNA molecular marker; lanes 1-3: TAT6a; lanes 4-6: TAT6b (TAT6 from A3Fa/Ab927R amplicons); lanes 7-9: TAT6-2. The bands indicated in red were excised from the gel and sequenced.

Table 3. 2: Representatives of sequenced amplicons from 340F/533R DGGE profiles

Sample/Band	Closest match	Accession #	Similarity (%)	Archaeal phylum
TAT6a/**A2**	*Haloterrigena limicola*	AB477224	92	Euryarchaeota
TAT6b/**B1**	Uncultured Desulfurococcales	DQ243729	85	Crenarchaeota
TAT6b/**B2**	Uncultured *Vulcanisaeta* sp.	DQ833775	82	Crenarchaeota
TAT6b/**B3**	Uncultured Desulfurococcales	DQ243729	89	Crenarchaeota
TAT6b/**B4**	*I. hospitalis*	CP000816	90	Crenarchaeota
TAT6b/**B5**	Uncultured *Vulcanisaeta* sp.	DQ833775	84	Crenarchaeota
TAT6b/**B6**	Uncultured *Vulcanisaeta* sp.	DQ833775	85	Crenarchaeota
TAT6-2/**C1**	*Halobacteriaceae* gen. sp.	AF478471	89	Euryarchaeota
TAT6-2/**C2**	*Natronococcus xinjiangense*	AF251285	85	Euryarchaeota

All three samples revealed unique banding patterns, with both TAT6a and TAT6-2 having sequences of euryarchaeal origin. The sequences were similar to *Haloterrigena limicola*, *Vulcanisaeta* sp. and *Halobacteriaceae* species. The TAT6b sample comprised crenarchaeotes from the order Thermoproteales, including sequences similar to *Vulcanisaeta* sp. and *I. hospitalis. H. limicola* species are chemo-organotrophic, aerobic haloarchaea, which grow optimally at NaCl concentrations of 3.1 M, pH 7.0 and at temperatures between 45-50 °C (Cui *et al*., 2006). *Vulcanisaeta* species belong to the Thermoproteaceae family and grow at extremely high temperatures between 85-90 °C in weakly acidic conditions (pH 4.0-4.5); and they have been identified in acidic hot springs (Itoh *et al*., 2002). It was also interesting to identify sequences with high similarity to *I. hospitalis* at the same site where nanoarchaeal sequences were previously identified (Casanueva *et al*., 2008). This suggests the possibility of the presence of nanoarchaeal-*Ignicoccus* symbionts at this location. However this can be further confirmed through FISH by the evidence of nanoarchaeal-*Ignicoccus* co-cultures.

Using UPGMA hierarchical cluster analysis, the DGGE fingerprints were analysed for their relatedness, in order to identify the closely related communities (Figure 3.5). The dendogram showed that the fingerprints were significantly different from one another as shown by the low similarity values, i.e. 27% for the TAT6b and TAT6-2 cluster and 5.4% for the TAT6a and TAT6b cluster. In contrast to the variability of the samples, the corresponding triplicates of each sample clustered together with the exception of the TAT6a (3) replicate, which was far apart from other replicates due to a missing DNA band. The triplicate clusters were supported by high similarity values within the range 77.8-100%. This also served as a measure of the reproducibility and reliability of the DGGE method, which implies that

the DGGE method is reliable and reproducible, and is able to provide quantitative information on sample variation and microbial community composition.

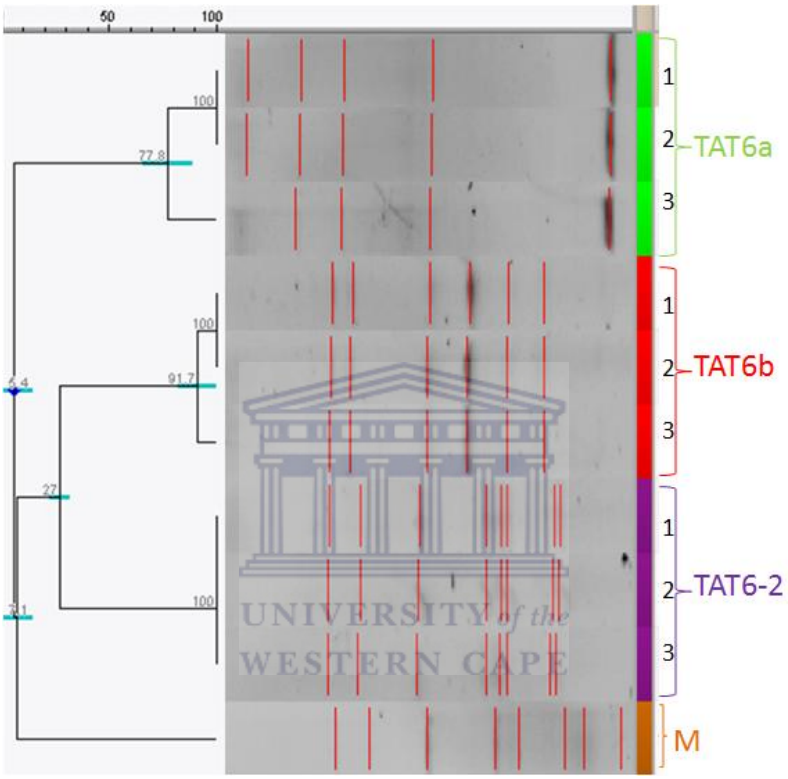

Figure 3.5: UPGMA hierarchical clustering of the DGGE fingerprints from the archaeal 16S rRNA gene (amplified with the 340F/533R primer set) of the TAT6a, TAT6b and TAT6-2 samples in triplicates. The dendogram is based on an archaeal community (340F/533R). The similarity values (numbers outside the branches) ranged between 5.4-100%. The triplicate clusters of each sample are differentiated by colour coding.

The dendogram clusters were subsequently superimposed on the MDS plot to map the relationship of the samples using the Bray-Curtis similarities of the three samples, and reducing the complex DGGE pattern to a point form clusters in a three-dimensional space. The MDS plot revealed a sample-specific pattern

of clustering corresponding to the dendogram in Figure 3.5 (Figure 3.6). The closer the points were to each other, the more similar were the replicate samples.

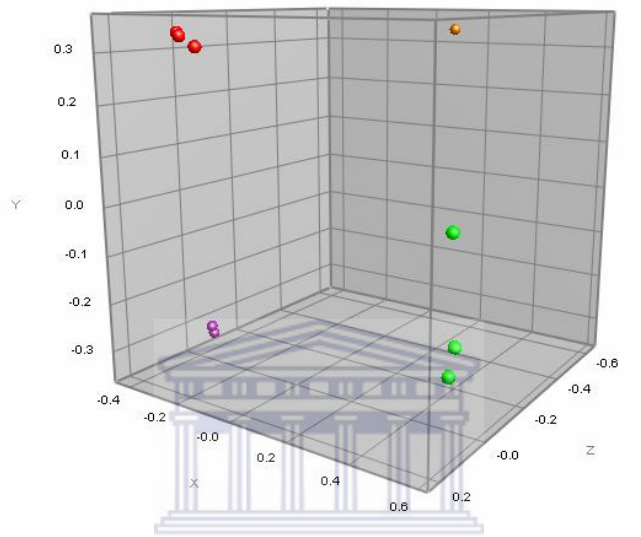

Figure 3.6: Coordinate space resulting from MDS analysis of an archaeal DGGE dendogram similarity matrix. The coloured dots correspond to the dendogram keys as indicated in Figure 3.5 and are clustered according to their relatedness.

Analysis of DGGE fingerprints showed that the microbial composition of the ETGF samples differed substantially, and most likely determined by the variable physicochemical conditions of the samples.

3.5 Assessment of microbial community using archaeal 16S rRNA gene clone libraries

3.5.1 Archaeal-specific 16S rRNA gene amplification

Archaeal 16S rRNA genes were successfully amplified using various archaeal primer sets as indicated in Table 2.2. Metagenomic DNA from two samples, TAT2-2 and TAT6-2, were amplified with the A571Fb/UA1204R primer set, yielding 634 bp fragments. DNA from other samples could not amplify with the A571Fb/UA1204R primer set. As a result, the A3Fa/Ab927R primer set was used as an alternative for crenarchaeote and eukarchaeote selective amplification, and generated 925 bp fragments. DNA from three samples (TAT2-2, TAT6-2 and TAT6) was able to amplify with the A3Fa/Ab927R primer set. However, only amplicons from TAT6 and TAT6-2 DNA were successfully cloned.

3.5.2 Nanoarchaeal-specific 16S rRNA gene amplification

Primer set A571Fb/N989R was used for nanoarchaeal-specific 16S rRNA gene amplification. Successful amplification was achieved only for DNA from TAT9-2 site, and generated 419 bp fragments.

3.5.3 ARDRA of phylotypes from 16S rRNA gene libraries

The archaeal and nanoarchaeal 16S rRNA gene amplicons that were successfully cloned and shown to contain the correct insert size by M13 PCR, were subjected to ARDRA. The heterogeneity of the phylotypes was determined as indicated in Figures 3.7-3.9. Rapid ARDRA analyses were conducted; however, only a selection is shown here.

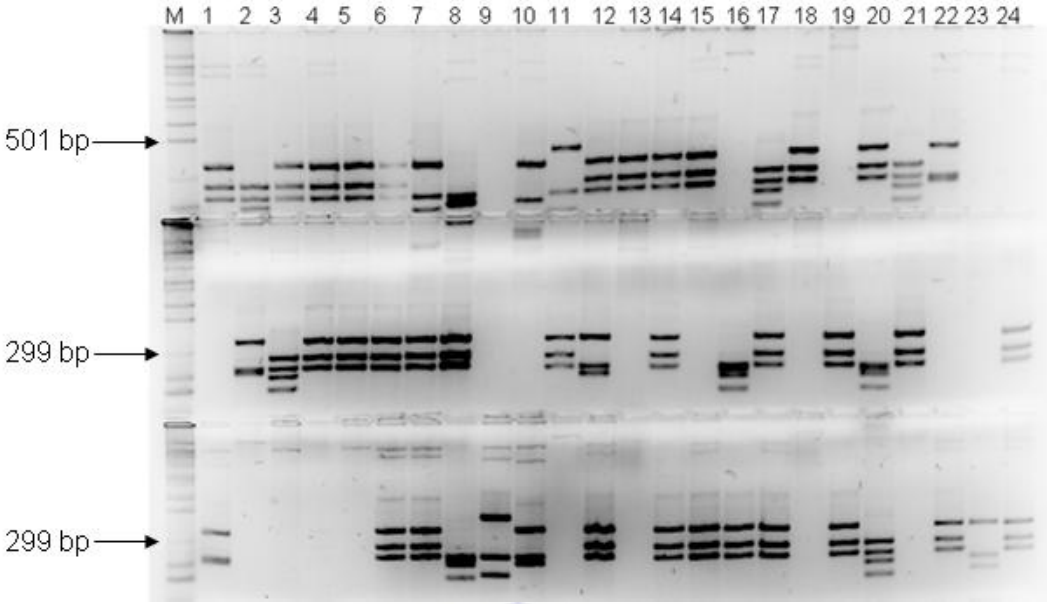

Figure 3.7: Representatives of ARDRA patterns generated from *Rsa*I digestion. Lane M: λ DNA marker; lanes 1-24: TAT6-2 phylotypes from the A3Fa/Ab927R library.

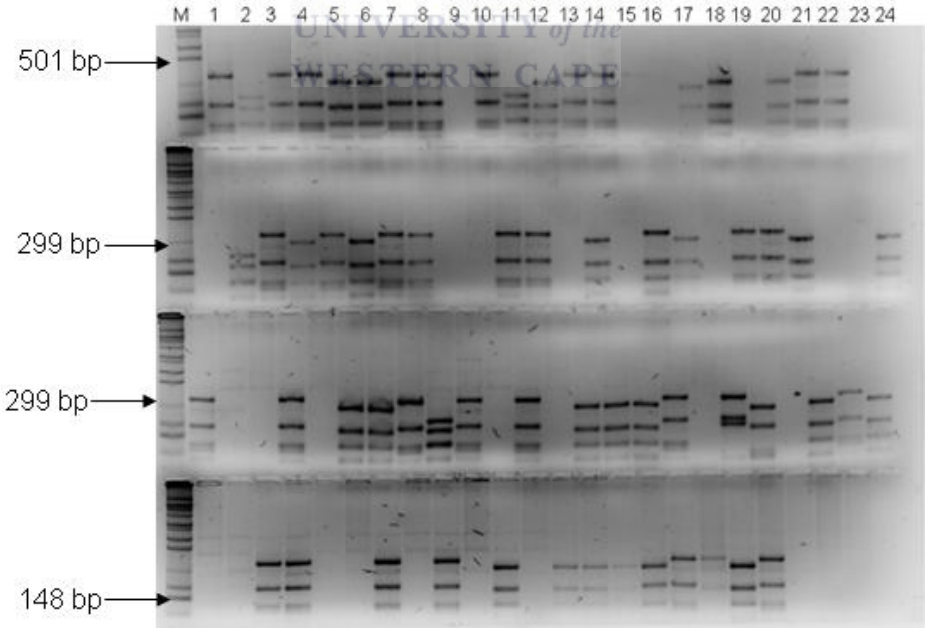

Figure 3.8: Representatives of ARDRA patterns generated from *Hae*III digestion. Lane M: λ DNA marker; lanes 1-24: TAT6-2 phylotypes from the A3Fa/Ab927R library.

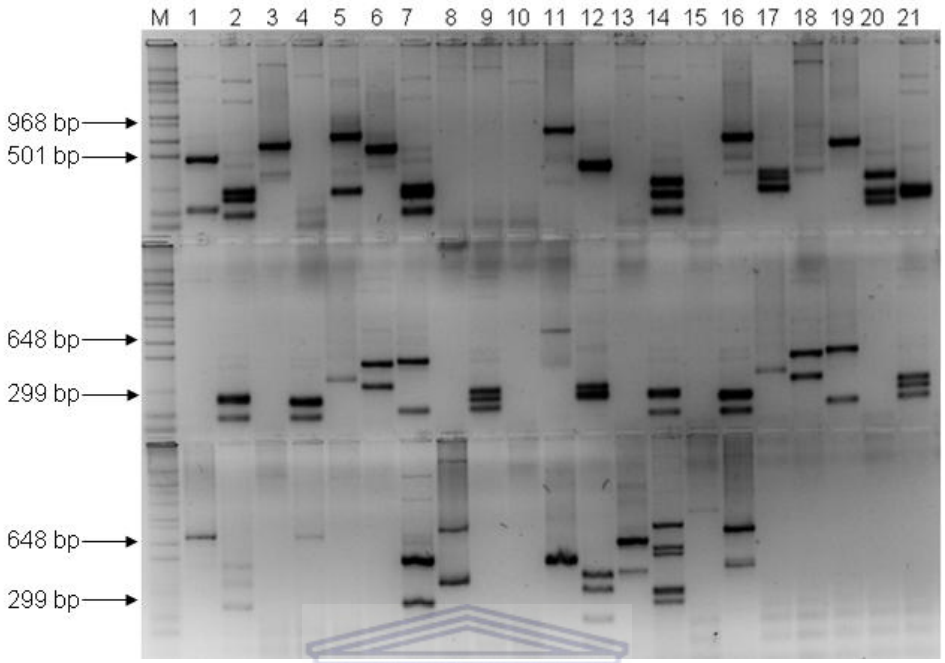

Figure 3.9: Representatives of ARDRA patterns generated from *Rsa*I digestion. Lane M: λ DNA marker; lanes 1-21: TAT9-2 phylotypes from the A571Fb/N989R library.

A large number of phylotypes from each of the 16S rRNA gene libraries were analysed by ARDRA and showed unique ARDRA patterns. A summary of successful amplifications with the respective accumulated ARDRA patterns is tabulated in Table 3.3. Twenty unique ARDRA patterns were observed for each of the A571Fb/UA1204R libraries. The A3Fa/Ab927R library yielded 8 and 5 unique patterns from TAT6-2 and TAT6 samples, respectively.

Table 3.3: A summary of successful archaeal 16S rRNA gene amplificons with the respective number of clones screened and the accumulated ARDRA patterns

Sample	Primer set	Fragment size (bp)	Number of clones screened	ARDRA patterns accumulated
TAT2-2	A571Fb/UA1204R	634	384	20
TAT6-2	A571Fb/UA1204R	634	288	20
TAT6	A3Fa/Ab927R	925	108	5
TAT6-2	A3Fa/Ab927R	925	188	8
TAT9-2	A571Fb/N989R	419	76	27*

* May be a result of artifacts

3.5.4. Comparative sequence analysis of archaeal 16S rRNA gene clone libraries

Representatives of phylotypes that revealed unique ARDRA patterns were sequenced and analysed. It was observed that some of the phylotypes that revealed unique ARDRA patterns were affiliated to the same taxonomic group.

3.5.4.1 Archaeal-specific A571Fb/UA1204R clone library

A total of 69 representatives were selected for sequencing based on ARDRA pattern groups (Table 3.4). Thirty-nine percent of the sequences recovered were represented by uncultured haloarchaea. Nanoarchaeotes were represented by 32% with 76-100% sequence similarity to uncultured nanoarchaeotes, Tok173 (EF562627) and Tok17C9 (EF562624). Uncultured archaea represented 26% of the sequenced phylotypes, whereas uncultured euryarchaeotes and bacteria were represented by only one phylotype each, contributing to 1.5% of the library. The presence of bacteria was not expected from this library, and suggests the possibility of mispriming.

Haloarchaea

The sodium chloride concentrations at the TAT2-2 and TAT6-2 sites were 23% and 19% (w/v), respectively, favouring the growth of haloarchaea. Haloarchaea usually form part of the predominant microorganisms in hypersaline environments, such as soda lakes and salterns (Oren, 1994). They are known to require a minimum salt concentration of 9% (w/v NaCl) for growth (Oren, 2002). The order Halobacteriales is composed of extreme haloarchaea which show optimum growth at 20% (w/v NaCl) and more (Purdy *et al*., 2004).

Previous studies have reported a high diversity of halophilic archaea from geothermal areas, which are normally associated with non-geothermal habitats (Ellis *et al*., 2008). According to Ellis *et al*. (2008), halophilic archaea survived temperatures above 75 °C for a short time (for approximately 5-30 min and longer). However, laboratory enrichments showed the highest growth of halophilic archaea to be 51 °C. This suggests that halophilic archaea exposed to hyperthermal environments can survive high temperatures, even though they are not able to grow optimally in these conditions. The functional contribution of haloarchaea in the ETGF microbial community is not known. However, haloarchaeal isolates with the ability to reduce elemental sulphur and sulphide have been reported (Elshahed *et al*., 2004). This suggests that haloarchaeal species that inhabit the ETGF sites may also be involved in sulphur reduction.

Nanoarchaea

The existence of nanoarchaea from the ETGF was not surprising. These phylotypes have been identified previously from this location (Casanueva *et al*., 2008). However, it is interesting that the closest affiliates were nanoarchaeal phylotypes from New Zealand (Casanueva *et al*., 2008). This further affirms that nanoarchaea are not distributed according to geographical origin.

Nanoarchaea were first known by a single isolate of *N. equitans*, which remains the only cultured nanoarchaeal isolate to date (Huber *et al.*, 2002). Previous studies have reported the existence of nanoarchaeal phylotypes retrieved from a wide spectrum of environmental geothermal biotopes around the globe, including non-thermal salterns (Casanueva *et al*., 2008; Hohn *et al.*, 2002; McCliment *et al.*, 2006). These results correlate with our findings of co-habitation of haloarchaea and nanoarchaea, and may indicate a possible nanoarchaeal-haloarchaeal relationship, similar to the *N. equitans*-*I. hospitalis* association. Such association may be explored by culturing these species and further investigation of their mode of interaction. Attempts to culture nanoarchaea were applied in this study and will be discussed in Chapter 5.

Table 3.4: Representatives of identified phylotypes from the A571Fb/UA1204R clone library

Clone	Closest match	Accession #	Similarity (%)	Archaeal phylum
T22p9	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22p10	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22p15	Uncultured archaeon	CU467152	94	NR
T22p19	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22pB4	Uncultured archaeon	FM242736	95	NR
T22C1pB2	Uncultured nanoarchaeote Tok173	EF562627	96	Nanoarchaea
T22C1pB5	Uncultured archaeon metC1	DQ082936	90	NR
T22C1pB8	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C1pB9	Uncultured archaeon metC1	DQ082936	90	NR
T22C1pC1	Uncultured archaeon	CU467152	91	NR
T22C1pC4	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C1pC5	Uncultured haloarchaeon	AM947464	95	Euryarchaeota
T22C1pC6	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C1pD8	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C1pE6	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C2pA7	Uncultured nanoarchaeote Tok173	EF562627	91	Nanoarchaea
T22C2pA8	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C2pA9	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C2pC7	Uncultured archaeon	DQ082936	91	NR
T22C2pD1	Uncultured euryarchaeote	FN391292	88	NR
T22C2pD2	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C2pD10	Uncultured nanoarchaeote Tok173	EF562627	90	Nanoarchaea
T22C2pE10	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C2pF9	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C3pA6	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C3pA8	Uncultured archaeon metC1	DQ082936	90	NR
T22C3pB4	Uncultured haloarchaeon	FN391225	87	Euryarchaeota
T22C3pC4	Uncultured archaeon metC1	DQ082936	91	NR
T22C3pC7	Uncultured haloarchaeon	FN391228	85	Euryarchaeota
T22C3pD1	Uncultured nanoarchaeote Tok173	EF562627	96	Nanoarchaea

NR: No rank.

Table 3.4: continued

Clone	Closest match	Accession #	Similarity (%)	Archaeal phylum
T22C3pE12	Uncultured nanoarchaeote Tok173	EF562627	97	Nanoarchaea
T22C3pG12	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C3pH6	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C4pA1	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C4pA6	Uncultured haloarchaeon	AM071500	68	Euryarchaeota
T22C4pA10	Uncultured archaeon	DQ082936	91	NR
T22C4pB5	Uncultured archaeon	DQ082936	90	NR
T22C4pB7	Uncultured archaeon	DQ082936	91	NR
T22C4pC2	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C4pC9	Uncultured nanoarchaeote Tok173	EF562627	95	Nanoarchaea
T22C4pD6	Uncultured haloarchaeon	AM947464	97	Euryarchaeota
T22C4pD7	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C4pE5	Uncultured archaeon	DQ082936	91	NR
T22C4pE12	Uncultured haloarchaeon	AM947464	98	Euryarchaeota
T22C4pF2	Uncultured archaeon	DQ082936	90	
T22C4pF9	Uncultured archaeon	DQ082936	85	NR
T22C4pG10	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T22C4pG11	Uncultured haloarchaeon	AM947464	95	Euryarchaeota
T22C4pH8	Uncultured archaeon	DQ082936	91	NR
T62C1pA4	Uncultured nanoarchaeote Tok17C9	EF562624	76	Nanoarchaea
T62C1pA5	Uncultured archaeon metC1	DQ082936	91	NR
T62C1pA6	Uncultured archaeon metC1	DQ082936	91	NR
T62C1pA7	Uncultured nanoarchaeote Tok173	EF562627	99	Nanoarchaea
T62C1pA10	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C1pB2	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C1pB4	Uncultured nanoarchaeote Tok17C9	EF562624	99	Nanoarchaea
T62C1pD5	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea

NR: No rank.

Table 3.4: continued

Clone	Closest match	Accession #	Similarity (%)	Archaeal phylum
T62C1pF6	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C2pA5	Uncultured haloarchaeon	FN391228	81	Euryarchaeota
T62C2pA8	Uncultured archaeon	FM210839	87	NR
T62C2pB6	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C2pB9	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C2pC2	Uncultured nanoarchaeote Tok17C9	EF562624	100	Nanoarchaea
T62C2pE1	Uncultured nanoarchaeote Tok173	EF562627	100	Nanoarchaea
T62C2pE5	Uncultured nanoarchaeote Tok17C9	EF562624	99	Nanoarchaea
T62C2pG6	Uncultured bacterium	EU455990	88	NR
T62C2pG11	Uncultured nanoarchaeote Tok17C9	EF562624	95	Nanoarchaea
T62C2pH4	Uncultured nanoarchaeote Tok173	EF562627	99	Nanoarchaea
T62C2pH5	Uncultured nanoarchaeote Tok173	EF562627	97	Nanoarchaea

NR: No rank.

The sequences retrieved from the A571Fb/UA1204R library clustered into two major groups, the nanoarchaea and the haloarchaea. As a result, two phylogenetic trees were constructed, which composed of GenBank sequences and sequences retrieved from this study (Figures 3.10 and 3.11). The nanoarchaea group clustered with high confidence with the *N. equitans* at 99% bootstrap support (Figure 3.10). A large number of sequences clustered within the uncultured haloarchaeal group with 100% bootstrap support (Figure 3.11). Representatives of uncultured archaea were also detected.

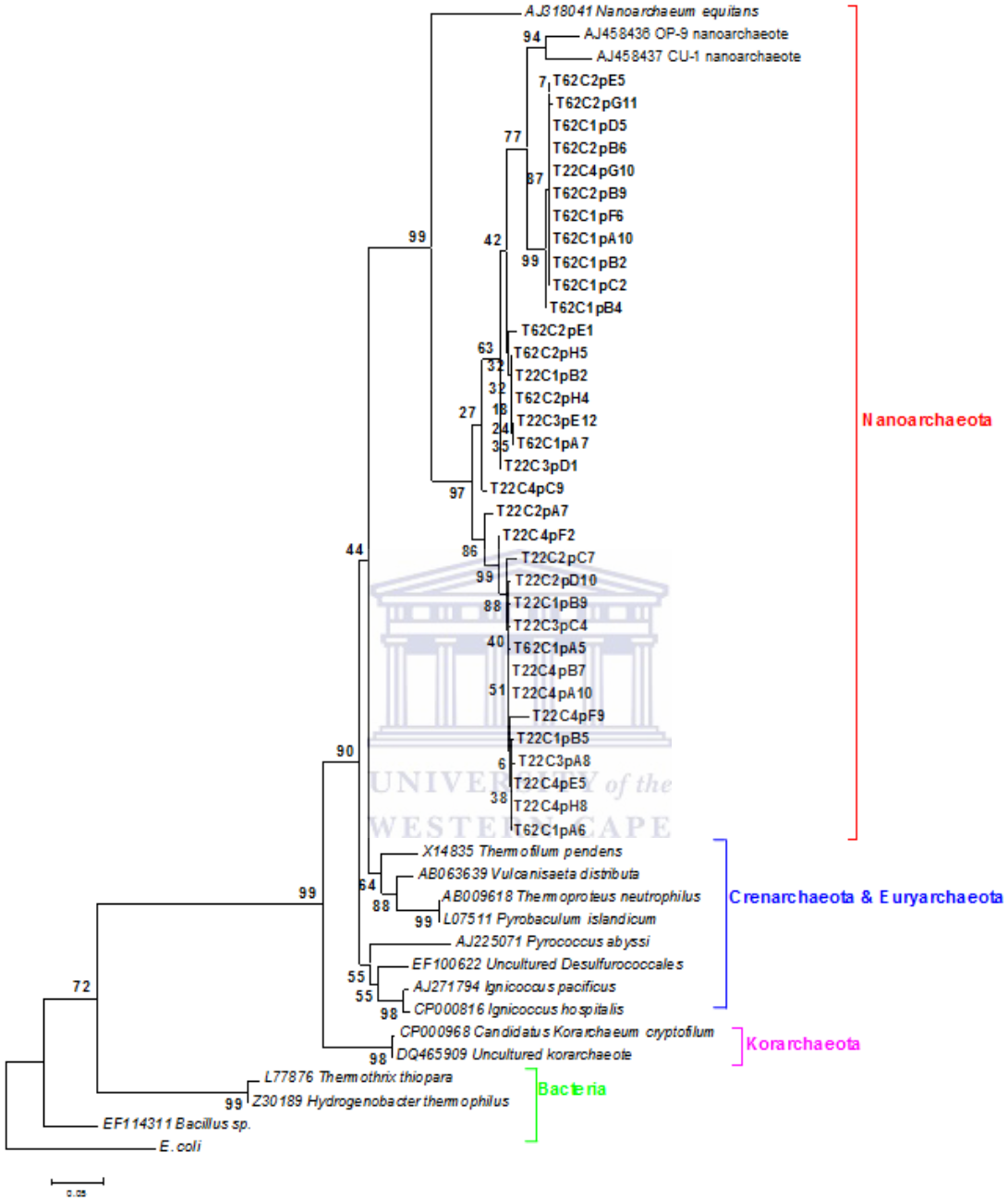

Figure 3.10: Phylogenetic tree constructed using MEGA 4 based on 438 bp 16S rRNA gene sequences amplified with the A571Fb/UA1204R primer set. The evolutionary history was inferred using the Neighbor-joining method. The evolutionary distances were computed using the maximum composite likelihood method; the scale bar represents the number of base substitutions per site. The numbers at branch points indicate bootstrap values of 1000 replicates. The sequences retrieved in this study are indicated in bold.

The use of the A571Fb/UA1204R primer pair was first reported by Baker *et al*. (2003). This primer set was selected for its ability to amplify 16S rRNA gene from Euryarchaea, Crenarchaea, Nanoarchaea and Korarchaea phyla. The results obtained revealed relatively broad archaeal primer specificity as shown by the presence of euryarchaeota and nanoarchaeota phylotypes. The presence of unclassified euryarchaeotes and uncultured bacterium was also detected with only a few representatives (4% each). The library was dominated by Euryarchaea (39%), which were represented by uncultured haloarchaea, followed by 32% uncultured nanoarchaeotes and 26% uncultured archaea. However, no evidence of Crenarchaeota or Korachaota was observed.

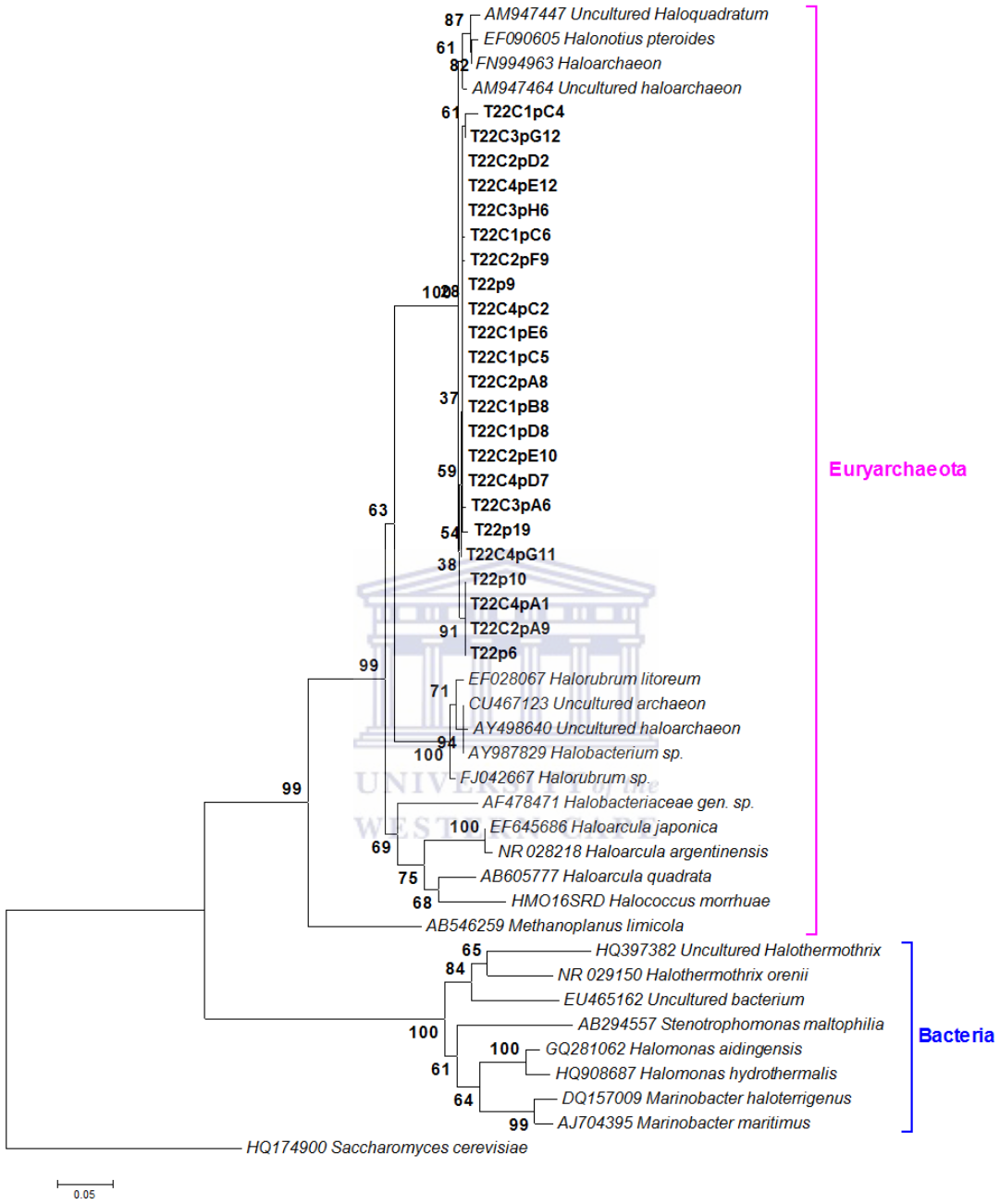

Figure 3.11: Phylogenetic tree constructed using MEGA 4 based on 436 bp 16S rRNA gene sequences amplified with the A571Fb/UA1204R primer set. The evolutionary history was inferred using the neighbor-joining method. The evolutionary distances were computed using the maximum composite likelihood method; the scale bar represents the number of base substitutions per site. The numbers at branch points indicate bootstrap values of 1000 replicates. The sequences retrieved in this study are indicated in bold.

The presence of haloarchaea was not surprising, in view of the high salinity levels in these hydrothermal systems. Full understanding of the significance of this archaeal group within the El Tatio microbial community, its means of survival and adaptation to hyperthermal habitats requires further investigation. This may also include cultivation and assessments of other parameter that are crucial for metabolic activities within the community.

3.5.4.2 Archaeal-specific A3Fa/Ab927R clone library

The systematic use of different primer combinations was exploited in order to improve community representation and to minimize bias. Primer set A3Fa/Ab927R had the advantage of generating a relatively large gene fragment and was used mainly to target Crenarchaea and Euryarchaea. Twenty-six representative phylotypes from the A3Fa/Ab927R clone library were selected for sequencing based on ARDRA pattern groups (Table 3.5). The A3Fa/Ab927R library was dominated by 34% Desulfurococcales, followed by 27% Halobacteriaceae. Thermoproteales and other uncultured archaea were also represented in relatively low numbers. The different community members were similar to *Thermoproteus tenax,* Halobacteriaceae sp., *Haloferax* sp., *Natrinema pallidum*, *Halogeometricum borinquense*, *Halorubrum* sp., uncultured haloarchaea and uncultured archaea as shown in Table 3.5. The library clustered into two significant categories, the Euryarchaeota, represented by Haloarchaea, and the Crenarchaeota, represented by Thermoproteales and Desulfurococcales (Figure 3.12). The Crenarchaea and Euryarchaea comprise a high proportion of Archaea, it was expected that the A3Fa/Ab927R primer set would amplify more crenarchaeal and euryarchaeal representatives. However, this was not the case and could be attributed to the specificity limitations of this primer set.

Table 3. 5: Representatives of phylotypes identified from the A3Fa/Ab927R clone library

Clone	Closest match	Accession #	Similarity (%)	Archaeal phylum
T6p3	Uncultured Desulfurococcales	EF100622	91	Crenarchaeota
T6p4	Uncultured Desulfurococcales	EF100622	94	Crenarchaeota
T6p14	Uncultured *Desulfurococcus* sp.	AF361213	94	Crenarchaeota
T6p15	*Thermoproteus tenax*	AY538162	91	Crenarchaeota
T6p23	Uncultured Desulfurococcales	EF100622	94	Crenarchaeota
T6p36	Uncultured *Desulfurococcus*	AF361213	94	Crenarchaeota
T6p37	Uncultured Desulfurococcales	EF100622	90	Crenarchaeota
T6p38	Uncultured archaeon	AJ969823	98	NR
T6p39	*Thermoproteus tenax*	AY538162	96	Crenarchaeota
T6p40	Uncultured Desulfurococcus sp.	AF361213	94	Crenarchaeota
T6p44	Uncultured Desulfurococcus sp.	AF361213	93	Crenarchaeota
T6p54	*Halobacteriaceae* gen. sp.	AJ002945	97	Euryarchaeota
T6p55	Uncultured Desulfurococcales	EF100622	93	Crenarchaeota
T62p1	*Haloferax* sp.	AM176545	99	Euryarchaeota
T62p2	*Halobacterium* sp.	AY987829	99	Euryarchaeota
T62p4	*Haloferax* sp.	AM176545	99	Euryarchaeota
T62p5	*Halorubrum* sp.	FJ042667	99	Euryarchaeota
T62p6	Uncultured archaeon	AJ969912	86	NR
T62p7	*Natrinema pallidum*	AF367370	97	Euryarchaeota
T62p8	*Halobacteriaceae* gen. sp.	AF478471	98	Euryarchaeota
T62p10	*Halobacteriaceae* gen. sp.	AJ002945	98	Euryarchaeota
T62p11	Uncultured haloarchaeon	AY498640	99	Euryarchaeota
T62p17	*Halogeometricum borinquense*	DQ853414	99	Euryarchaeota
T62p21	*Halobacteriaceae* gen. sp.	AF478471	99	Euryarchaeota
T62p22	*Halobacteriaceae* gen. sp	AJ002945	98	Euryarchaeota
T6p96	*Halobacteriaceae* gen. sp.	AJ002945.1	96	Euryarchaeota

NR: no rank.

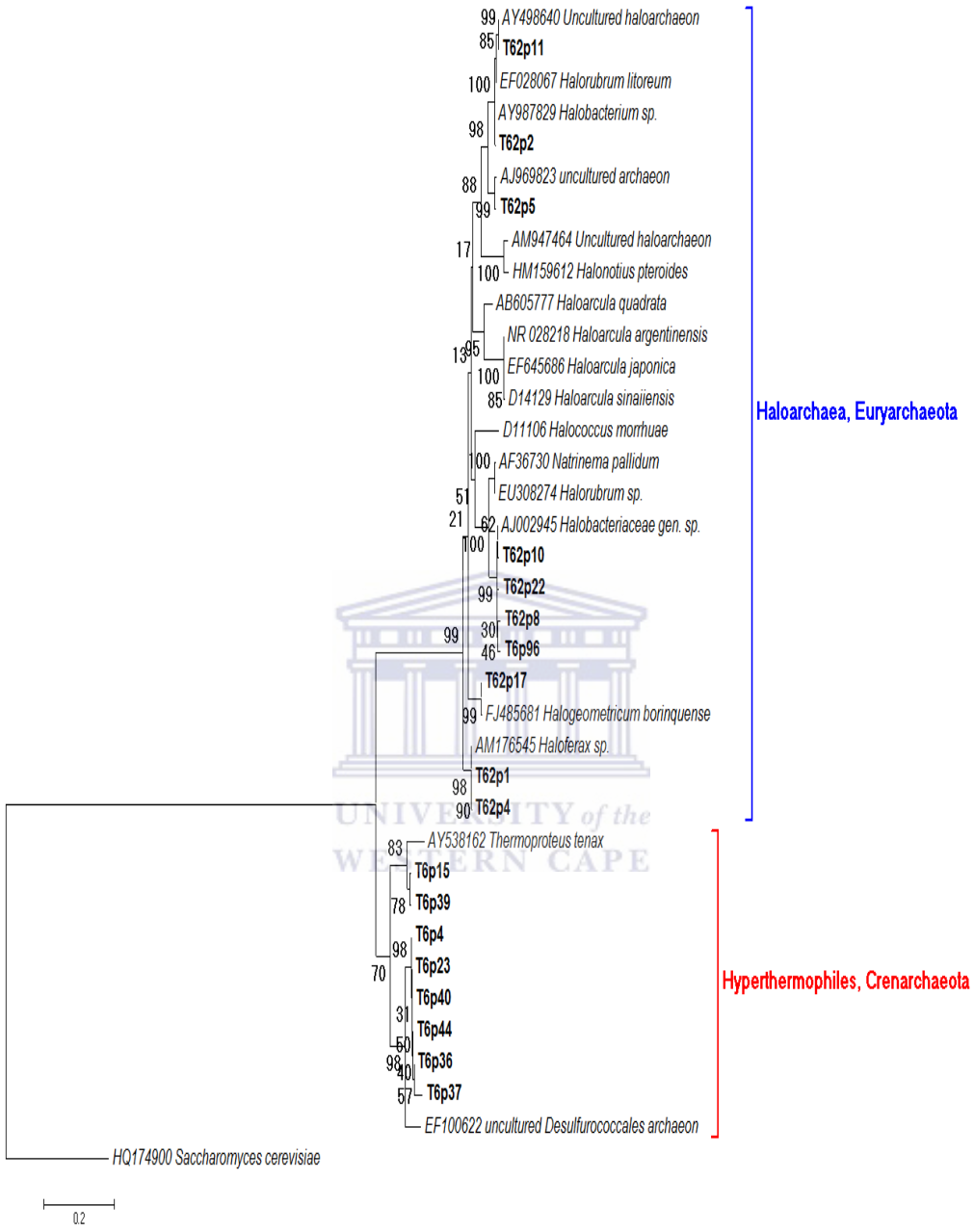

Figure 3.12: Phylogenetic tree constructed using MEGA 4 based on 602 bp 16S rRNA gene sequences amplified with the archaeal A3Fa/Ab927R primer set. The evolutionary history was inferred using the neighbor-joining method (Saitou & Nei, 1987). The evolutionary distances were computed using the maximum composite likelihood method; the scale bar represents the number of base substitutions per site. The numbers at branch points indicate bootstrap values of 1 000 replicates. Phylotypes isolated in this study are indicated in bold.

The presence of haloarchaea has been discussed in Section 3.5.4.1.

Thermoproteus tenax

Thermoproteus tenax is a hyperthermophilic anaerobic thermoacidophile chemolithotrophic archaeon, able to grow in the presence of CO_2, H_2 and S^0, and also chemo-organotrophically in the presence of S^0 and carbohydrates (Siebers *et al*., 1998). *T. tenax* belongs to the order Thermoproteales of the Thermoproteaceae family. A sequence with high similarity to *T. tenax* was detected from TAT6 (85 °C, pH 6.5). The ETGF is thus a suitable habitat for optimum growth of *T. tenax* and related microorganisms.

Desulfurococcales

Over a third of the sequences identified belonged to the Desulfurococcales. Several members of this order have been isolated from a variety of shallow and deep-sea hydrothermal locations (Huber *et al*., 2006). Members of the Desulfurococcales are the most hyperthermophilic archaeal microorganisms known, and one isolate is able to grow poorly at 121 °C, which is currently thought to be the 'upper temperature limit' of life (Kashefi & Lovely, 2003). Members of the genus *Ignicoccus* also belong to this group and are obligate chemolithothophic sulphur reducers, with moderate halophiles and moderate acidophiles (Paper *et al.*, 2007; Huber *et al.*, 2000). *I. hospitalis* is the only member known to host *N. equitans*. Even though no *I. hospitalis* representatives were identified from the clone libraries, but on PCR-DGGE sequences (Table 3.2), it is therefore most likely that nanoarchaeal-*Ignicoccus* symbionts exist in this field.

3.5.4.3 "Nanoarchaeal-specific" A571Fb/N989R clone library

The A571Fb/N989R primer set has been reported previously to amplify nanoarchaeal sequences (Casanueva *et al.*, 2008). In this study, A571Fb/N989R primer set was used to screen for nanoarchaeotes from the ETGF. The A571Fb/N989R primer set was successful in retrieving nanoarchaeal sequences, including archaea and unexpected bacterial sequences (Table 3.5). A total of 24 phylotypes were sequenced, comprising 38% *Hydrogenobacter* and 29% nanoarchaeotes, which shared 97-100% similarity with a Chilean uncultured nanoarchaeote clone, TA91N9A2 (EF562618), and an uncultured nanoarchaeote clone, TOK17C9 (EF562624) from New Zealand. Other, less represented taxa included *Geoglobus* sp., *Thermothrix thiopara*, uncultured Verrucomicrobia, uncultured archaea and uncultured bacteria. Four putative novel sequences were also identified; these are T92pA12, T92p40, T92p41, and T92p55. These clones could match only <100 bp of the 419 bp query sequence submitted to the GenBank sequence database. These sequences will be submitted to the GenBank database.

A phylogenetic tree was constructed based on 29 taxa as shown in Figure 3.13. Nanoarchaeal sequences retrieved from this study were compared to the nanoarchaeal sequences in the GenBank database.

Nanoarchaea

The presence of Nanoarchaea has been discussed in Section 3.5.4.1.

Hydrogenobacter

Hydrogenobacter species have been reported to be one of the early branching bacterial species, and tend to share some phylogenetic similarities with archaea. As a result, *Hydrogenobacter* species have become easily accessible with archaeal primers (Shima *et al*., 1994).

Table 3.6: Representatives of phylotypes identified from the A571Fb/N989R clone library

Clone	Closest match	Accession #	Similarity (%)	Phylum
T92pA11	Uncultured nanoarchaeote	EF562624	99	Nanoarchaea
T92pA12	Uncultured archaeon	EU118585	92	NR
T92pB1	*Geoglobus* sp.	FM242725	92	Euryarchaeota
T92pB2	Uncultured bacterium	FM213511	100	NR
T92pB6	Uncultured nanoarchaeote	EF562624	100	Nanoarchaea
T92pB7	Uncultured nanoarchaeote	EF562624	99	Nanoarchaea
T92p25	Uncultured nanoarchaeote	EF562618	98	Nanoarchaea
T92p26	*Hydrogenobacter* sp.	AM259501	98	Aquificae
T92p29	Uncultured nanoarchaeote	EF562618	97	Nanoarchaeota
T92p31	*Hydrogenobacter* sp.	AM259501	98	Aquificae
T92p38	*Thermothrix thiopara*	L77876	97	Proteobacteria
T92p40	Uncultured Verrucomicrobia	GQ261849	94	Verrucomicrobia
T92p41	Uncultured rumen archaeon	HM211722	80	NR
T92p43	Uncultured bacterium	AM921478	100	NR
T92p44	*Hydrogenobacter* sp.	AM259501	98	Aquificae
T92p45	*Hydrogenobacter* sp.	AM259501	97	Aquificae
T92p50	*Hydrogenobacter* sp	AM259501	98	Aquificae
T92p52	*Hydrogenobacter* sp.	AM259501	97	Aquificae
T92p54	Uncultured nanoarchaeote	EF562618	99	Nanoarchaeota
T92p55	Uncultured archaeon	DQ791688	97	NR
T92p57	*Hydrogenobacter* sp.	AM259501	97	Aquificae
T92p60	*Hydrogenobacter* sp.	AM259501	97	Aquificae
T92p67	Uncultured nanoarchaeote	EF562618	99	Nanoarchaeota
T92p74	*Hydrogenobacter* sp.	AM259501	98	Aquificae

NR: No rank.

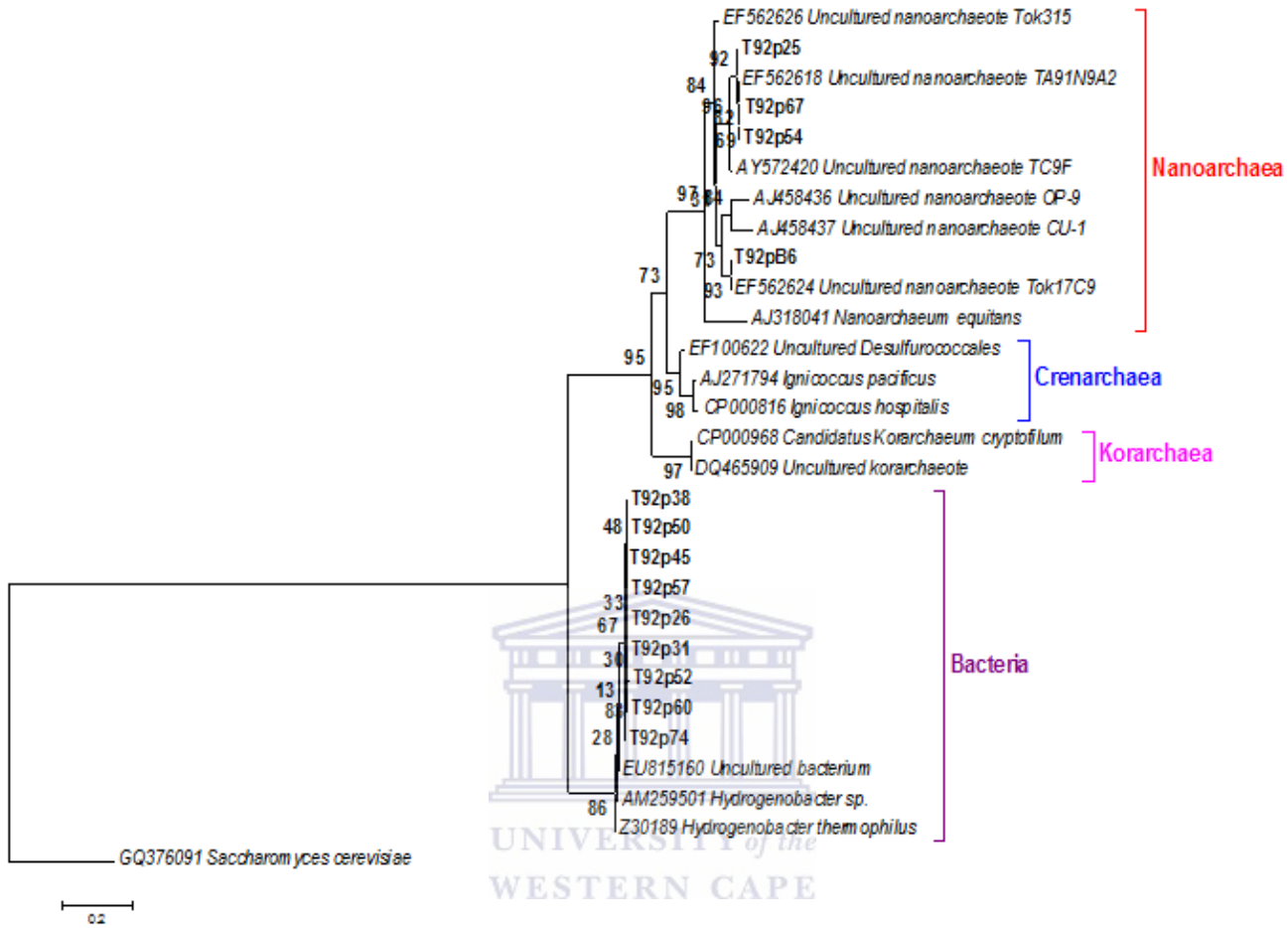

Figure 3.13: Phylogenetic tree constructed using MEGA 4 based on 404 bp 16S rRNA gene sequences amplified with the A571Fb/N989R primer set. The evolutionary history was inferred using the neighbor-joining method. The evolutionary distances were computed using the maximum composite likelihood method; the scale bar represents the number of base substitutions per site. The numbers at branch points indicate bootstrap values of 1000 replicates. Phylotypes isolated in this study are indicated in bold letters.

The A571Fb/N989R primer combination was initially designed for specific amplification of nanoarchaeal phylotypes (Casanueva *et al.*, 2008). However, the use of the A571Fb/N989R primer set in this study has shown its ability to amplify a broader phylogenetic spectrum than was previously thought, including members of the Euryarchaea and Bacteria (Table 3.6).

The amplification of non-nanoarchaeal sequences by the A571Fb/N989R primer set was unexpected, especially when the stringency of PCR annealing conditions was not changed, thus suggesting the possibility of mispriming. Whether the non-nanoarchaeal sequences were identified as a result of mispriming or compromised primer specificity due to novel sequence discoveries, this can only be resolved by continuous assessment of the primer sets (Baker & Cowan, 2004). Consequently, this could aid in escalating discoveries of novel microorganisms and also minimize bias associated with analysis of microbial diversity.

The variation of A571Fb/N989R priming sites among the different taxonomic groups amplified was analysed as shown in Figure 3.14. Not much variation was observed within the A571Fb primer, with the exception of T92p44 (*Hydrogenobacter*), T92A11 & T92B7 (uncultured nanoarchaeotes), and *E. coli*, which showed 56%, 56% and 28%, respectively. This was somehow expected, as the A571Fb primer was designed as a "universal" primer with the ability to amplify Crenarchaea, Euryarchaea, Nanoarchaea and Korarchaeota (Baker *et al.*, 2003). In contrast to the A571Fb primer, the N989R primer revealed significant variation, amounting to 61%, 72% and 50% variation in the same comparison of sequences. While the possibility of mispriming cannot be ruled out, the quality of sequences retrieved from samples from the TAT9-2 site, for which both forward and reverse primers could be identified, justify the potential of the A571Fb/N989R primer pair to amplify not only Nanoarchaea but also Archaea and Bacteria. Given the "universal" effect of the A571Fb primer, the specificity of N989R primer becomes compromised.

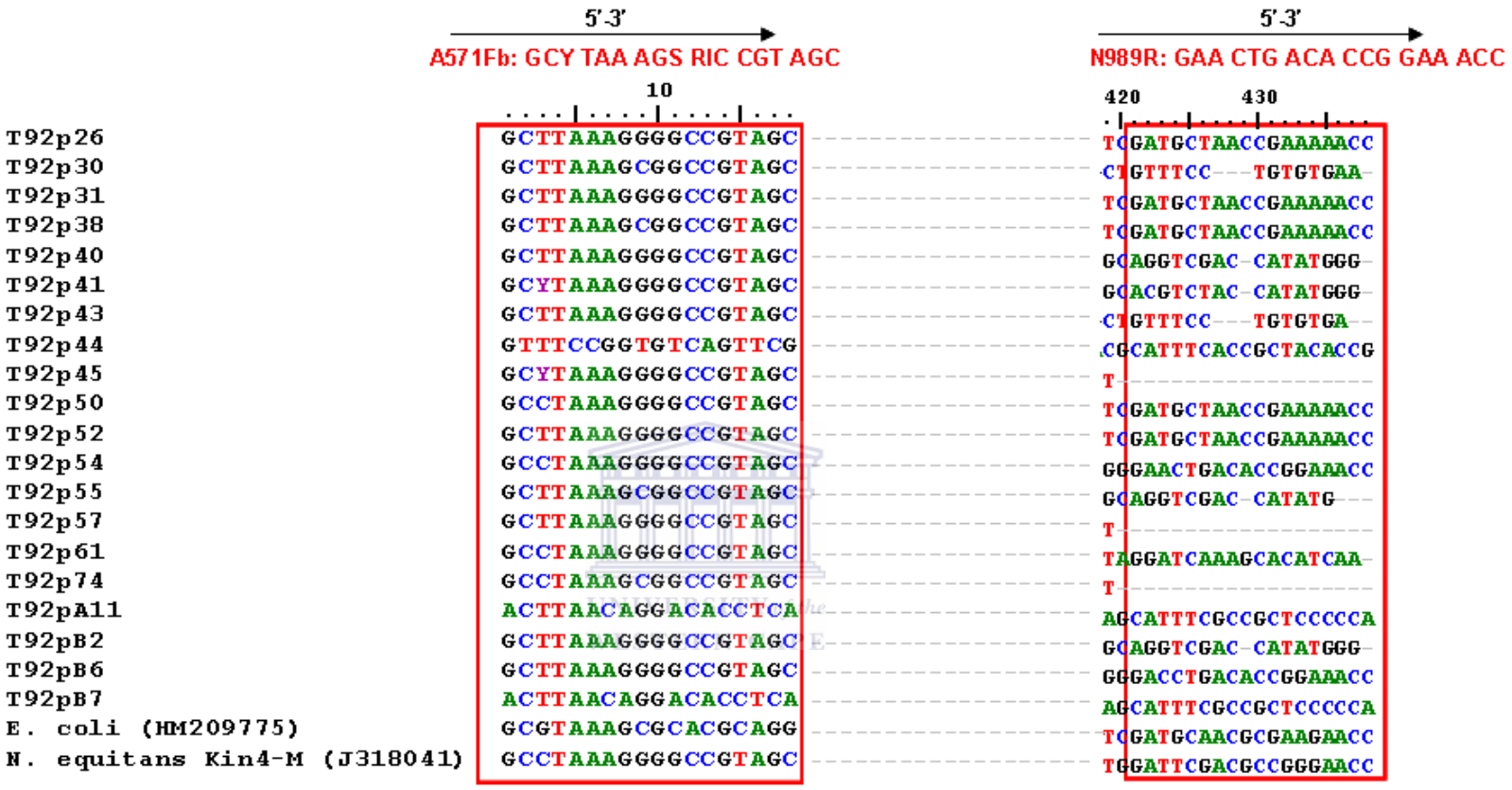

Figure 3.14: Alignment of 16S rRNA gene sequences to identify variation in the A571Fb and N989R priming sites.

3.6 Discussion and conclusion

The diversity of the microbiota from the ETGF sampling sites has been investigated by comparative phylogenetic analysis. This study presents the first El Tatio report on systematic microbial diversity analysis and characterization using a collection of archaeal, nanoarchaeal and bacterial primer sets. Irrespective of the challenges encountered, such as limited samples, low DNA yields and unsuccessful amplification for some of the samples, several attempts to improve DNA yield and amplification were conducted. These include use of Taguchi PCR optimizing algorithm, rigorous DNA extraction methods, etc. Consequently spurious amplification was greatly reduced and successful amplification was obtained using DreamTaq™ (Fermentas) with an optimized buffer which has most likely contributed in chelating PCR inhibitors. As a result of these improvements in amplification, significant microbial diversity was observed in the ETGF samples.

A combination of PCR-based molecular analysis tools, such as 16S rRNA gene clone libraries, ARDRA and DGGE, were used to compare microbial diversity and distribution from seven sites within the geyser field. Archaeal, nanoarchaeal and bacterial 16S rRNA genes were successfully amplified and analysed based on clone libraries and DGGE profiles. The archaeal library was divided into two based on primer selectivity, namely the A571Fb/UA1204R and A3Fa/Ab927R libraries. The nanoarchaeal library was constructed using the A571Fb/N989R primer set. The DGGE fingerprints were constructed on the basis of archaeal and bacterial analysis corresponding to 340F/533R and 341F/534R primer sets, respectively. Analysis of different replicate samples with varying physicochemical parameters from the geyser field revealed variations in the composition of the archaeal and bacterial communities. The DGGE

archaeal sequence results corresponded to those from the clone libraries, with additional representatives observed. These included *I. hospitalis*, uncultured *Vulcanisaeta* sp., *N. xinjiangense* and *H. limicola*.

The microbial diversity from the clone libraries differed significantly, with Nanoarchaea and Haloarchaea dominant in the A571Fb/UA1204R library. Desulfurococcales accompanied by Haloarchaea were dominant in the A3Fa/Ab927R library. Whereas, the A571Fb/N989R library was characterized by Nanoarchaea, similar to those previously detected from New Zealand and the ETGF (Casanueva *et al*., 2008). Microbial sequences retrieved from DGGE fingerprints corresponded with those from the clone libraries, and were also characterized by Desulfuroccales, Haloarchaea and a mixed pool of mesophilic bacteria. Overlapping microbial distribution was observed across the samples, and suggests that the microorganisms may have originated from common sources, thus resulting in the cosmopolitan microbial diversity observed. It was also evident from the sequence results that temperature and salinity were the most important physicochemical conditions contributing to the composition of the microbial community of the ETGF. These results reflect a cosmopolitan microbial diversity, which also imply the microbiota's ability to adapt to a variety of temperature, pH, salinity and redox gradients. The source of the ETGF is thought to be a mixture of hot magmatic brine (~185 000 mg/L Cl^- and 105 000 mg/L Na^+ at 190 °C with dilute, steam-heated meteoric water (Landrum *et al*., 2009), which explains the dominance of haloarchaea and hyperthermophiles in this geyser field.

What has made the ETGF a special research focus for study is the fact that it is one of the least investigated geothermal systems, especially in the field of microbial ecology (Fernandez-

Turiela *et al.*, 2005). Most of the research conducted has focused on geology and geochemistry for mining purposes, and the search for geothermal resources for power generation (Fernandez-Turiela *et al.*, 2005). None of the studies conducted at the ETGF have investigated its overall microbial diversity, even though there has been evidence of microbial activity in the site. As a result, very little was known about the microbial composition of the ETGF, which was the basis of the novelty of this study. Furthermore, previous studies reported only a small amount of microbial-related data, which are qualitative and quantitatively outweighed by the data presented in this study.

Formations of "silica oncoids" made up of amorphous silica around the ETGF hot springs have been reported (Jones and Renaut, 1997). These silicious grains are termed "silica oncoids" due to the consequences of microbial intervention, which includes the presence of filamentous microbes, mucous and spores. Cyanobacterial mats, *Chloroflexus*-like filaments, and *Synechococcus*-like microbes associated with siliceous sinter deposits were also detected (Fernandez-Turiela *et al.*, 2005). These studies were based on microscopic observations with no evidence from molecular analysis, and have concluded that the ETGF microbial composition is moderately diverse and characterized by Cyanobacteria and Diatoms (Fernandez-Turiela *et al.*, 2005). Only Aquificales and Nanoarchaea were detected through molecular analysis (Ferrera *et al*., 2007; Casanueva *et al*., 2008).

The physicochemical properties of the ETGF are made complex by the chemistry of the volcanic activities which characterize the site as a NaCl-rich geyser field. This includes high concentrations of Ca, Li, As, and Cs as well as 95% of gases made up of CO_2 and H_2S as one of the main gases. The silica-rich saline waters of the ETGF are thus a significant source of

toxic elements such as As in the form of arsenate (AsO_4^{3-}) and Sb in the form of antimony trioxide (Sb_2O_3) (Landrum *et al*., 2009). Such toxicity is thought to have a negative effect on the proliferation and distribution of microorganisms in this field, and thus results in only moderate microbial diversity. Microbial adaptation to mitigate toxic aqueous arsenic has been reported; only a few microorganisms are known to utilize arsenic as an energy source (Silver & Phung, 2005; Oremland & Stolz, 2003). While this study did not investigate the metabolic activities of microorganisms at the ETGF, it is likely that even the toxic elements are involved in metabolic processes while, on the other hand, the microbial community has a high tolerance of the toxic elements present. The high levels of silica, NaCl and other compounds also affected DNA extraction as high precipitation of these components was observed during DNA extration. These elements most likely contributed as PCR inhibitors in the unsuccessful PCR reactions

A number of mesophiles and thermophiles were observed from both bacterial and archaeal representatives. These could be attributed to a variety of temperature, pH, salinity and redox gradients, and suggest the presence of complex microbial populations. Most geothermal springs harbour diverse microbial communities, which change along geochemical gradients. Indeed, the findings presented here correspond with several studies which have reported microbial isolates from habitats whose environmental conditions deviated from their optimum growth conditions. Some of these microorganisms are able to survive latently in these foreign habitats, while others can adapt and grow optimally. The indigenous source and ecological significance of some microbial representatives may not be clearly understood. However, it has been suggested that hydrological shifts and subsurface waters may play a role in shaping

geothermal microbial communities, thereby becoming a conduit for the dispersal of the microorganisms (which may originate elsewhere) within the geothermal habitats (Tin *et al*., 2011).

Given the dynamics of microbial evolution, it is generally acceptable that various factors such as natural selection, fitness, adaptive evolution and mutations contribute significantly in shaping the microbial diversity of an environment (see Section 1.2). Such factors can be driven by natural events, including interaction between microbial species. How the non-indigenous members of the community affect the relationship of indigenous members will depend on the antagonistic or synergistic interaction mode between these two groups. Furthermore, the needs, for example the metabolic requirements, could also determine the nature of interactions. In addition, due to fitness, the indigenous microorganisms are more likely to have a competitive advantage as environmental conditions would be more favourable for optimum growth compared to their invading counterparts who have to adapt for survival. Thus, irrespective of the above, the ETGF remains an interesting environment to harbour diverse microbial communities and both PCR-DGGE and 16S rRNA gene clone libraries have been useful tools for investigating and revealing the microbial consortium in this biotope.

The 16S rRNA gene analysis of microbial population of the ETGF has greatly expanded our knowledge of microbial diversity and distribution in this geyser field. The microbial consortium analysed in this study was moderately diverse, with a total of only 30 OTUs identified from both clone libraries and PCR-DGGE fingerprints. However, these results show only a limited coverage of the diversity, which may not necessarily be a true reflection of the total *in situ* microbial diversity. A high proportion of microbial representatives have certainly

been left out or remain undetected due to bias associated with sample preservation, DNA extraction, PCR and cloning, among other things. Furthermore, use of advanced screening techniques, such as high-throughput sequencing, would be able to refine these findings further, such that absent taxa could then be identified in contrast to those missed by gene amplification.

Chapter 4

Comparative statistical analysis of microbial diversity from the El Tatio geyser field

4.1 Introduction

Statistical analysis of microbial diversity is fundamentally comparative. It provides a basis for a more objective and informative interpretation of the diversity, that is not only limited to what can be observed but also gives detailed statistically significant, logical and insightful analysis (Fromin *et al.*, 2002). To date, a variety of statistical analysis tools have been developed and used to identify statistically significant microbial groups and relationships (Smalla *et al.*, 2007). These include multivariate statistical tools, such as principal component analysis (PCA), multi-dimensional scaling (MDS), self-organizing maps (SOM), and additive main effects and multiplicative interaction (AMMI). The variations among the microbial communities have also been linked to physicochemical differences using canonical correspondence analysis (CCA) and redundancy analysis (RDA) (Blackwood & Paul, 2003; Magalhaes *et al.*, 2008). Species richness has been estimated using parametric and non-parametric richness estimation methods, such as Chao1 and ACE (Kemp & Aller, 2003). For the purposes of this study only PCA, CCA, MDS and species richness estimators will be discussed.

The data required for analysis of microbial diversity and ecology can be generated using various tools and techniques, including metagenomic clone libraries and fingerprinting

techniques such as PCR-DGGE, terminal-restriction fragment length polymorphism (T-RFLP) and automated ribosomal intergenic spacer analysis (ARISA).

Clone libraries generated from the 16S rRNA gene amplicons, including fingerprinting techniques such as PCR-DGGE, have been used extensively for studying microbial ecology from various habitats including the YNP hot springs and hydrothermal vents (Ferris *et al*., 1996; Hamamura *et al*., 2005; Lau *et al*., 2009; Muyzer *et al*., 1995). PCR-DGGE allows simultaneous analysis of multiple samples, thereby permitting comparison of microbial community fingerprints, which are viewed from the banding patterns, and can also be used for statistical analysis. The banding patterns generated reflect the state of the microbial community and relate more to the relative abundance of the dominant members of the population than to the total richness (Muyzer & Smalla, 1998). The presence of the bands can also be evaluated in relation to the effects of the physicochemical conditions of the ecosystem from which samples were collected. The total number of bands can then be used as a measure of species richness and the dominant phylotypes can be identified using band intensity (Müller *et al.*, 2001; Widmer *et al.*, 2001).

T-RFLP is used to analyse polymorphic lengths of PCR amplicons digested with restriction enzymes, hence the name terminal-restriction fragments (TRFs). ARISA makes use of the highly intergenic region between the 16S and 23S rRNA genes in the rRNA operon (ITS). ITS is characterized by the variability in the length and nucleotide sequence; different bacterial genotypes can be discriminated on the basis of the migration distance (Okubo & Sugiyama, 2009). T-RFLP and ARISA have been suggested to be more sensitive than PCR-DGGE and

have been shown to detect even less abundant bacterial species (Hewson, & Fuhrman, 2004; Marsh, 1999). However, only 16S rRNA gene clone libraries and PCR-DGGE fingerprints data could be generated, which form the basis of the statistical data analyzed in this study.

4.2 Statistical analysis for 16S rRNA gene clone libraries and PCR-DGGEfingerprints

4.2.1 Principal component analysis (PCA)

Principal component analysis (PCA) allows the reduction of large data sets to smaller linear combinations which can be viewed in Euclidean or ordination plots. A proportion of variance accounted for by the principal components (variables) can then be viewed in these plots; the more closely the samples appear in the plots, the more they are related. PCA has found application in various microbial ecology data sets. These include analysing seasonal changes in the microbial diversity of different geographical biotopes (Merrill & Halverson, 2002). PCA can be used to visualize relationships among microbial community fingerprints and to find hypothetical variables or components which are responsible for variance in multidimensional data sets (Schwartz *et al.*, 2007). PCA has also been used, to establish relationships between the seasonal, spatial PCR-DGGE banding profiles and temperature ranges of samples from the Ranong hot spring in Thailand, and between microbial community composition, activity and geochemical variables (Sompong *et al.*, 2006).

4.2.2 Multi-dimensional scaling (MDS)

Like PCA, multi-dimensional scaling (MDS) is a data reduction technique that has been used to interprete complex data sets. MDS reduces complex data to a point in a two or three dimentional space. It has been used to visualise relationships among samples based on fingerprinting profiles such as those derived from PCR-DGGE and RFLP analysis (Terahara *et al.*, 2004). The closer the points representing the samples to one another, the more similar are the corresponding samples.

4.2.3 Canonical correspondence analysis (CCA)

Canonical correspondence analysis (CCA) models the species response to environmental variations using mathematical simulations. The model then enables interpretation of the abundance and occurrence of species based on environmental impact. CCA is most appropriate to test the correlation between changes in the community compositon and the environment (Cao *et al.*, 2006). It has been used in a variety of analyses of microbial assemblages, where the environmental conditions influencing microbial diversity were determined. A few examples include the use of CCA in fingerprinting data sets such as those derived from ARISA (Yannarell & Triplett, 2005), PCR-DGGE (Salles *et al*., 2004; Sapp *et al*., 2007), and T-RFLP (Córdova-Kreylos *et al*., 2006; Klaus *et al*., 2007). Furthermore, CCA has been used to determine specific species or OTUs that respond to particular environmental variables (Yannarell & Triplett, 2005), and using those species as environmental indicators, for example of pollution (Córdova-Kreylos *et al*., 2006). Furthermore, interactions between various abiotic

and biotic factors have been analysed using CCA in order to explain microbial dynamics in marine ecology and subsequently in identifying those species that responded to those factors (Sapp *et al*., 2007).

4.3 Estimating microbial diversity

Several statistical methods are available for estimating microbial diversity; these include rarefaction curves (Dunbar *et al.*, 1999; Moyer *et al.*, 1998), species richness estimators (Kroes *et al.*, 1999; Nübel *et al.*, 1999a), and diversity indices (McCaig *et al.*, 1999; Nübel *et al.*, 1999b). In addition to PCA and MDS, use of diversity indices, richness and evenness estimators provides additional insight into the similarities and dissimilarities of microbial communities (Schütte *et al.*, 2008).

4.3.1 Rarefaction and richness estimators

A variety of diversity comparisons and richness estimation approaches have been developed, such as rarefaction curves and richness estimators (Bills & Polishook, 1994; Dunbar *et al.*, 1999; Moyer *et al.*, 1998). Rarefaction curves plot the cumulative species richness as a function of the number of phylotypes or individuals sampled. This gives an estimate of the expected number of species based on randomly collected samples (Crist & Veech, 2006). However, this approach has the potential to generate incorrect estimates of relative species richness when sample sizes are small (Hughes & Hellmann, 2005).

Microorganisms have not been sampled exhaustively as yet; as a result, species richness estimators such as Chao1 and ACE were derived. These estimators predict the number of species in a sample based on the rare species sampled (Colwell & Coddington, 1994), and are discussed below.

4.3.2 S_{Chao1} and S_{ACE}

Parametric and non-parametric richness estimators have been widely used for OTUs comparisons between environments (Bohannan & Hughes, 2003). Non-parametric richness estimators such as Chao1 and ACE estimate unobserved OTUs in a community. These estimators consider the proportion of species that have been observed beforehand in relation to those observed only once. The disadvantage with Chao1 and ACE is that they both underestimate the true richness for small samples. Species richness estimators also take into account the shape of the cumulative curve to determine total richness (Hughes *et al.*, 2001). In theory, these estimators can predict relative total richness and are useful for assessing microbial diversity (Krebs, 1989; Seber, 1973). S_{Chao1} is calculated as;

$$S_{Chao1} = S_{obs} + \frac{F_1^2}{2(F_2+1)} - \frac{F_1 F_2}{2(F_2+1)^2}$$

where S_{obs} is the number of species observed in the community, and F_1 and F_2 are the numbers of species occurring either once or twice. S_{Chao1} is suitable for libraries or communities where species are relatively rare.

S_{ACE} is calculated as;

$$S_{ACE} = S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{F_1}{C_{ACE}} \gamma_{ACE}^{2}$$

where F_1 is the number of species occurring only in the community, S_{rare} is the number of species occurring 10 times or fewer, S_{abund} is the number of species occurring >10 times, and γ_{ACE}^{2} is the coefficient of variation of F_1. C_{ACE} is an estimate of sample coverage, defined as the proportion of individuals of relatively rare species (<10) that occur more than once in the community or library (Chao *et al.*, 1993). C_{ACE} is also appropriate for data sets where the presence of some phylotypes or species is more frequent, and S_{Chao1} and S_{ACE} mostly correlate with one another when most species occur mostly once or twice (Kemp & Aller, 2003).

The disadvantage of parametric species richness estimators is that they require large data sets compared to their non-parametric counterparts. Generally, species richness estimates are dependent on community composition, which may not always reflect the true representation of the microorganisms in the sample (Kemp & Aller, 2004). Both sampling errors and PCR artefacts, such as mismatches and mispriming, can also affect the library composition, thereby creating inaccuracy and bias in the estimates (Reysenbach *et al.*, 1992; Suzuki & Giovannoni, 1996). Measures to test bias in species richness estimates can be difficult to implement in microbial diversity (Kemp & Aller, 2004). Such tests would require comparison of the true richness against sample estimates, which can be possible only if the community has been adequately sampled. However, bias can be minimized by the use of multiple techniques in order to increase the robustness of comparative analysis (Nübel *et al.*, 1999a). As mentioned

previously, high-throughput sequencing is potentially the most useful technique to minimise most of the bias introduced by PCR and clonal-based methods. Furthermore, the robust high-throughput sequencing data guarantee stable estimates, because richness estimates mostly favour large data sets.

It is generally agreed that a large proportion of microbial species have not been cultured and no microbial community has been screened exhaustively (Hughes *et al.*, 2001). In view of the ubiquitous nature of microorganisms, even the statistical methods discussed in this chapter are subject to some level of bias. Nevertheless, there is no doubt that these methods contribute significantly in improving the resolution of species richness.

4.3.3 Diversity indices

Diversity indices measure the species diversity in a community, by providing information about species composition and also by taking into account the relative abundance of different species. Diversity indices enable the reduction of large amounts of data from species abundance into single quantitative expressions (Sager & Hasler, 1969). These are coupled to evenness and dominance measures, where evenness is a measure of the abundance of similar species and dominance indicates the extent to which species dominate in an environment. There are various ways of describing species abundance relationships, and these are expressed by diversity indices, examples of which are outlined in Table 4.1 (Lexerød & Eid, 2006).

Table 4.1: Summary of the diversity indices as described by Lexerød & Eid (2006)

Index	Index value range	Reference
Margalef index	[0,∞]	(Clifford & Stephenson, 1975)
Shannon index	[0,ln(S)]	(Shannon, 1948)
Gini coefficient	[0;1]	(Gini, 1912)
Simpson index	[0;1]	(Simpson, 1949)
Berger-Parker index	[0,1]	(Berger & Parker, 1970)
Shannon evenness	[0,1]	(Pielou, 1969)

Any measure chosen for analysis should emphasize at least one or more components of diversity, whether richness or evenness. It is also crucial to have a clear understanding of the objectives of the study before choosing the appropriate diversity index and to ensure proper sampling for the purpose of the study.

Shannon's and Simpson's diversity indices, which are particularly relevant to this study, are described here. Both Shannon's and Simpson's indices are less sensitive to the sampling effort and respond more to the evenness of species distribution, while increasing with increasing diversity. This combination of indices is based on their simplest measures, which rely on the total number of individual species represented in the sample. Most diversity indices are coupled with an evenness index. Shannon's and Simpson's evenness accounts for the observed value of the diversity index divided by the number of species in the sample. The closer the evenness index is to 1, the more evenly distributed are the species in the sample.

4.3.3.1 Shannon's index

Shannon's index was originally developed in information theory (Shannon & Weaver, 1949), and later gained application in ecological diversity (Margalef, 1957). Shannon's index is also referred to as the Shannon-Wiener or Shannon-Weaver index as a result of the joint authorship of Shannon and Weaver's book, including Shannon's precursory research for Wiener's work in the late 1940s (Spellerberg & Fedor, 2003; Shannon & Weaver, 1949). The 'Shannon index' or 'Shannon & Wiener index' have been suggested as the most appropriate terms to use in this context (Spellerberg & Fedor, 2003). Shannon's index accounts for both the abundance and evenness of species in a community (Spellerberg & Fedor, 2003). It is calculated as follows:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i - [(S-1)/2N]$$

where S is the total number of species, n_i is the number of individuals belonging to the *i*th species, N is the total number of individuals and p_i is n_i /N, the proportion of individuals of the total sample belonging to the *i*th species. The equation is simplified as:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i$$

When H′ = 0, it means there is one species in the sample. Shannon's index has been used extensively as a benchmark measure for diversity analysis.

4.3.3.2 Simpson's index

Simpson's index was the first diversity measure proposed (Simpson, 1949). The index measures the probability that two phylotypes picked at random belong to different species (Krebs, 1978). Simpson's index is calculated as:

$$D = \sum_{i=1}^{S} p_i^2$$

where S and p_i are as described for Shannon's diversity index (see Section 4.3.3.1). Values near zero correspond to a more diverse or heterogeneous community, whereas those near 1.0 imply a more homogeneous community. However, this index is commonly expressed as:

$$D = 1 - \sum_{i=1}^{S} p_i^2$$

This expression enables the value of D to range from 0 to 1; D is directly proportional to diversity and increases as the assemblage becomes more even. Simpson's index is one of the most meaningful and robust diversity indices available, capturing the variance of pecies abundance (Magurran, 2004). In contrast to Shannon's diversity index, Simpson's index performs well for both general diversity statistics and as a measure of evenness. A separate measure of evenness ($E_{1/D}$) can be calculated by dividing the reciprocal form of Simpson's diversity index by the number of species present (Krebs, 1999; Smith & Wilson, 1996). In this case, the evenness index should be at maximum when all species are equally abundant, and decrease towards zero as the relative abundances diverge from evenness (Meerman, 2004).

4.4 Aims and objectives

The aim of this chapter was to statistically analyze the characterized microbial diversity of the ETGF using a combination of statistical tools traditionally used by microbial ecologists. These include cluster analysis, MDS, PCA, CCA, estimates of species richness and diversity indices. The rationale behind this analysis was to compare and simultaneously assess the relative microbial diversity of samples from different ETGF sites. Furthermore, it was to evaluate the correlation of the microbial diversity with changes in environmental conditions such as pH, temperature, salinity, conductivity and total dissolved solids.

4.5 Statistical analysis of ETGF microbial community composition based on 16S rRNA clone libraries

The composition of the El Tatio microbial community was determined by sequencing representatives from the constructed 16S rRNA gene clone libraries and PCR-DGGE bands as described in Chapter 3. The composition of archaeal, nanoarchaeal and bacterial representatives was calculated as a proportion of the total phylotypes sequenced (Figure 4.1).

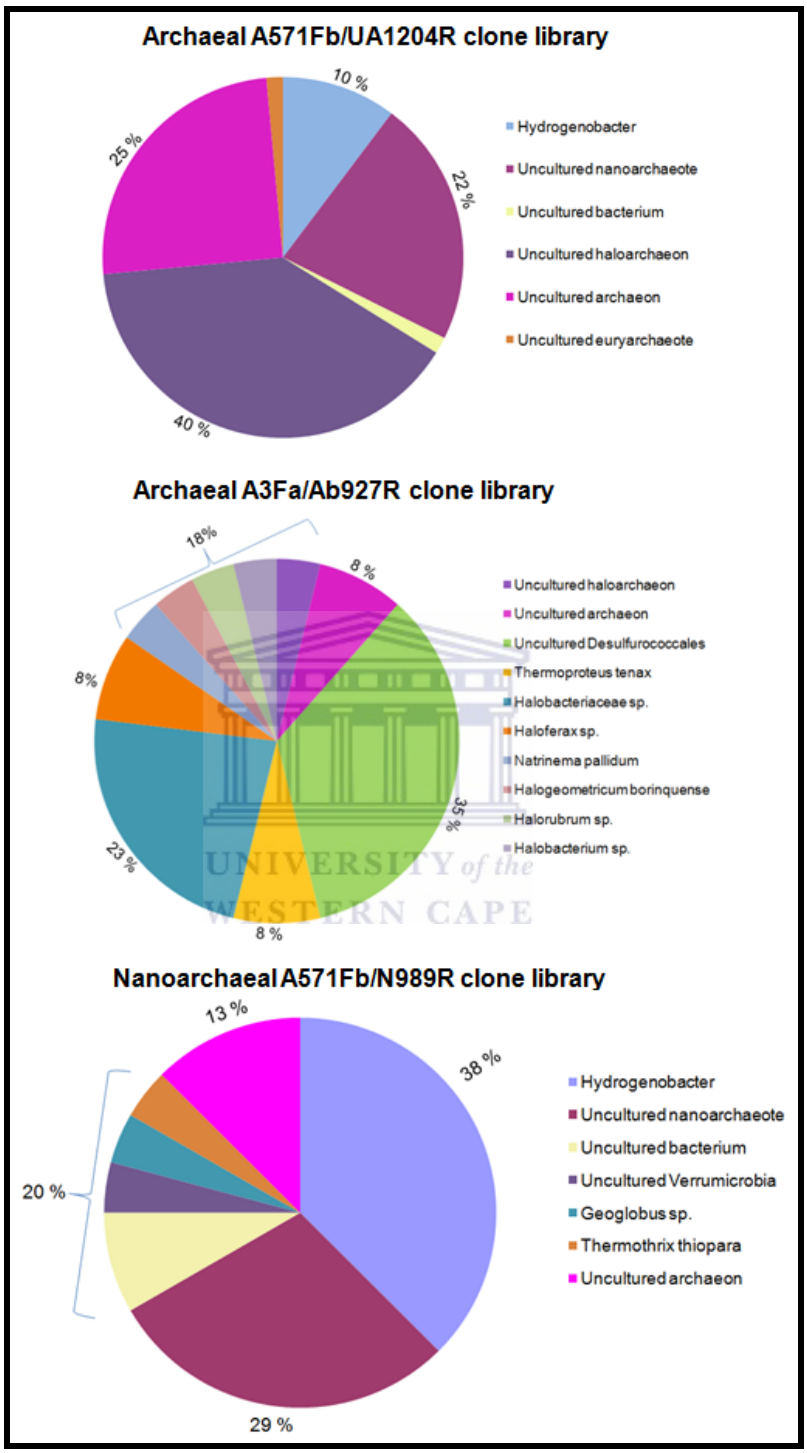

Figure 4.1: Pie chart diagrams displaying the percentage composition of archaeal, nanoarchaeal and bacterial phylotype representatives from the 16S rRNA gene clone libraries.

4.6 Statistical estimation of species richness and library coverage

To determine the quality of the clone libraries and whether they were screened sufficiently to generate stable phylotype richness estimates, and to make rigorous statistical comparisons, non-parametric total species richness and abundance-based coverage estimators S_{Chao1} and S_{ACE} were calculated using a web-based model: (http://www.aslo.org/lomethods/free/2004/0114a.html) (Kemp & Aller**,** 2004). The S_{Chao1} and S_{ACE} richness estimators were supported by C_{ACE} and Good's coverage estimators based on the number of sequenced representatives. These coverage indices were used to estimate the proportion of phylotypes present in the samples. Figures 4.2-4.6 show the behaviour of S_{Chao1} and S_{ACE} richness estimators, and C_{ACE} and Good's coverage estimators on ETGF 16S rRNA gene clone libraries.

All of the libraries analysed reached an asymptote in at least one of the richness estimators used, and were considered 'large enough' to yield unbiased estimates. S_{Chao1} was more stable (it assumed a more distinct asymptotic curve) than S_{ACE}, which seemed to be more sensitive to the low sampling effort (small library size).

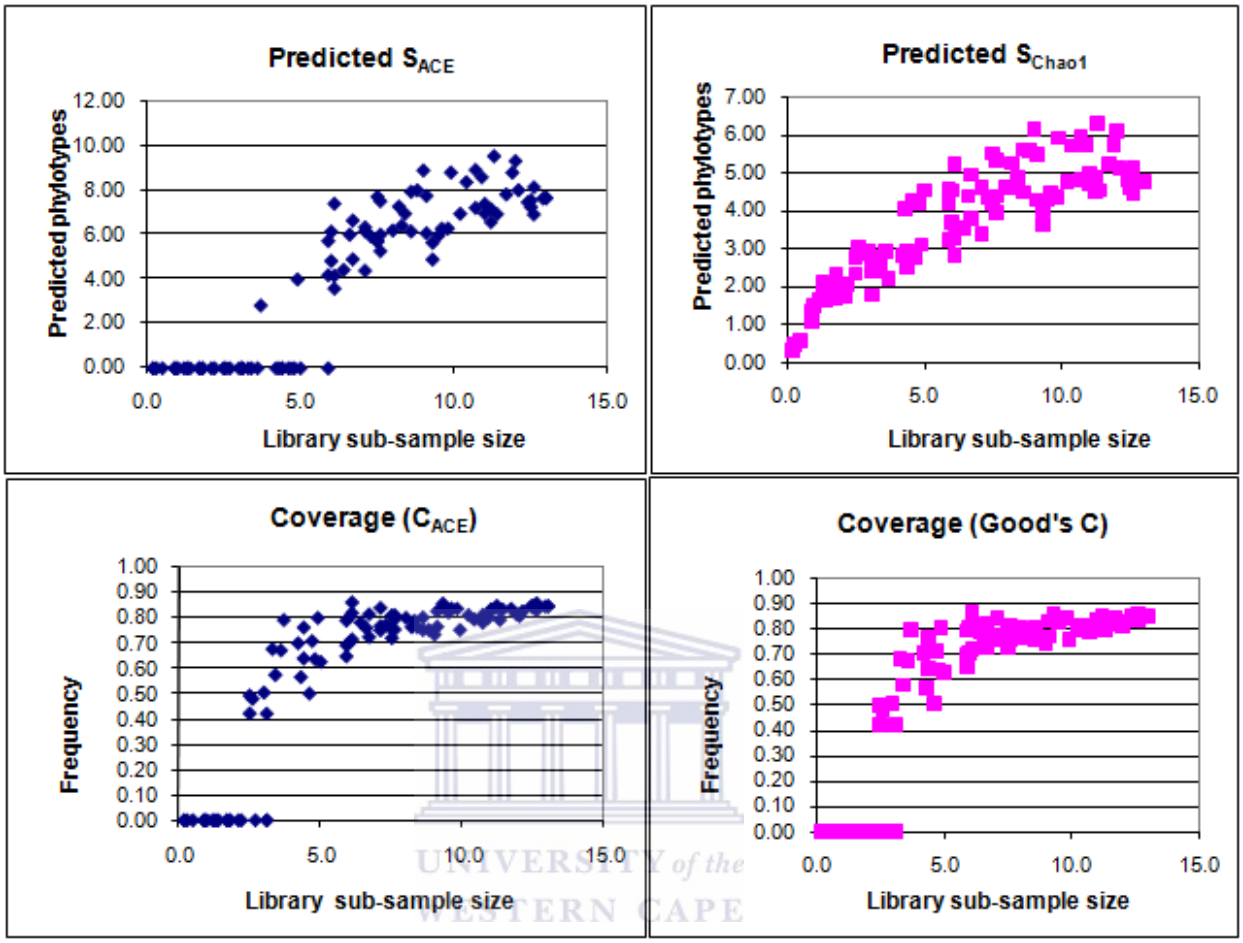

Figure 4.2: Species richness estimates of the TAT6 sample based on S_{ACE} and S_{Chao1}, and supported by C_{ACE} and Good's coverage estimators. The estimates of richness and coverage were plotted against the library sub-sample size. Each point represents the mean of 10 replicate sub-samples of the A3Fa/Ab927R sequence data.

While more than 50% of the sequenced representatives from all the clone libraries were assigned to OTUs, the estimated species richness was more than the observed estimates for both S_{ACE} andS_{Chao1}. This was only for 41-71% (C_{ACE}) and 65-83% (Good's) sample coverage of the total diversity (Table 4.2). Considering that the actual species diversity in the samples may have been in millions, even though most microorganisms have not yet been identified or

even detected, the estimates clearly indicate underestimation of the total species richness, irrespective of the doubling capacity estimated compared to the observed species richness. However, it is important to note that these estimates are standardized based on the sample size, and were conducted on limited sequenced representatives. Therefore, the ability to obtain stable estimates at such small size of the clone libraries was a positive remark. This implies that a more extensive sampling effort, which could be achieved by high-throughput sequencing, would improve the quality of the estimates.

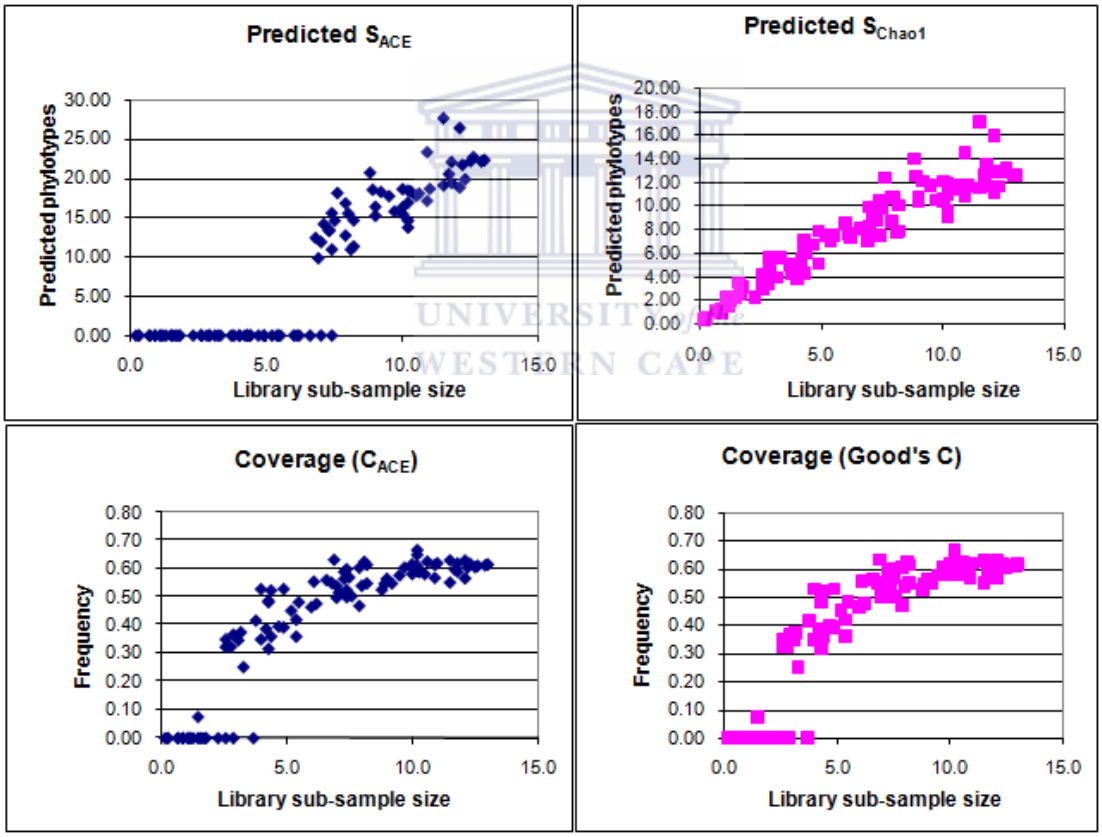

Figure 4.3: Species richness estimates of the TAT6-2 sample based on S_{ACE} and S_{Chao1}, and supported by C_{ACE} and Good's coverage estimators. The estimates of richness and coverage were plotted against the library sub-sample size. Each point represents the mean of 10 replicate sub-samples of the archaeal A3Fa/Ab927R sequence data.

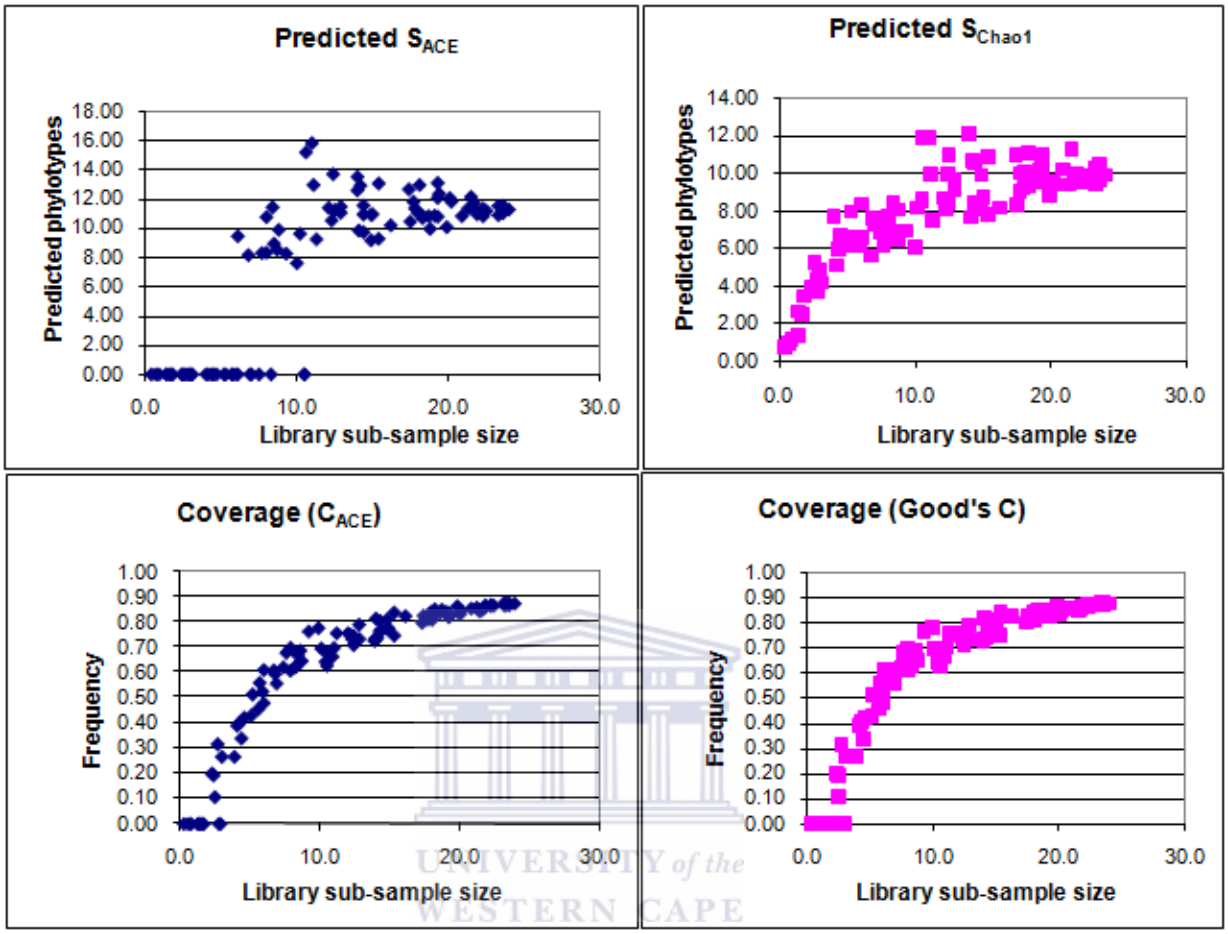

Figure 4.4: Species richness estimates of the TAT9-2 sample based on S_{ACE} and S_{Chao1}, and supported by C_{ACE} and Good's coverage estimators. The estimates of richness and coverage were plotted against the library sub-sample size. Each point represents the mean of 10 replicate sub-samples of the nanoarchaeal A571Fb/N989R sequence data.

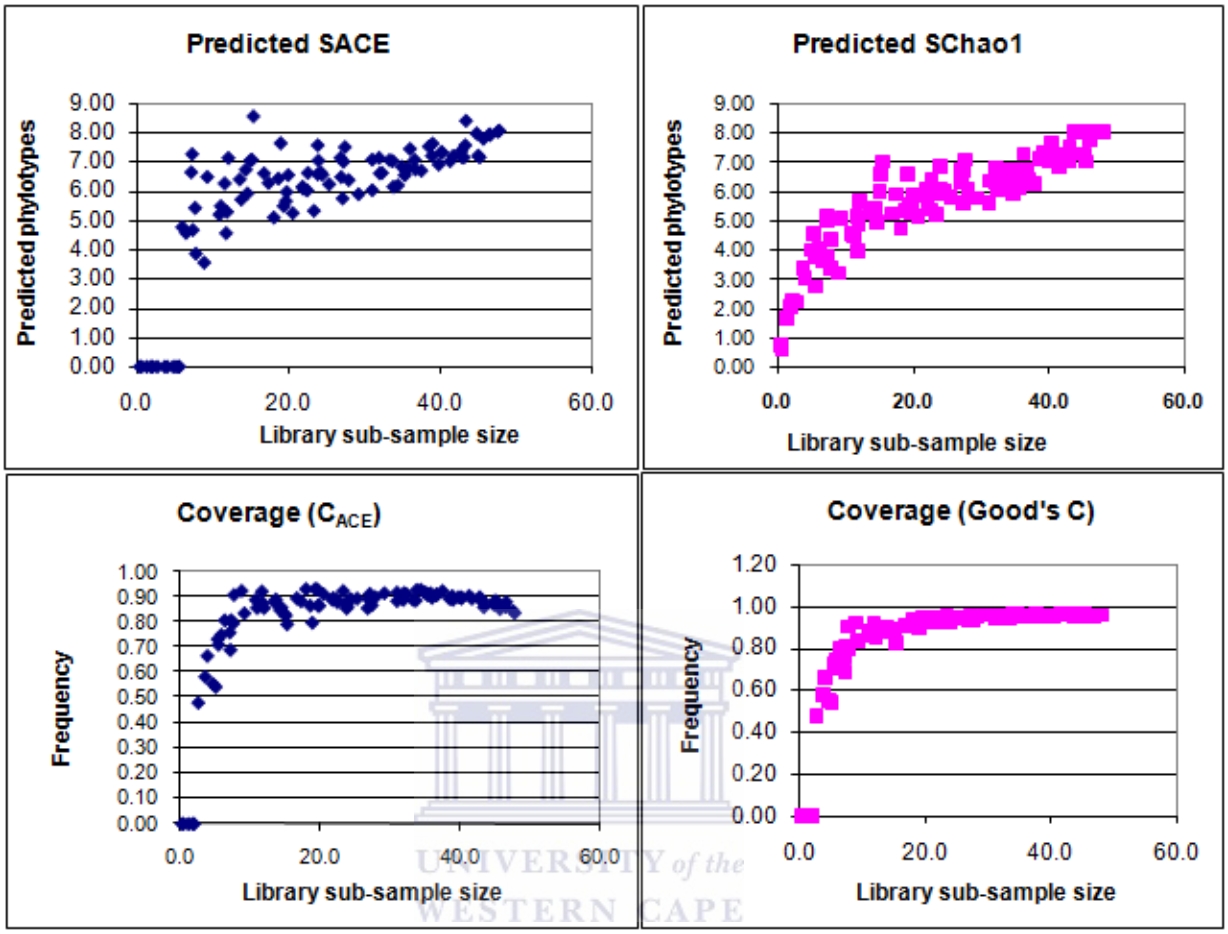

Figure 4.5: Species richness estimates of the TAT2-2 sample based on S_{ACE} and S_{Chao1}, and supported by C_{ACE} and Good's coverage estimators. The estimates of richness and coverage were plotted against the library sub-sample size. Each point represents the mean of 10 replicate sub-samples of the archaeal A571Fb/UA1204R sequence data.

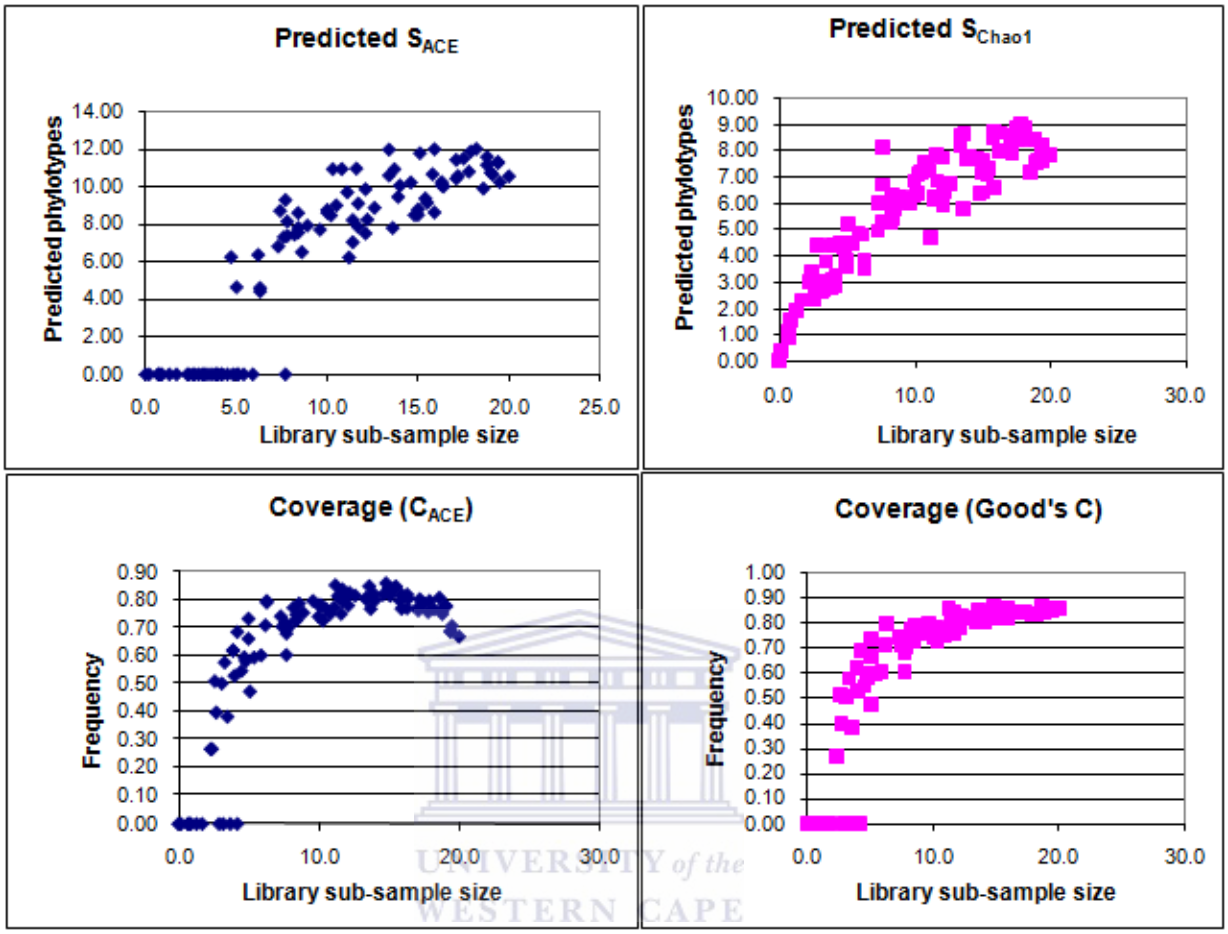

Figure 4.6: Species richness estimates of the TAT6-2 sample based on S_{ACE} and S_{Chao1}, and supported by C_{ACE} and Good's coverage estimators. The estimates of richness and coverage were plotted against the library sub-sample size. Each point represents the mean of 10 replicate sub-samples of the archaeal A571Fb/UA1204R sequence data.

A good estimator is one that remains stable and is independent of sampling effort (Gotelli & Colwell, 2001). S_{Chao1} proved to be such an estimator by showing more stability than S_{ACE} and by not being affected by sampling effort. In this study, S_{ACE} was sensitive to sampling effort, particularly evident when the estimation of species richness was based on the sequenced PCR-DGGE fragments (results not shown). The PCR-DGGE fragments were not well sequenced,

and as a result the coverage of the component samples was poor and did not yield stable richness estimates, but revealed highly uneven frequency distributions and unstable estimates.

Table 4.2 summarizes species richness estimates of clone libraries that describe the relative richness and diversity of the represented species in the microbial communities of the ETGF. Both S_{ACE} and S_{Chao1} were slightly different from each other, with only one sample [TAT2-2 (A571Fb/UA1204R)] showing equal richness estimates (Table 4.2). The richness estimates were both higher than the observed phylotypes, indicating that more species still need to be identified from the samples. Ignoring the estimates from the under-sequenced PCR-DGGE profiles, the coverage estimates from the clone libraries indicated that more than 41% of total diversity was detected in the clone libraries, with a maximum of 83% coverage. These are satisfactory estimates considering the small size of the libraries and the fact that a large number of microorganisms remain undetected. These estimates could be significantly improved by increasing the sampling effort.

The estimation of species richness was useful in showing that the libraries were screened sufficiently to yield stable estimates that were relevant to further statistical analysis and comparison.

Table 4.2: Summary of species richness and coverage estimates based on clone libraries. N denotes the number of sequenced representatives, S is the total number of species, or OTUs, and E denotes evenness (E = H / ln(S)). The ratio of OTU and the value of the estimate is directly proportional to stability.

Clone library	N	S	E	Simpson	Shannon	S_{ACE}	S_{Chao1}	C_{ACE} [a]	Good[b]
TAT6 (A3Fa/Ab927R)	13	4	0.68	0.49	0.94	8	5	54	83
TAT6-2 (A3Fa/Ab927R)	13	7	0.84	0.73	1.63	22	13	41	65
TAT9-2 (A571Fb/N989R)	24	8	0.86	0.79	1.79	11	10	71	82
TAT2-2 (A571Fb/UA1204R)	48	6	0.72	0.65	1.31	8	8	71	75
TAT6-2 (A571Fb/UA1204R)	20	6	0.74	0.64	1.33	11	8	50	68

[a] C_{ACE} coverage percentage $[1-(F_1/N_{rare})] \times 100$.
[b] Good's coverage percentage $[1-(n_1/N)] \times 100$.

Table 4.3: Diversity indices based on the presence (1) or absence (0) of archaeal and bacterial PCR-DGGE bands. S denotes the total number of species in the sample (assuming that each band represents a unique species), N is the total number of bands, and E stands for evenness [E = H / ln(S)]

Sample	**S***	**E**	**Simpson**	**Shannon**
TAT1-1	10	1	0.90	2.302585
TAT2-2	14	1	0.93	2.639057
TAT6	9	1	0.89	2.197225
TAT6-2	12	1	0.92	2.484907
TAT9-1	19	1	0.95	2.944439
TAT10-1	16	1	0.94	2.772589

*S = N (the total number of bands).

4.7 Comparison of microbial distribution based on diversity indices

Shannon's and Simpson's diversity indices were used to determine and compare microbial diversity in the samples from the ETGF sites. The response of the diversity indices to species richness and evenness is shown in Tables 4.2 and 4.3. Table 4.2 shows diversity indices for archaeal (A3Fa/Ab927R and A571Fb/UA1204R) and nanoarchaeal (A571Fb/N989R) 16S rRNA clone libraries. The diversity indices of archaeal and bacterial PCR-DGGE fingerprints are given in Table 4.3.

A total of 118 phylotypes were sequenced from the clone libraries. Both Shannon's and Simpson's indices followed a similar trend, in which those for TAT9-2 (A571Fb/N989R) were found to indicate greater diversity than for the other libraries, trailed by TAT6-2 (A3Fa/Ab927R), TAT2-2 (A571Fb/UA1204R), TAT6-2 (A571Fb/UA1204R) and TAT6 (A3Fa/Ab927R). However, the fact that primer bias may also have contributed to the increased diversity of the TAT9-2 sample cannot be ruled out. Samples TAT6 and TAT6-2 (A3Fa/Ab927R) showed slightly different indices, even though the sampling size was the same, thus suggesting significant differences in these samples, which were not affected by differences in sampling capacity and primer bias. A similar trend was observed with archaeal (A571Fb/UA1204R) libraries, for which differences in sampling capacity did not affect species diversity nor did primer selection.

Comparison of the archaeal and bacterial PCR-DGGE fingerprints showed sample TAT9-1 to be more diverse (0.95) with only bacteria present, followed by TAT10-1 (0.94), TAT2-2 (0.93), TAT6-2 (0.92), TAT1-1 (0.90) and TAT6 (0.89) (Table 4.3). TAT6 and TAT6-2 were the only samples that were amplified with both archaeal (340F/533R) and bacterial (341F/534R) PCR-DGGE primers; the rest of the samples were amplified only with the (340F/533R) primer set. These findings suggest that bacterial species were dominant over their archaeal counterparts. However, this was proved otherwise by the sequencing results, from which only 13 bacterial species were identified (Table 3.1) over a total of 30 species assigned to OTUs, and is thus attributed to amplification bias.

While the sequenced representatives differed considerably among the samples, the species were slightly evenly distributed, with almost all evenness indices approaching 1 (Table 4.2). Diversity indices of PCR-DGGE fingerprints were based on the presence (1) or absence (0) of DNA bands, because only a few bands were sequenced. As before, each band was assumed to represent a unique species. Using the presence or absence approach, both archaeal and bacterial PCR-DGGE fingerprints showed S to be equal to N and thus theoretically the samples approached perfect evenness, i.e. E = 1, which may not necessarily have been true. In this case the evenness indices were ignored (Tables 4.3).

4.8 Multivariate statistical analysis of the ETGF microbial communities

On the basis of the data presented in Chapter 3, the phylogenetic affiliation of the OTUs detected favoured growth at high temperatures and high saline environments. Thus, temperature and salinity seemed to be the significant factors that influenced the microbial diversity of the ETGF samples. In order to determine the variance in these samples and the effect of environmental conditions on the distribution and abundance of microbial species, PCA and CCA plots were constructed. PCA plots were useful in revealing relations between the sample variables, whereas the CCA plots were effective in determining relationships between the microbial community and the physicochemical factors. An analysis of variance (ANOVA) was used to test for statistically significant differences of microbial distribution within the samples. The ANOVA test is normally used to compare data and to test the hypothesis of no difference between the means of two or more treatment groups or samples (Bewick *et al*., 2004).

4.8.1 Principal component analysis of the ETGF microbial diversity

PCA was performed on the 16S rRNA clone libraries and PCR-DGGE fingerprint data to determine the degree of variance among the microbial communities at the various ETGF sites. The PCA plot of sequenced species from both 16S rRNA clone libraries and PCR-DGGE fingerprints data account for 62.7%, with PC 1 explaining 37.1% and PC 2 25.6% of the samples variance, respectively (Figure 4.7). Only 37.3% could not be explained by the

principal components, which is a positive remark. Based on the PCA plot, TAT1-1 and TAT10-1 were closer to each other along the PC 1 axis. This relatedness was based, however, on the distribution of species (one species on each sample) and not on similarities of species. Samples TAT2-2, TAT6, TAT6-2 and TAT9-2 were all scattered further apart from each other across the PCA plot, indicating significant differences ($P < 0.05$) of microbial diversity on ETGF samples (Appendix 1C). These results were also confirmed by the correlation tests as shown in Appendix 1D.

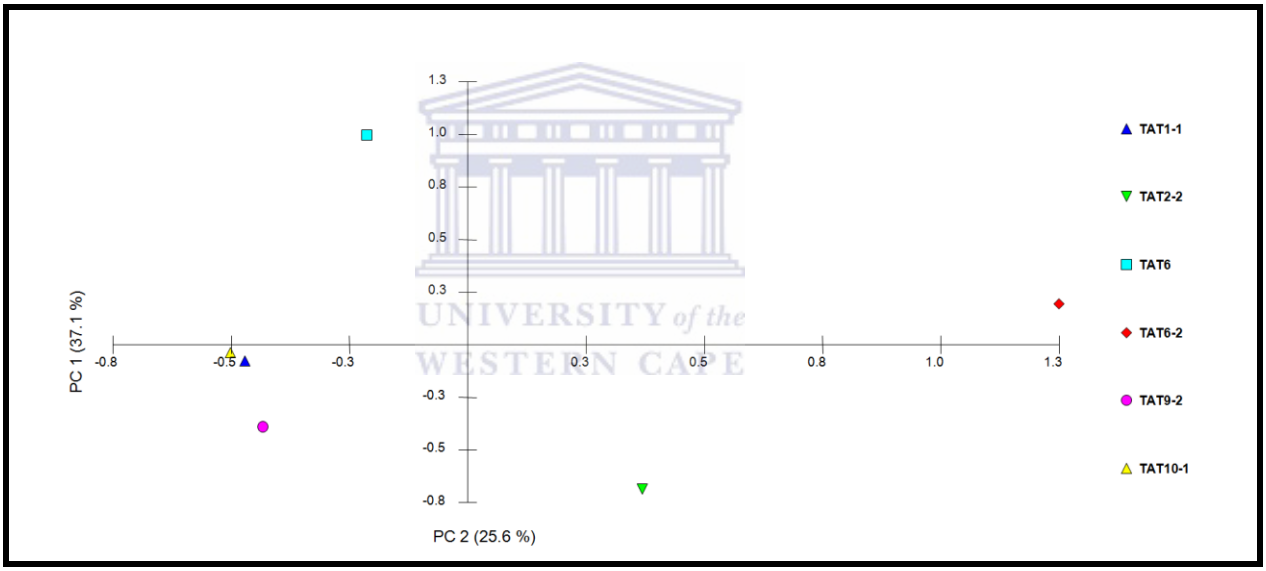

Figure 4.7: PCA plot based on the sequenced species from 16S rRNA clone libraries and PCR-DGGE profiles, revealing a 2D projection of axis 1 (x-axis) and 2 (y-axis) principal coordinates. The percentage values on the axes are cumulative percentage variation, which defines the maximum variability.

Using PCR-DGGE fingerprint data, a similar trend of variance was explained by PCA (62.5%), with PC 1 accounting for 43.1% and PC 2 19.4%, respectively (Figure 4.8). However, in this case the sample relatedness differed from that explained by PCA of sequenced species, and is

attributed to primer selectivity and sequencing sampling effort. Samples TAT2-2 and TAT10-1 were more closely related and formed a cluster with TAT1-1 and TAT9-1 along the PC 1 axis. This cluster was based on the bacterial bands shared among the samples. Samples TAT6 and TAT6-2 stood out individually and were negatively correlated (-0.1) (Appendix 2D). However, no statistically significant difference was observed from these two samples, probably because they had only one band difference, i.e., nine and eight bands for TAT6 and TAT6-2, respectively. The bacterial PCR-DGGE fingerprints showed significant differences ($P < 0.05$), whereas the archaeal and bacterial PCR-DGGE fingerprints combined showed no statistically significant differences ($P > 0.05$) (Appendix 2C).

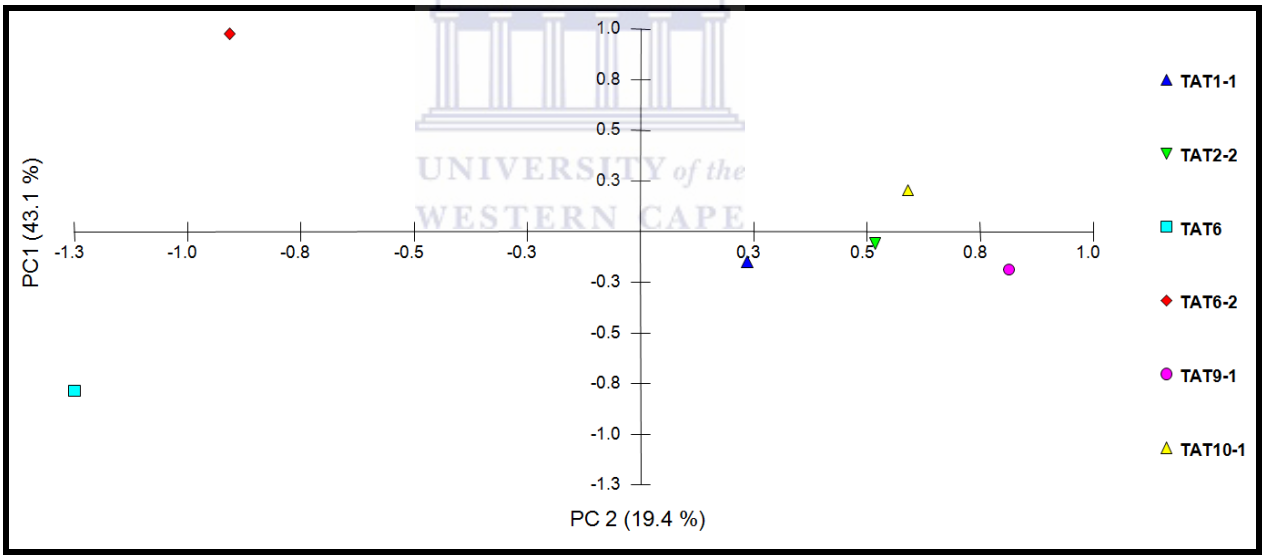

Figure 4.8: PCA plot based on archaeal and bacterial PCR-DGGE fingerprints on a 2D projection of axes 1 (x-axis) and 2 (y-axis) principal coordinates. The percentage values on the axes are cumulative percentage variation, which defines the maximum variability.

4.8.2 Canonical correspondence analysis of ETGF microbial diversity

The linkage of the microbial communities with environmental parameters was analysed by canonical correspondence analysis (CCA). CCA allowed distinction of the influence of the environmental factors such as temperature, salinity, pH, conductivity and total dissolved solids on the microbial diversity of the different ETGF samples.

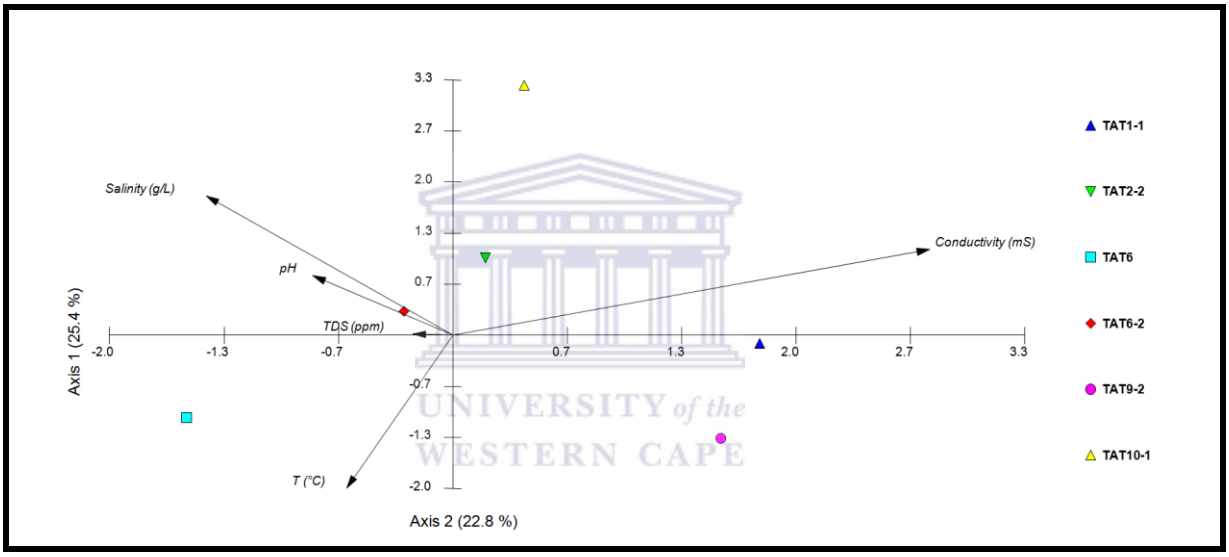

Figure 4.9: The CCA ordination diagram based on the sequenced species from 16S rRNA clone libraries and PCR-DGGE fingerprints. The CCA plots showed the distribution of species as determined by the environmental parameters. The arrows indicate the direction of maximum change of the environmental factors, where the length of the arrow is proportional to the rate of change.

Figure 4.9 indicates the influence of the physicochemical parameters on the distribution of sequenced microbial species from 16S rRNA clone libraries and PCR-DGGE fingerprints. CCA accounted for only 48.2% of cumulative variance, with axis 1 explaining 25.4% and axis 2 22.8%. Conductivity had the greatest influence on the sample variance, followed by salinity,

temperature, pH and, to a lesser extent, by TDS. CCA supported the distinction between the ETGF samples, which separated into four groups, with the first group consisting of TAT1-1 and TAT9-2, the second group comprising TAT6-2 and TAT2-2, and the third and fourth groups were represented by each of TAT101 and TAT6 samples. The microbial composition of sample TAT6-2 was more influenced by salinity and pH, and less so by TDS. This ordination corresponded to the sequencing results, which showed the haloarchaea dominance on TAT6-2 sample. Group 1 and group 4 samples were slightly influenced by conductivity and temperature, and this was supported by the presence of hyperthermophiles such as *Hydrogenobacter*, *Geoglobus* and nanoarchaea species.

When compared to CCA results for PCR-DGGE fingerprints, a similar pattern of environmental influence was observed, in which conductivity, followed by pH, salinity and temperature gradients had the strongest influence on the microbial composition of the samples (Figure 4.10). However, in this case, samples separated into three groups, with one cluster consisting of TAT1-1, TAT2-2, TAT9-1 and TAT10-1. This cluster was mostly influenced by salinity and conductivity. The other two groups were represented by TAT6 and TAT6-2 as individual samples. Microbial diversity of the TAT6-2 sample was influenced by pH, with only a small effect of temperature. No significant environmental impact was observed for the microbial composition of TAT6 sample (Figure 4.10). As a result, more physicochemical factors need to be investigated, such as those analysed by Fernandez-Turiel *et al*. (2005) and Landrum *et al*. (2009), in order to better understand the microbial distribution in the ETGF.

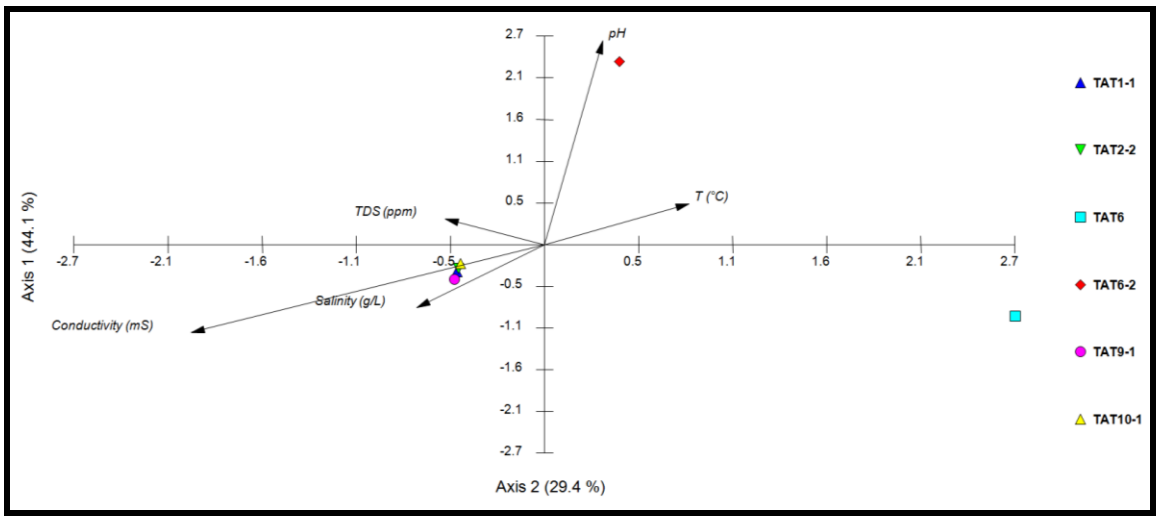

Figure 4.10: The CCA ordination diagram of bacterial and archaeal PCR-DGGE fingerprints based on the presence and absence of bands. The CCA plots showed the distribution of species as determined by the environmental parameters. The arrows indicate the direction of maximum change of the environmental factors, where the length of the arrow is proportional to the rate of change.

4.9 Discussion and conclusion

The ETGF samples revealed moderate microbial diversity, represented by a total of 30 individual species assigned to OTUs from archaeal, nanoarchaeal and bacterial taxonomic groups. The low microbial diversity of species associated with hot springs is thought to be due mainly to the effects of geochemistry. One example is the hot springs of St Lucia, Lesser Antilles, known to be sulphur- and boron-rich, where Aquificales dominated the bacterial diversity in the pools. Silica-depositing thermal springs, the Eclipse geyser and Spindle spring of YNP, were also reported to contain limited phylotypic diversity, and comprised only *Thermocrinis ruber*. It is not clear what factors contribute to such behaviour. However, it was

suggested that it is most likely to be influenced by geochemical differences, such as gas and fluid chemistry (Blank *et al.*, 2002).

The ETGF has been considered to be one of the analogues or model systems for understanding ancient Precambrian microbial system (Landrum *et al.*, 2009). This is because of the extreme conditions associated with this geyser field. These conditions include high concentrations of arsenic and silica, high levels of solar radiation (such as UV-A, UV-B and UV-C), high temperature, high altitude, a low latitude and a variety of pH gradients ranging from acidic to alkaline (Landrum *et al.*, 2009; Phoenix *et al.*, 2006). Such extreme conditions are also believed to be responsible for the limited microbial diversity found in this geyser field, where proliferation of microorganisms may be inhibited by one or more of these extreme conditions.

This study showed the variability of the different ETGF sites as expressed by the PCA plots. Microbial community relatedness to physicochemical parameters was demonstrated and salinity and temperature appeared to be the most important contributors to the microbial distribution across the geyser field. This was also confirmed by the sequencing results. However, CCA ordinations showed conductivity and pH to have also influenced the microbial diversity of the ETGF. TDS was shown to contribute least to microbial diversity.

In conclusion, the microbial diversity detected from the ETGF sites seems to be an under-representation of a much broader microbial spectrum, that could not be justified by the primer selection and the various techniques used in this study. A more extensive study of the ETGF may be required. This could involve analysis of more sampling sites with a variety of

physicochemical conditons, in order to ensure a robust comparison and analysis. Furthermore, use of high-throughput sequencing would be highly advantageous in processing an increased sampling effort and ensuring that the ETGF microbial community is screened exhaustively. A multidisciplinary approach would also be feasible in linking these parameters to the assemblage and distribution of microbial life forms thriving in these extreme environments.

Chapter 5

Culture-dependent screening of novel nanoarchaeal species

5.1 Introduction

Nanoarchaea are members of the fourth archaeal phylum, currently represented by only one culturable isolate, *Nanoarchaeum equitans* (Huber *et al*., 2002). However, several nanoarchaeal phylotypes from environmental samples have been detected from geothermal and non-thermal biotopes around the globe, including the El Tatio geyser field (Hohn *et al*., 2002; McCliment *et al*., 2006; Casanueva *et al*., 2008).

N. equitans can grow only in a co-culture with its host, *Ignicoccus hospitalis*, suggesting a host-parasite or host-symbiont relationship that is yet to be resolved. The *N. equitans* genome lacks most of its metabolic and biosynthetic genes, and depends mainly on its host for growth (Waters *et al*., 2003). *I. hospitalis* is an obligate chemolithotroph that grows exclusively by sulphur reduction (Paper *et al*., 2007). It is able to grow at temperatures between 73 and 98 °C, in 0.5-5% (w/v) NaCl and over a pH range of 4.5-7.0. No growth of *I. hospitalis* has been detected below 70 °C or above 100 °C, or at pH values below 4 or above 7.5. Further attempts to co-culture *N. equitans* and *I. hospitalis* have been reported to be difficult, with only onesuccessful sub-culture out of 50 attempts (Paper *et al*., 2007). The challenge therefore lies in developing novel culturing methods that support growth of novel nanoarchaeal-host cells. Successful culturing of these novel nanoarchaeal species would be advantageous for whole

genome sequencing and comparative genome analysis. It could also improve the resolution of nanoarchaeal phylogenetic placement and give insight into microbial evolution.

Genomic studies of culturable microbial isolates have provided insight into comparative evolutionary biology. However, this knowledge is limited because more than 99% of prokaryotes remain uncultured (Amann *et al*., 1995; Nelson, 2003; Ward & Fraser, 2005). As much as metagenomic studies continue to reveal a tremendous wealth of novel microorganisms, it is still imperative to pursue these "unculturable" microorganisms, and develop novel methods for isolating the yet "unculturable" and undetected prokaryotes. In responding to this challenge, different techniques were used in this study in search of novel nanoarchaeal species from enriched cultures. These include fluorescent *in situ* hybridization (FISH), flow cytometry and whole genome amplification (WGA). The combination of these techniques had the potential to improve the chances of isolating individual nanoarchaeal cells. FISH would contribute to the identification of nanoarchaea using fluorescently labelled probes. Flow cytometry would be useful in separating the individual cells from mixed enriched cultures, and WGA, through multiple-strand displacement amplification (MDA) (Figure 5.1), could amplify the whole genome of the separated cells, to be used for whole genome sequencing and large insert cloning, such as fosmid library construction (Paulson *et al*., 1999).

Flow cytometry facilitates the analysis, counting and sorting of large numbers of individual cells (Kruger *et al*., 2002). This technique enhances detection of light scatter and fluorescence (FACS) emission signals, which are digitized and displayed as frequency distributions, providing information for cell classification (Givan, 1992). In this study, forward (FSC) and

side (SSC) scatter were used to sort the cells. These are the size-measuring parameters based on scattered light used by flow cytometry.

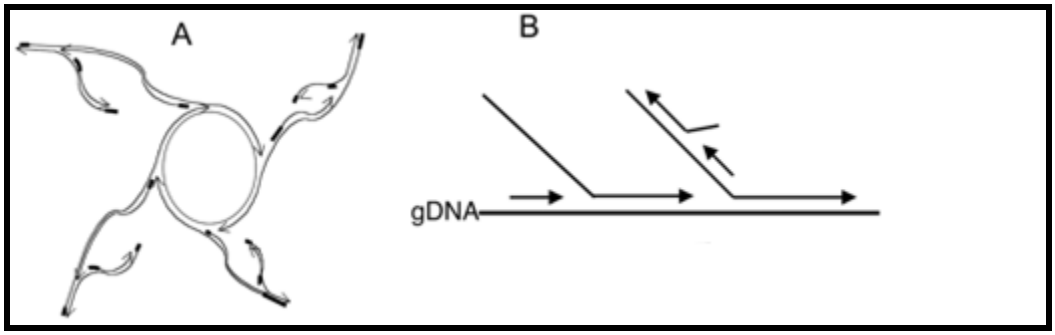

Figure 5.1: MDA strategy of random-primed rolling circle amplification. The arrowheads indicate 3′ ends and the thickened regions indicate primer locations; (A) indicates MDA on circular DNA template and (B) shows MDA on a linear genomic DNA template. Courtesy of Dean *et al*. (2002).

WGA is based on the ability of a high fidelity polymerase such as Φ29 DNA polymerase to cause strand displacement using random primers that anneal at random initiation points. Consequently, large amounts of high molecular weight hyperbranched DNA structures, referred to as wgaDNA are generated through these random events (Aviel-Ronen *et al*., 2006). WGA has been shown to amplify DNA isolated from as little as individual cell material (Raghunathan *et al*., 2005). This approach was adopted in an attempt to separate cells by flow cytometry and to amplify them using MDA.

5.2 Aims and objectives

The aims and objectives of the studies reported in this chapter were:

- To use an anaerobic manifold fermenter system for culturing nanoarchaea-host cells from El Tatio geyser field samples.

- To trace the presence of the nanoarchaeal cells by PCR and further verify the nanoarchaeal-host association by FISH.

- To extract genomic DNA of co-cultured cells for comparative genome analysis, in order to determine their novelty and phylogenetic relatedness to the known nanoarchaeal-*Ignicoccus* isolates.

5.3 Enrichment of hyperthermophilic samples

Samples were enriched anaerobically at 80 °C and 85 °C. Cell cultures were sampled over 24-hour intervals for DNA extraction, microscopic observations, DAPI staining and FISH analysis.

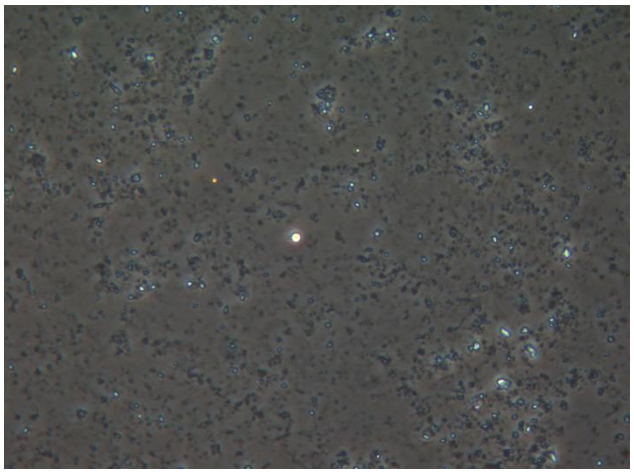

Figure 5.2: Micrograph of DAPI (4',6-diamidino-2-phenylindole) stained enrichment mixed cultures.

Microscopic observations on DAPI stained cultures revealed mixed microbial cells of different morphologies and sizes, including rods and cocci (Figure 5.2). However, no distinction could be made between live and dead non-motile cells, because DAPI is known to stain both viable and dead cells (Shimakita *et al*., 2007).

5.4 Electron microscopy

The mixed-cell cultures were screened by electron microscopy for the presence of nanoarchaea. Free-living and attached nano-sized ‘particles’ were detected in enrichment samples from sites TAT2-2, TAT9-1 and TAT10-1 (Figure 5.3).

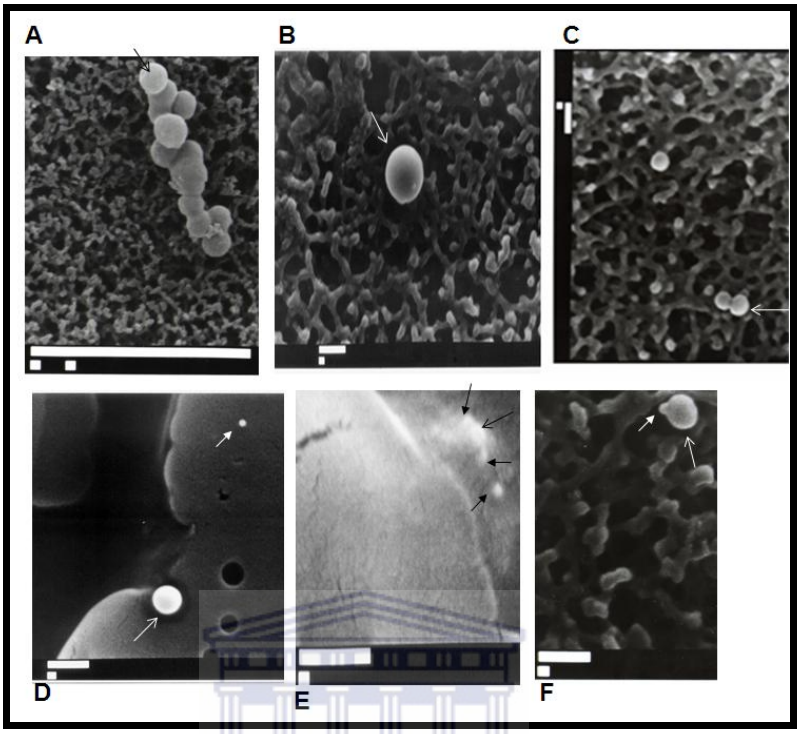

Figure 5.3: Electron micrographs generated from enrichment samples of TAT2-2, TAT9-1 and TAT10-1 sites. The nano-sized 'particles' are indicated by arrows. The scale bars represent 1 micrometre (where one square block is reflected under the scale bar) and 2 micrometres (where two square blocks are reflected under the scale bar).

Figure 5.3 shows a variety of coccoid 'particles' that may represent archaeal and nanoarchaeal cells. Using the magnification and the scale bar measurements, the coccoid particles were calculated to range in diameter from 0.1 to 1.2 µm. The diameters of the coccoid particles (as indicated by arrows) were calculated as follows: block A: 0.25 µm; B: 1.2 µm; C: 0.8 µm; D: 0.25 µm and 0.63 µm; E: 0.15 µm and 0.23 µm; and F: 0.2 and 0.5 µm; respectively. At this stage it is not clear whether these 'particles' are living cells or just artifacts. However, the small coccoid 'particles' are compatible with *N. equitans* dimensions, which fall within 0.4 µm diameter range and could contain less than 1% of the *E. coli* cell bio-volume (Huber *et al*.,

2003). FISH and sequencing results would give a more accurate indication of whether these coccoid 'particles' are indeed living cells or just artifacts.

Variations in cell size have been observed for archaeal and bacterial cell cultures (Stetter, 1996; Stetter & Zillig, 1985). Some examples include *Thermofilum* (0.15-0.17 μm), *Thermoproteus* (0.4 μm), *Desulfurococcus* (0.5-15 μm) (Stetter, 1999), *N. equitans* and *Dehalococcoides ethenogenes* (0.4-0.5 μm), etc. (Huber *et al*., 2002; Dworkin & Falkow, 2006). The minimal genome of microorganisms has also been associated with the smallest cell bio-volumes (Gregory, 2003). As a result, several estimates have been derived for the minimal genome required for a free-living cell. Approximately 250 genes were identified as being sufficient to sustain growth of a "modern-type" cell (Mushegian & Koonin, 1996). A comparison of a minimal genome size for self-replicating organisms, such as *Mycoplasma genitalium* and its closest relative *M. pneumoniae*, suggested that 265-350 genes were sufficient for cell survival and growth (Hutchison *et al*., 1999). The sequencing of *N. equitans* has further reduced the size of the previously known minimal genome [that of *M. genitalium* with 580 kb (~ 469 genes)]. Both of these genomes are thought to have lost most of their genes through an extensive genome reduction process (Waters *et al*., 2003; Glass *et al*., 2006). This suggests that there is a high possibility of identifying even smaller (in terms of cell and genome size) living entities than the ones that are already known. Furthermore, this support the finding of nanosized particles (whether smaller or greater than the known *N. equitans* cell size) as probable nanoarchaeal cells.

5.5 Fluorescence *in situ* hybridization (FISH)

The cultured cells from the El Tatio samples were subjected to FISH analysis and viewed on an Axioplan 2 imaging fluorescence microscope (Zeiss) with built-in emission filters for DAPI, rhodamine and CY3 (cyanine) (Figures 5.4-5.6). Fluorescence was detected from all three fluorochromes, and this could be attributed to the mixed archaeal cultures. However, no distinctive cell morphologies or nanoarchaeal cells were observed. Evidence from light and DAPI staining (Figures 5.2 and 5.5) indicated the presence of mixed cell morphologies, and suggested that the fluorescence was derived from true cells and not just artifacts. However, attempts to trace nanoarchaeal cells by PCR from enriched cultures were not successful, suggesting that nanoarchaeal cells might have been lost through an enrichment process where, culturing conditions may not have been favourable for nanoarchaeal growth. Alternatively, nanoarchaeal cells might have been lost during the centrifugation stages of DNA extraction experiments.

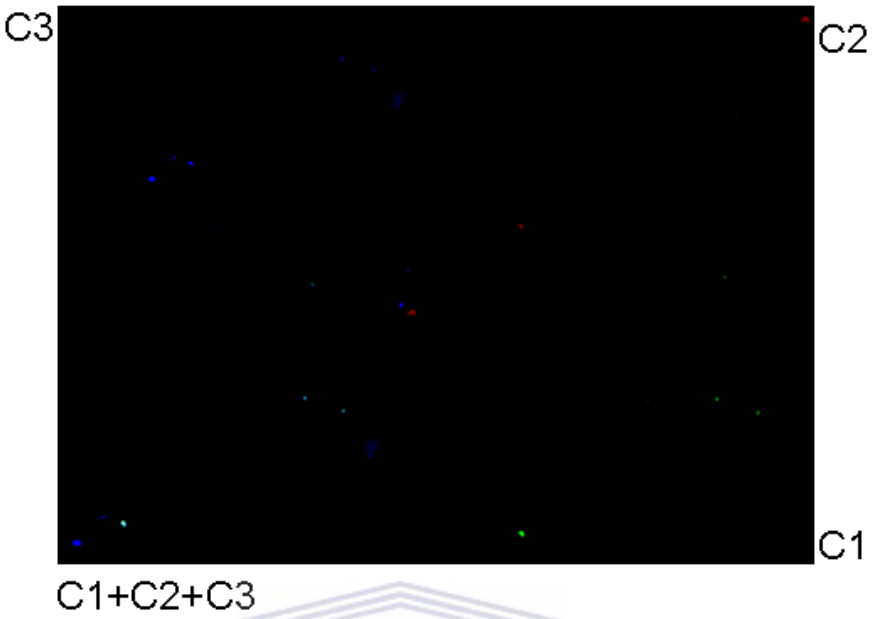

Figure 5.4: Amalgamated FISH micrographs of TAT6-2 cultures. C1: Cy3 (green), C2: rhodamine (red), C3: DAPI (blue), C1+C2+C3: combination of all three fluorochromes.

Figure 5.5: FISH micrographs of TAT6-2 cultures stained with DAPI.

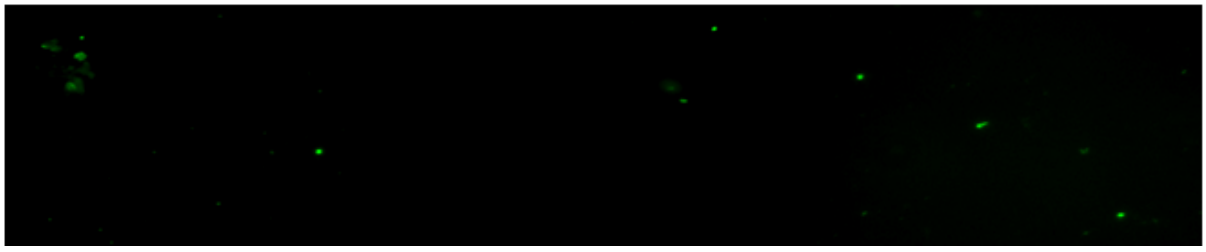

Figure 5.6: FISH micrographs of TAT6-2 cultures stained with Cy3.

It is evident from the results obtained in this study that more work still needs to be done to improve nanoarchaeal screening methods. This work should include optimal culturing conditions for nanoarchaeal enrichment, together with FISH using high resolution electron microscopy for refined imaging for nanosized or smaller cells.

5.6 Cell sorting by flow cytometry

To overcome the challenge of possible loss of nanoarchaeal cells through DNA extraction, flow cytometry was used to separate the cell enrichment cultures according to cell size. This approach was chosen as a result of the presence of nano-sized 'particles' as shown by electron microscopy. Using forward and side light scatter, large cells were scattered higher on the FSC axis (R15), compared to their smaller counterparts that scattered lower on the FSC axis (R1) (Figure 5.7).

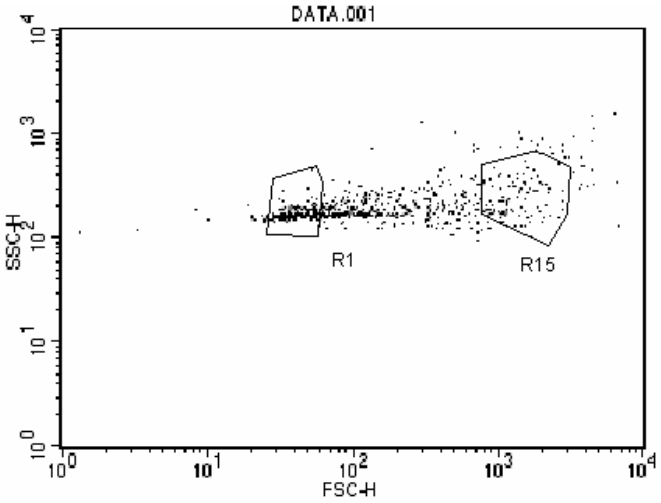

Figure 5.7: Illustration of flow cytometry sorted cells using TAT2-2 enrichment cultures. R1: region 1 gate for smaller cell and R15: region15 for bigger cells.

According to the scattergram (Figure 5.7), gate R1 was created to facilitate the separation of smaller cells from large cells located in gate R15. Cells recovered from the gate R1 and R15 fractions were processed for whole genome amplification (WGA) through multiple displacement amplification (MDA) (see Section 2.11). One hundred and ninety-two (192) wells from both R1 and R15 gates containing 1, 5, 10 and 20 cells/fraction were sorted for DNA extraction.

5.7 DNA extraction

Genomic DNA was successfully extracted from enriched cultures in amounts that ranged between 160 ng and 1.3 µg total gDNA. No quantifiable amounts of DNA were observed from the cell fractions separated by flow cytometry (results not shown). This could be the result of various factors such as the recovery of artifacts during cell sorting, the resistance of cells to lyse, low DNA extraction yields due to the presence of only limited cell numbers and loss of cells collected due to high volumes used for DNA extraction (Raghunathan *et al*., 2005).

5.8 Whole genome amplification of enriched cultures

5.8.1 WGA using the REPLI-G® kit

Whole genome amplification has been shown to be effective in amplifying gDNA isolated from individual cells (Raghunathan *et al*., 2005). Consequently, large amounts of high molecular weight DNA structures, referred to as wgaDNA, are generated through the random amplification events using random primers and high fidelity polymerase such as Φ29 DNA

polymerase (Aviel-Ronen *et al*., 2006). Due to minute amounts of gDNA were extracted from the enriched cell cultures, this approach was adopted to amplify any DNA extracted with minimal amplification bias. The wgaDNA could then be used for large insert cloning.

WGA was conducted on gDNA from enriched samples using the REPLI-G® kit (QIAGEN) according to the manufacturer's instructions. The gDNA was successfully amplified, resulting in high molecular weight DNA (Figure 5.8). TAT2-2, SHS-G, positive control and the negative control samples yielded 69, 78, 75 and 72 µg, respectively (Figure 5.8).

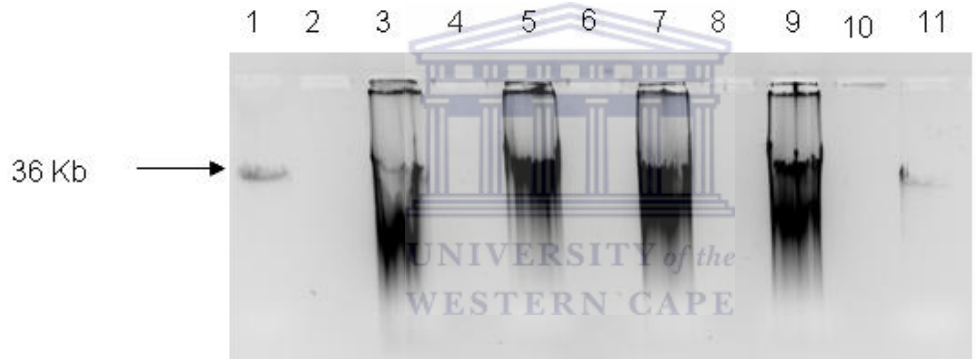

Figure 5.8: WGA using the REPLI-G® kit. Lanes 1 and 11: 100 ng fosmid control DNA; lane 3: TAT2-2; lane 5: SHS-G (control); lane 7: control DNA from the kit; lane 9: negative control; lanes 2, 4, 6, 8 and 10 were left empty to allow space for gel slicing.

Owing to the sensitivity of this method, even trace amounts of contaminating DNA can be amplified, and the formation of primer dimers can result in background synthesis. As shown in Figure 5.8, high molecular weight DNA was obtained. However, the negative control (lane 9) produced equally large yields of high molecular weight DNA. The generation of a high DNA background in the negative control has been observed in previous studies, and is believed to be caused by artifactual synthesis such as self-priming or primer dimer formation, DNA

contamination or both (Raghunathan *et al*., 2005; Hutchison & Venter, 2006; Zhang *et al*., 2006).

5.8.2 WGA using Φ29 DNA polymerase

To overcome the background formation caused by the REPLI-G® kit, Φ29 DNA polymerase was used with modified random heptamers as primers to avoid primer re-annealing, and was found to be effective in both the synthesis of high molecular weight DNA and in removing the background amplification (Figure 5.9). However, the amplification yield using modified random heptamers as primers was quantitatively less compared to REPLI-G amplification, and can be attributed to the absence of primer dimer formation, and the quality and quantity of the DNA template used.

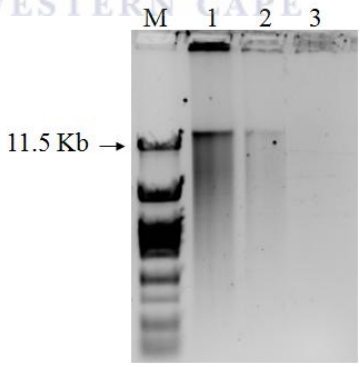

Figure 5.9: Amplification of gDNA from enriched culture samples using Φ29 DNA polymerase. Lane M: λ *Pst*I DNA marker; lane 1: TAT9-1; lane 2: TAT10-1; lane 3: negative control.

Even though no quantifiable amounts of DNA were observed from the eluate of cell fractions separated by flow cytometry, the eluate was also used as template for MDA, so that any trace

amounts of DNA present could be amplified. However, only enriched culture samples could amplify with Φ29 DNA polymerase, suggesting that high quality DNA template was most efficient for successful MDA. Previous reports have indicated that the Φ29 DNA polymerase can efficiently amplify from as little as 0.03 ng DNA template (Dean *et al*., 2002). On the basis of the results obtained in this study, however, this has not been the case. The template DNA is most like to have been the main limiting factor contributing to the failure of WGA, including other factors such as DNA polymerase inhibitors.

5.8.3 Purification of wgaDNA

A significant amount of wgaDNA was lost during purification using low-melting-point agarose. In the preparation of fosmid libraries, at least 2.5 μg of 40 kb DNA fragments are required for efficient library construction. In various amplification reactions, 1-6.9 μg was obtained. Gel electrophoresis showed that wgaDNA products migrated below the 40 kb fosmid control DNA benchmark; they were about 11.5 kb and not suitable for fosmid library construction (Figure 5.9). In order to avoid low cloning efficiency and the formation of chimeric clones, wgaDNA products could not be used for fosmid library construction.

Tools capable of generating data from single cells are limited because of the difficulties involved when dealing with small cell quantities and volumes to be analysed, including the disruption or lysis of cells (Brown & Audet, 2008). Cell lysis at a small scale or at a single-cell level can be a challenge especially when high cell dilutions are used, due to the possibility of the loss of cells during lysis. To circumvent this problem, special devices and protocols have

been developed for the control and handling of picolitre cell volumes (Irimia *et al*., 2004). Unfortunately, it was not possible to implement these methods in this study.

5.8.4 16S rRNA gene amplification from wgaDNA

In order to test whether the wgaDNA was efficient for other molecular applications such as PCR, the wgaDNA was amplified with nanoarchaeal targeting primers, A571Fb/N989R. No nanoarchaeal amplicons were detected from wgaDNA derived from enriched culture samples. However, amplification was obtained with a control wgaDNA of environmental sample (wgaDNA extracted from environmental samples) SHS-G from the Ethiopian Shalla hot springs (Figure 5.10), proving that wgaDNA can be used successfully for other PCR-based applications and to recover sequences of taxonomic groups of interest.

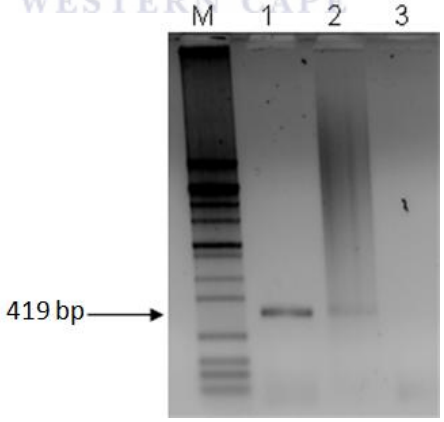

Figure 5.10: 16S rRNA gene amplification using wgaDNA as a template. Lane M: DNA marker; lane 1: SHS-G; lane 2: nanoarchaeal clone T173 as positive control; lane 3: negative control.

Detection of a nanoarchaeal PCR signal from environmental wgaDNA, but not from enriched samples, suggests that the nanoarchaeal cells might have been lost during cultivation and/or DNA extraction. Alternatively, nanoarchaeal cells might have been present in such low amounts that they could not be detected by amplification.

5.9 Impact of culturing limitations on nanoarchaea

From the results obtained in this study, it is clear that the effects of culturing bias had a negative impact on this study. It is generally known that hyperthermophiles are difficult to culture, due to their "extreme" environments, unusual and complex metabolic requirements. Most hyperthermophiles isolated thus far have generated relatively low biomass yields, which have consequently limited the detailed analysis of the biomolecules of these taxonomic groups (Marybeth *et al*., 2001). Inadequate or limited knowledge of factors affecting growth has also contributed to the lack of progress in developing effective culturing techniques (Kim & Lee, 2003). New culturing systems are being developed such as automated robotic systems that could be used for culturing both aerobic and anaerobic hyperthermophiles. Furthermore, these systems enable simultaneous screening and monitoring of various culturing conditions in one experimental run. For example, a new culturing system that makes use of a gas substrate under constant pressure has been reported (Uzarraga *et al*., 2011). This culturing system, made up of stainless steel jars, an automatic gas dispenser, propylene microplates and a robotic platform, offers an alternative culturing method to the conventional systems such as glass serum bottles, ceramic or stainless steel fermentors, and dialysis membrane reactors, (Hicks and Kelly, 1999).

Using these new culturing systems, nanoarchaea could be co-cultured, thereby resulting in new perspectives on the diversity of the Nanoarchaeota. This could be achieved by using the *N. equitans* genome as a benchmark to model the metabolic activities of nanoarchaeotes. In-depth analysis of environmental physicochemical data could be coupled with the metabolism model for engineering and developing these culturing methodologies. A clear understanding of nanoarchaeal metabolism is required, in order to achieve effective culturing that could generate sufficient biomass yields (Blumentals *et al*., 1990). Understanding the association of *I. hospitalis* and *N. equitans* could give insight into the metabolic requirements of nanoarchaeotes. At this stage *N. equitans* is known to be highly dependent on the presence of and contact with *I. hospitalis* (Burghardt *et al*., 2009), and the lack of metabolic capacity of *N. equitans* may contribute largely to the limitations on cell proliferation of nanoarchaeotes (Waters *et al*., 2003).

It is envisage that more time is needed for the thorough investigation of the *I. hospitalis* and *N. equitans* association before successful culturing of other novel nanoarchaeal representatives can be achieved. Meanwhile, high-throughput sequencing systems also offer alternative means to access genomic data and to gain insight into the physiology and metabolism of these microorganisms.

Chapter 6

General discussion

6.1 Introduction

In this study the microbial diversity of bacterial, archaeal and nanoarchaeal species from the ETGF was described using various molecular and statistical methods, such as 16S rRNA clone libraries, PCR-DGGE and multivariate statistical analysis tools. This has enabled comparison of the phylogenetic diversity and spatial distribution of species associated with seven sites within the geyser field. In addition, enriched cultures were screened for the presence of nanoarchaeal cells, where PCR and FISH were conducted to trace the presence of nanoarchaeal cells. MDA was also used to amplify the genomic DNA of enriched cell cultures, in order to generate sufficient DNA template to screen for nanoarchaeal genomes.

Chapter 1 reviews the traditional and modern technologies used in microbial ecology. Background information on the techniques and on the analytical tools used in this study is also explained. Chapter 2 describes in detail the materials and methods used in this work, while Chapters 3 to 5 discussed the results accumulated in this study. In this chapter, the broader context of the results generated is summarized.

6.2 Sampling

Samples were obtained from seven various sites within the El Tatio geyser field, Chile, courtesy of Dr Jenny Blamey. Unfortunately, only limited physicochemical data was obtained during sampling. Reference of a broader physicochemical data set, referred to in this study, has been analysed elsewhere (Fernandez-Turiela *et al.*, 2005; Landrum *et al.*, 2009; Tassi *et al.*, 2005).

6.3 DNA extraction

Metagenomic and genomic DNA extraction from environmental samples and enriched cultures has become a versatile and fruitful technique in microbial ecology. It has enabled the discovery and characterization of a number of microorganisms that were previously unknown, some of which are yet uncultured. In this study, high molecular weight DNA from environmental samples and enriched cultures was successfully extracted in sufficient purity and quality to allow amplification of the targeted DNA fragments. However, difficulties in obtaining sufficiently high DNA yields were experienced from both environmental and enriched samples, and consequently prevented large insert cloning, which requires large amounts of high-molecular-weight DNA. The low microbial diversity, DNA shearing effect and the silicified soil samples are believed have been the main causes of low DNA yields. For this reason, alternative methods to obtain high DNA yields were applied, such as MDA.

6.4 DNA amplification

The 16S rRNA gene was successfully amplified using various primer sets that targeted bacterial, archaeal and nanoarchaeal taxonomic groups. PCR amplification using nanoarchaeal targeting primer was initially difficult, yielding spurious amplification. PCR optimization reactions therefore had to be conducted, in which amplification conditions had to be modified, such that optimum amplification was obtained. Using the Taguchi PCR optimizing algorithm, successful amplification was obtained.

MDA was also conducted to amplify and generate high yields of the limited DNA extracted. However, this technique was discontinued due to lack of reproducibility that may have been caused by the quality of the DNA and the sensitivity of the method.

6.5 PCR-DGGE fingerprint analysis

PCR-DGGE fingerprint analysis was conducted on archaeal and bacterial DGGE profiles, constructed using archaeal (340F/533R) and bacterial (341F/534R) primer sets. The analysis demonstrated considerable differences in the community structure of the samples analysed. Bacterial PCR-DGGE fingerprints were shown to have the most significant differences ($P < 0.05$). In more than 100 unique DNA fragments that were identified through PCR-DGGE fingerprints and which theoretically may represent individual species, only a small fraction of the predominant community members could be excised and re-amplified for sequencing. The less dominant groups could not be sequenced due to their faint bands that could not be excised. Consequently, this has limited access to a broader microbial diversity of the ETGF and also contributed to the limited representation of the ETGF microbial diversity.

The sequenced represantatives from PCR-DGGE bands generated 16 unique species assigned to OTUs, which were mainly characterized by Desulfuroccales, Haloarchaea and a mixed pool of mesophilic bacteria. The cosmopolitan nature of this microbial diversity suggests an ability to adapt to a variety of temperature, pH, salinity and redox gradients.

6.6 Analysis of 16S rRNA gene clone libraries

The 16S rRNA clone libraries were constructed using archaeal (A571Fb/UA1204R; A3Fa/Ab927R) and nanoarchaeal (A571Fb/N989R) primer sets. The amplicons generated were successfully cloned for sequencing. Sequencing each individual phylotype present in the library would be laborious and expensive. Therefore, ARDRA analysis was conducted in order to monitor the heterogeneity of the phylotypes and to minimize the sequencing effort. A number of phylotypes that revealed unique ARDRA patterns were identified which were used to select representatives for sequencing. The sequencing results from the 16S rRNA gene clone libraries contributed 14 unique species assigned to OTUs.

6.7 Phylogenetic analysis

The microbial diversity of the ETGF samples analysed in this study yielded a total of 30 OTUs. The sequences from both the clone libraries and PCR-DGGE fingerprints complemented each other, and were dominated by hyperthermophiles and halophilic members of both archaea and bacterial domains. Four novel sequences were also identified from the TAT9-2 sample and could match only <100 bp of the database sequences from the 419 bp query sequence submitted. These sequences will be submitted to the GenBank database for future reference.

The presence of mesophilic microbial representatives was unexpected, even though it has been reported at most hot springs. This can be associated with a number of factors, such as varying temperature gradients within the geyser field, contamination, mispriming, and dispersal by natural forces such as rain and volcanic activities.

This study is the first of its kind to investigate a wide survey of microbial diversity in the ETGF using molecular techniques. Most studies of this geyser field were based on geological and geochemical systems for mining purposes and the search for geothermal resources for power generation (Fernandez-Turiela *et al*., 2005). This study was able to give an overview of the microbial community composition of the ETGF, and to provide a descriptive analysis of the differences and similarities among the different sites within the geyser field. Temperature and salinity were shown to be the most influential environmental conditions impacting the microbial community structure of the ETGF, which is believed to originate from a hot magmatic brine with steam-heated meteoric water (Landrum *et al*., 2009), and thus explains the dominance of haloarchaea and hyperthermophiles in this field. In addition, the ETGF has been reported to contain silica-rich saline water with high concentrations of Ca, Li, As, and Cs, and thus becoming a significant source of toxic elements (Landrum *et al*., 2009). Consequently, the moderate microbial diversity of the ETGF can be attributed to such toxicity, which is believed to have an adverse effect on the proliferation of microorganisms in the region.

6.8 Culture-dependent screening of novel nanoarchaeal species

The detection of nanoarchaeotes and *Ignicoccus* species from the ETGF has triggered interest in culturing the nanoarchaeotes. This was to search for novel nanoarchaeal species and their

association or interaction with other microorganisms and/or potential hosts. Using PCR amplification to trace the presence of nanoarchaeotes from the enriched samples gave no nanoarchaeal signals. Even repeated attempts to amplify the gDNA of enriched cultures through MDA were in vain. These results reflect previous studies where attempts to harvest *N. equitans-I. hospitalis* co-cultures have been reported to be difficult, with only one out of 50 attempts being successful (Paper, *et al*. 2007). Novel culturing methods that support the growth of novel nanoarchaeal-host cells are thus required. These would be advantageous for the identification of novel nanoarchaeal species and comparative genome analysis of these species with *N. equitans* and other archaeal taxonomic groups. Consequently, this could improve the resolution of nanoarchaeal phylogenetic placement and give insight into microbial evolution.

6.9 Recommendations and future work

It is recommended that future work that focuses on the evaluation of the microbial diversity of the ETGF should include more samples for comparative analysis. Temporal studies would be required, where changes in physicochemical gradients could be traced over certain time intervals and be used to assess the impact of factors that contribute to shaping the microbial community structure of the ETGF. This could be instrumental in tracking consistency of the identified microbial species. These studies, coupled with assessment of the metatranscriptome would also enable identification of the relationship between the indigenous microbial community versus the foreign taxonomic groups associated with adventitious events.

Furthermore, high-throughput sequencing could be used to generate robust sequencing data that will enable efficient estimation and representation of the overall microbial diversity and also minimize most of the bias introduced by PCR and cloning-based techniques.

Failure to culture nanoarchaeal isolates emphasizes the need for further work in this area. This would require a combination of multidisciplinary techniques that include investigating the metabolic activities, developing new culturing systems (e.g. robotic systems) and enrichment media that would mimic the natural environment, and favour the growth of the nanoarchaeotes. These systems would then enhance sufficient recovery of high molecular weight nanoarchaeal DNA and enable progressive whole genome sequencing, detection of novel genes and metabolic processes and would advance our understanding of nanoarchaeal diversity, its placement in the phylogenetic tree of life, and the significance of this phylogenetic group in the history of evolution.

Appendices

APPENDIX 1A

Principal component analysis of ETGF samples based on sequenced representatives from both 16S rRNA clone libraries and PCR-DGGE fingerprints

Analysis of 33 species x 6 samples

Tolerance of eigenanalysis set at 1E-007

Eigenvalues

	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
Eigenvalues	2.437	1.682	1.257	0.912	0.28
Percentage	**37.105**	**25.611**	19.138	13.888	4.258
Cum. percentage	37.105	**62.716**	81.854	95.742	100

PCA variable loadings

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Haloterrigena limicola	-0.039	0.267	0.11	0.173	-0.058
Uncultured Desulfurococcales	-0.039	0.267	0.11	0.173	-0.058
Uncultured *Vulcanisaeta* sp.	-0.039	0.267	0.11	0.173	-0.058
Ignicoccus hospitalis	-0.039	0.267	0.11	0.173	-0.058
Halobacteriaceae sp.	0.192	0.319	-0.027	0.079	-0.004
Natronococcus xinjiangense	0.231	0.052	-0.137	-0.093	0.053
Uncultured *Halobacterium* sp.	0.231	0.052	-0.137	-0.093	0.053
Haloferax volcanii	0.231	0.052	-0.137	-0.093	0.053
Haloarchaeon	0.26	0.135	0.232	0.198	-0.076
Uncultured *Pseudomonas* sp.	-0.025	-0.194	0.303	-0.096	0.529
Uncultured bacterium	0.132	-0.256	-0.154	0.042	-0.543
Halorubrum aidingense	0.192	0.319	-0.027	0.079	-0.004
***Haloferax* sp.**	-0.039	0.267	0.11	0.173	-0.058
Aquatic bacterium	0.068	-0.184	0.259	0.119	-0.072
***Delftia* sp.**	0.068	-0.184	0.259	0.119	-0.072
***Micromonospora* sp.**	-0.087	-0.02	-0.034	-0.237	-0.583
***Maricaulis* sp.**	0.231	0.052	-0.137	-0.093	0.053
***Brevibacillus* sp.**	0.231	0.052	-0.137	-0.093	0.053
Uncultured *Bacillus* sp.	0.231	0.052	-0.137	-0.093	0.053
Uncultured *Shewanella* sp.	0.299	-0.132	0.122	0.025	-0.019
Solibacillus silvestris	0.068	-0.184	0.259	0.119	-0.072
***Hydrogenobacter* sp.**	-0.08	-0.104	-0.241	0.254	0.058
Uncultured nanoarchaeote 173	0.299	-0.132	0.122	0.025	-0.019
Uncultured nanoarchaeote TA91N9A2	-0.08	-0.104	-0.241	0.254	0.058

Uncultured nanoarchaeote TOK 17C9	0.219	-0.236	-0.119	0.279	0.04
Uncultured archaeon MetC1	0.299	-0.132	0.122	0.025	-0.019
***Geoglobus* sp.**	-0.08	-0.104	-0.241	0.254	0.058
Thermothrix thiopara	-0.08	-0.104	-0.241	0.254	0.058
Uncultured archaeon	0.18	0.031	-0.009	0.452	-0.018
Uncultured euryarchaeote	0.068	-0.184	0.259	0.119	-0.072
Natrinema pallidum	0.231	0.052	-0.137	-0.093	0.053
Halogeometricum borinquense	0.231	0.052	-0.137	-0.093	0.053
Uncultured *Verrucomicrobia* sp.	-0.08	-0.104	-0.241	0.254	0.058

PCA case scores

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	-0.474	-0.076	-0.096	-0.483	-0.364
TAT2-2	0.372	-0.691	0.728	0.242	-0.045
TAT6	-0.215	1.003	0.308	0.352	-0.036
TAT6-2	1.258	0.195	-0.384	-0.191	0.033
TAT9-2	-0.436	-0.392	-0.678	0.517	0.036
TAT10-1	-0.505	-0.039	0.123	-0.439	0.376

APPENDIX 1B

Canonical correspondence analysis of ETGF samples based on sequenced representatives from both 16S rRNA clone libraries and PCR-DGGE fingerprints

Data file analyis of 33 variables x 6 cases
Environmental data file analysis of 5 variables
(physicochemical parameters) x 6 cases (samples)
Tolerance of eigenanalysis set at 1E-007

Scores scaled by species

Variable	**Weighted mean**	**Weighted SD**	**Inflation factor**
T (°C)	83.694	2.14	4.258
Conductivity (mS/cm)	31.465	8.062	2.414
TDS (ppm)	1192.539	160.933	9.773
pH	6.965	0.762	6.362
Salinity (g/L)	1.81	0.899	1.927

Eigenvalues

	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
Eigenvalues	0.728	0.654	0.594	0.544	0.349

Percentage	**25.363**	**22.808**	20.708	18.972	12.149
Cum. Percentage	25.363	**48.171**	68.879	87.851	100
Cum.Constr. percentage	25.363	48.171	68.879	87.851	100
Spec.-env. correlations	1	1	1	1	1
CCA variable scores					
	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Haloterrigena limicola	-1.542	-1.074	0.146	0.942	-0.064
Uncultured Desulfurococcales	-1.542	-1.074	0.146	0.942	-0.064
Uncultured *Vulcanisaeta* sp.	-1.542	-1.074	0.146	0.942	-0.064
Ignicoccus hospitalis	-1.542	-1.074	0.146	0.942	-0.064
Halobacteriaceae sp.	-0.912	-0.38	0.142	-0.102	0.273
Natronococcus xinjiangense	-0.282	0.313	0.139	-1.146	0.61
Uncultured *Halobacterium* sp.	-0.282	0.313	0.139	-1.146	0.61
Haloferax volcanii	-0.282	0.313	0.139	-1.146	0.61
Haloarchaeon	-0.545	0.08	-0.001	0.046	-0.271
Uncultured *Pseudomonas* sp.	0.302	2.124	-0.906	2.099	1.563
Uncultured bacterium	0.81	-0.032	0.823	0.095	-0.077
Halorubrum aidingense	-0.912	-0.38	0.142	-0.102	0.273
***Haloferax* sp.**	-1.542	-1.074	0.146	0.942	-0.064
Aquatic bacterium	0.189	1.002	-0.287	0.342	-1.358
***Delftia* sp.**	0.189	1.002	-0.287	0.342	-1.358
***Micromonospora* sp.**	1.778	-0.105	4.372	1.072	0.253
***Maricaulis* sp.**	-0.282	0.313	0.139	-1.146	0.61
***Brevibacillus* sp.**	-0.282	0.313	0.139	-1.146	0.61
Uncultured *Bacillus* sp.	-0.282	0.313	0.139	-1.146	0.61
Uncultured *Shewanella* sp.	-0.046	0.657	-0.074	-0.402	-0.374
Solibacillus silvestris	0.189	1.002	-0.287	0.342	-1.358
***Hydrogenobacter* sp.**	1.554	-1.34	-0.931	0.113	0.189
Uncultured nanoarchaeote 173	-0.046	0.657	-0.074	-0.402	-0.374
Uncultured nanoarchaeote TA91N9A2	1.554	-1.34	-0.931	0.113	0.189
Uncultured nanoarchaeote TOK 17C9	0.487	-0.008	-0.36	-0.231	-0.186
Uncultured archaeon MetC1	-0.046	0.657	-0.074	-0.402	-0.374
***Geoglobus* sp.**	1.554	-1.34	-0.931	0.113	0.189
Thermothrix thiopara	1.554	-1.34	-0.931	0.113	0.189
Uncultured archaeon	-0.02	-0.275	-0.233	0.063	-0.156
Uncultured euryarchaeote	0.189	1.002	-0.287	0.342	-1.358
Natrinema pallidum	-0.282	0.313	0.139	-1.146	0.61
Halogeometricum borinquense	-0.282	0.313	0.139	-1.146	0.61
Uncultured Verrucomicrobia sp.	1.554	-1.34	-0.931	0.113	0.189

CA case scores

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	1.778	-0.105	4.372	1.072	0.253
TAT2-2	0.189	1.002	-0.287	0.342	-1.358
TAT6	-1.542	-1.074	0.146	0.942	-0.064
TAT6-2	-0.282	0.313	0.139	-1.146	0.61
TAT9-2	1.554	-1.34	-0.931	0.113	0.189
TAT10-1	0.415	3.246	-1.524	3.856	4.483

Site scores, constrained by env. data

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	1.778	-0.105	4.372	1.072	0.253
TAT2-2	0.189	1.002	-0.287	0.342	-1.358
TAT6	-1.542	-1.074	0.146	0.942	-0.064
TAT6-2	-0.282	0.313	0.139	-1.146	0.61
TAT9-2	1.554	-1.34	-0.931	0.113	0.189
TAT10-1	0.415	3.246	-1.524	3.856	4.483

Canonical coefficients

	Spec. Axis 1	Spec. Axis 2	Spec. Axis 3	Spec. Axis 4	Spec. Axis 5
T (°C)	0.58	-1.309	1.41	-0.004	0.468
Conductivity (mS/cm)	1.291	0.044	0.8	-0.024	0.324
TDS (ppm)	-1.456	1.302	-1.761	1.334	1.038
pH	1.29	-0.102	1.081	-1.844	-0.345
Salinity (g/L)	-0.694	0.496	0.612	0.786	0.453

Interset correlations between env. variables and site scores

	Envi. Axis 1	Envi. Axis 2	Envi. Axis 3	Envi. Axis 4	Envi. Axis 5
T (°C)	-0.176	-0.57	-0.003	-0.237	0.767
Conductivity (mS/cm)	0.791	0.319	-0.026	0.521	-0.005
TDS (ppm)	-0.066	0.005	-0.265	-0.368	0.889
pH	-0.232	0.221	0.113	-0.756	0.56
Salinity (g/L)	-0.409	0.518	0.713	0.139	-0.192

Intraset correlations between env. variables and constrained site scores

	Envi. Axis 1	Envi. Axis 2	Envi. Axis 3	Envi. Axis 4	Envi. Axis 5
T (°C)	-0.176	-0.57	-0.003	-0.237	0.767
Conductivity (mS/cm)	0.791	0.319	-0.026	0.521	-0.005
TDS (ppm)	-0.066	0.005	-0.265	-0.368	0.889
pH	-0.232	0.221	0.113	-0.756	0.56
Salinity (g/L)	-0.409	0.518	0.713	0.139	-0.192

Biplot scores for env. variables

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
T (°C)	-0.176	-0.57	-0.003	-0.237	0.767
Conductivity (mS/cm)	0.791	0.319	-0.026	0.521	-0.005

TDS (ppm)	-0.066	0.005	-0.265	-0.368	0.889
pH	-0.232	0.221	0.113	-0.756	0.56
Salinity (g/L)	-0.409	0.518	0.713	0.139	-0.192

Centroids of env. variables					
	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
T (°C)	-0.004	-0.015	0	-0.006	0.02
Conductivity (mS/cm)	0.203	0.082	-0.007	0.134	-0.001
TDS (ppm)	-0.009	0.001	-0.036	-0.05	0.12
pH	-0.025	0.024	0.012	-0.083	0.061
Salinity (g/L)	-0.203	0.257	0.354	0.069	-0.095

APPENDIX 1C

ANOVA of ETGF samples based on sequenced representatives from both 16S rRNA clone libraries and PCR-DGGE fingerprints

SUMMARY

Group	**Count**	**Sum**	**Average**	**Variance**
TAT1-1	33	2	0.060606	0.058712
TAT2-2	33	12	0.363636	0.238636
TAT6	33	9	0.272727	0.204545
TAT6-2	33	17	0.515152	0.257576
TAT9-2	33	8	0.242424	0.189394
TAT10-1	33	1	0.030303	0.030303

ANOVA

Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	5.540404	5	1.108081	6.789942	7.46E-06	2.261138
Within Groups	31.33333	192	0.163194			
Total	36.87374	197				

SS denotes sum of squares; df: degree of freedom; MS: mean of squares; F: F-statistic and F crit: critical values of F-statistics

APPENDIX 1D

Correlation test of El Tatio samples based on sequenced representatives

	TAT1-1	TAT2-2	TAT6	TAT6-2	TAT9-2	TAT10-1
TAT1-1	1					
TAT2-2	0.072002	1				
TAT6	-0.15554	-0.18002	1			
TAT6-2	-0.0077	0.103129	-0.08664	1		
TAT9-2	0.152665	0.013363	-0.18764	-0.15864	1	
TAT10-1	-0.0449	0.233854	-0.10825	-0.18222	-0.1	1

APPENDIX 2A

Principal component analysis of bacterial and archaeal PCR-DGGE fingerprints based on presence or absence of bands

Analysis of 39 variables (DNA bands) x 6 cases (samples)

Tolerance of eigenanalysis set at 1E-007

Eigenvalues

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Eigenvalues	3.791	1.708	1.544	1.008	0.748
Percentage	**43.085**	**19.406**	17.551	11.46	8.498
Cum. percentage	43.085	**62.491**	80.042	91.502	100

PCA variable loadings

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Bac-A	0.097	-0.05	-0.279	0.056	-0.071
Bac-C	0.097	-0.05	-0.279	0.056	-0.071
Bac-D	0.167	0.005	-0.209	-0.304	0.005
Bac-E	0.257	-0.05	0.081	0.009	-0.049
Bac-F	0.121	0.247	-0.143	-0.025	0.316
Bac-G	0.159	-0.065	-0.157	0.277	0.256
Bac-H	0.167	0.005	-0.209	-0.304	0.005
Bac-I	0.097	-0.05	-0.279	0.056	-0.071
Bac-J	0.257	-0.05	0.081	0.009	-0.049
Bac-K	0.229	-0.011	-0.087	-0.084	0.332
Bac-L	0.187	-0.105	0.012	0.369	-0.125
Bac-M	0.187	-0.105	0.012	0.369	-0.125
Bac-N	0.195	-0.035	-0.041	-0.212	-0.376
Bac-O	0.097	-0.05	-0.279	0.056	-0.071
Bac-P	0.167	0.005	-0.209	-0.304	0.005
Bac-Q	0.159	-0.065	-0.157	0.277	0.256

Bac-R	0.087	0.223	-0.097	-0.152	-0.392
Bac-S	0.16	-0.001	0.36	-0.048	0.022
Bac-T	0.149	0.208	0.026	0.068	-0.065
Bac-U	0.16	-0.001	0.36	-0.048	0.022
Bac-V	0.024	0.297	0.136	-0.081	0.387
Bac-X	0.16	-0.001	0.36	-0.048	0.022
Bac-Y	0.07	0.054	0.07	-0.361	0.076
Bac-Z	0.229	-0.011	-0.087	-0.084	0.332
Arch-B	-0.149	-0.208	-0.026	-0.068	0.065
Arch-C	-0.257	0.05	-0.081	-0.009	0.049
Arch-E	-0.149	-0.208	-0.026	-0.068	0.065
Arch-F	-0.149	-0.208	-0.026	-0.068	0.065
Arch-G	-0.108	0.258	-0.056	0.059	-0.016
Arch-I	-0.108	0.258	-0.056	0.059	-0.016
Arch-J	-0.149	-0.208	-0.026	-0.068	0.065
Arch-K	-0.149	-0.208	-0.026	-0.068	0.065
Arch-L	-0.108	0.258	-0.056	0.059	-0.016
Arch-N	-0.108	0.258	-0.056	0.059	-0.016
Arch-O	-0.257	0.05	-0.081	-0.009	0.049
Arch-Q	-0.149	-0.208	-0.026	-0.068	0.065
Arch-R	-0.108	0.258	-0.056	0.059	-0.016
Arch-S	-0.108	0.258	-0.056	0.059	-0.016
Arch-V	-0.149	-0.208	-0.026	-0.068	0.065

PCA case scores

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	0.238	-0.15	0.581	0.209	-0.638
TAT2-2	0.523	-0.059	0.422	0.496	0.547
TAT6	-1.262	-0.794	-0.088	-0.153	0.108
TAT6-2	-0.916	0.986	-0.193	0.134	-0.026
TAT9-1	0.821	-0.19	-0.962	0.127	-0.118
TAT10-1	0.596	0.207	0.241	-0.813	0.127

APPENDIX 2B

Canonical correspondence analysis of bacterial and archaeal PCR-DGGE fingerprints based on presence or absence of bands

Data file analysis of 39 variables (DNA bands) x 6 cases (samples)

Environmental data file analysis of 5 variables (physicochemical parameters) x 6 cases (samples)

Tolerance of eigenanalysis set at 1E-007

Scores scaled by species

Variable	**Weighted mean**	**Weighted SD**	**Inflation Factor**
T (°C)	83.875	1.9	16.363

Conductivity (mS/cm)	40.96	12.067	22.175
TDS (ppm)	1300.318	258.872	63.228
pH	6.79	0.557	2.182
Salinity (g/L)	2.374	0.647	16.055

Eigenvalues

	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
Eigenvalues	0.907	0.603	0.27	0.167	0.107
Percentage	**44.158**	**29.381**	13.147	8.123	5.191
Cum. percentage	44.158	**73.539**	86.686	94.809	100
Cum. constr. percentage	44.158	73.539	86.686	94.809	100
Spec.-env. correlations	1	1	1	1	1

CCA variable scores

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Bac-A	-0.51	-0.439	1.615	-0.235	0.305
Bac-C	-0.51	-0.439	1.615	-0.235	0.305
Bac-D	-0.492	-0.341	0.543	0.804	0.056
Bac-E	-0.491	-0.332	-0.247	-0.094	0.11
Bac-F	-0.261	0.339	0.128	0.111	-0.371
Bac-G	-0.497	-0.369	0.459	-0.657	-0.675
Bac-H	-0.492	-0.341	0.543	0.804	0.056
Bac-I	-0.51	-0.439	1.615	-0.235	0.305
Bac-J	-0.491	-0.332	-0.247	-0.094	0.11
Bac-K	-0.489	-0.327	0.129	0.176	-0.514
Bac-L	-0.497	-0.361	-0.153	-0.739	0.211
Bac-M	-0.497	-0.361	-0.153	-0.739	0.211
Bac-N	-0.493	-0.342	-0.097	0.235	0.698
Bac-O	-0.51	-0.439	1.615	-0.235	0.305
Bac-P	-0.492	-0.341	0.543	0.804	0.056
Bac-Q	-0.497	-0.369	0.459	-0.657	-0.675
Bac-R	-0.264	0.327	-0.041	0.156	0.538
Bac-S	-0.485	-0.296	-0.868	-0.047	0.045
Bac-T	-0.308	0.202	-0.172	-0.092	0.1
Bac-U	-0.485	-0.296	-0.868	-0.047	0.045
Bac-V	-0.178	0.598	-0.367	0.227	-0.596
Bac-X	-0.485	-0.296	-0.868	-0.047	0.045
Bac-Y	-0.473	-0.243	-0.53	1.844	-0.193
Bac-Z	-0.489	-0.327	0.129	0.176	-0.514
Arch-B	2.657	-0.909	-0.023	0.013	-0.008
Arch-C	1.541	0.714	0.052	-0.035	0.025
Arch-E	2.657	-0.909	-0.023	0.013	-0.008
Arch-F	2.657	-0.909	-0.023	0.013	-0.008
Arch-G	0.425	2.337	0.126	-0.083	0.059
Arch-I	0.425	2.337	0.126	-0.083	0.059
Arch-J	2.657	-0.909	-0.023	0.013	-0.008
Arch-K	2.657	-0.909	-0.023	0.013	-0.008
Arch-L	0.425	2.337	0.126	-0.083	0.059
Arch-N	0.425	2.337	0.126	-0.083	0.059
Arch-O	1.541	0.714	0.052	-0.035	0.025
Arch-Q	2.657	-0.909	-0.023	0.013	-0.008
Arch-R	0.425	2.337	0.126	-0.083	0.059

Arch-S	0.425	2.337	0.126	-0.083	0.059
Arch-V	2.657	-0.909	-0.023	0.013	-0.008

CCA case scores

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	-0.496	-0.346	-1.375	-0.904	1.982
TAT2-2	-0.484	-0.299	-0.698	-1.08	-1.654
TAT6	2.657	-0.909	-0.023	0.013	-0.008
TAT6-2	0.425	2.337	0.126	-0.083	0.059
TAT9-1	-0.51	-0.439	1.615	-0.235	0.305
TAT10-1	-0.473	-0.243	-0.53	1.844	-0.193

Site scores, constrained by env. data

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
TAT1-1	-0.496	-0.346	-1.375	-0.904	1.982
TAT2-2	-0.484	-0.299	-0.698	-1.08	-1.654
TAT6	2.657	-0.909	-0.023	0.013	-0.008
TAT6-2	0.425	2.337	0.126	-0.083	0.059
TAT9-1	-0.51	-0.439	1.615	-0.235	0.305
TAT10-1	-0.473	-0.243	-0.53	1.844	-0.193

Canonical coefficients

	Spec. Axis 1	Spec. Axis 2	Spec. Axis 3	Spec. Axis 4	Spec. Axis 5
T (°C)	2.553	-1.135	-2.13	1.824	0.834
Conductivity (mS/cm)	1.607	-1.159	-3.56	2.356	0.171
TDS (ppm)	-4.206	1.765	5.767	-3.012	-0.303
pH	0.035	0.889	-1.077	0.441	0.192
Salinity (g/L)	-2.194	0.958	2.459	-1.909	0.794

Interset correlations between env. variables and site scores

	Envi. Axis 1	Envi. Axis 2	Envi. Axis 3	Envi. Axis 4	Envi. Axis 5
T (°C)	0.288	0.183	0.502	0.616	0.501
Conductivity (mS/cm)	-0.705	-0.396	-0.092	0.559	0.16
TDS (ppm)	-0.198	0.116	0.446	0.865	0.011
pH	0.117	0.919	0	0.365	0.093
Salinity (g/L)	-0.255	-0.285	-0.338	-0.526	0.681

Intraset correlations between env. variables and constrained site scores

	Envi. Axis 1	Envi. Axis 2	Envi. Axis 3	Envi. Axis 4	Envi. Axis 5
T (°C)	0.288	0.183	0.502	0.616	0.501
Conductivity (mS/cm)	-0.705	-0.396	-0.092	0.559	0.16
TDS (ppm)	-0.198	0.116	0.446	0.865	0.011
pH	0.117	0.919	0	0.365	0.093
Salinity (g/L)	-0.255	-0.285	-0.338	-0.526	0.681

Biplot scores for env. variables

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
T (°C)	0.288	0.183	0.502	0.616	0.501
Conductivity (mS/cm)	-0.705	-0.396	-0.092	0.559	0.16
TDS (ppm)	-0.198	0.116	0.446	0.865	0.011

pH	0.117	0.919	0	0.365	0.093
Salinity (g/L)	-0.255	-0.285	-0.338	-0.526	0.681

Centroids of env. variables

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
T (°C)	0.007	0.004	0.011	0.014	0.011
Conductivity (mS/cm)	-0.208	-0.117	-0.027	0.165	0.047
TDS (ppm)	-0.039	0.023	0.089	0.172	0.002
pH	0.01	0.075	0	0.03	0.008
Salinity (g/L)	-0.069	-0.078	-0.092	-0.143	0.186

APPENDIX 2C

ANOVA of El Tatio samples based on bacterial and archaeal PCR-DGGE fingerprints: Single factor

SUMMARY

Groups	**Count**	**Sum**	**Average**	**Variance**
TAT1-1	39	10	0.25641	0.195682
TAT2-2	39	14	0.358974	0.236167
TAT6	39	9	0.230769	0.182186
TAT6-2	39	12	0.307692	0.218623
TAT9-1	39	19	0.487179	0.25641
TAT10-1	39	16	0.410256	0.248313

ANOVA

Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	1.82906	5	0.365812	1.641171	0.150043	2.253643
Within Groups	50.82051	228	0.222897			
Total	52.64957	233				

SS denotes sum of squares; df: degree of freedom; MS: mean of squares; F: F-statistic and F crit: critical values of F-statistics

APPENDIX 2D

Correlation test of El Tatio samples based on bacteria and archaeal PCR-DGGE fingerprints

	TAT1-1	**TAT2-2**	**TAT6**	**TAT6-2**	**TAT9-1**	**TAT10-1**
TAT1-1	1					
TAT2-2	0.539878	1				
TAT6	-0.32163	-0.40988	1			
TAT6-2	-0.13702	-0.15145	-0.10143	1		
TAT9-1	0.250027	0.340013	-0.53385	-0.31634	1	
TAT10-1	0.465287	0.462541	-0.45683	-0.10426	0.334268	1

References

Abraham, W.-R., Strömpl, C., Meyer, H., *et al*. (1999). Phylogeny and polyphasic taxonomy of Caulobacter species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *International Journal of Systematic Bacteriology* **49**, 1053-1073.

Aditiawati, P., Yohandini, H., Madayanti, F., *et al*. (2009). Microbial diversity of acidic hot spring (Kawah Hujan B) in geothermal field of Kamojang area, West Java-Indonesia. *Open Microbiology Journal* **3**, 58-66.

Ahn, S., Kim, T. H., Lee, S., *et al*. (2009). The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Researc*h **19**, 1622-1629.

Alain, K., Querellou, J. (2009). Cultivating the uncultured: limits, advances and future challenges. *Extremophiles*, **13**, 583-594.

Allen, E. E. & Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology* **3**, 489-498.

Alm, E. W. & Stahl, D. A. (1996). Extraction of microbial DNA from aquatic sediments. In: Akkermans, A. D. L., van Elsas, J. D. & de Bruijn, F. J. (ed). *Molecular Microbial Ecology Manual* 1.1.5, Kluwer, Dordrecht, The Netherlands, 1-29.

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

Amann, R., Snaidr, J., Wagner, M., *et al*. (1996). *In situ* visualization of high genetic diversity in a natural microbial community. *Journal of Bacteriology* **178**, 3496-3500.

Amann, R. I., Stromley, J., Devereux, R., *et al*. (1992). Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. *Applied and Environmental Microbiology* **58**, 614-623.

Aminin, A. L. N., Madayanti, W. F., Pingkan, A., *et al*. (2008a). Simple enrichment and independent cultures to expand bacterial community analysis from Gedongsongo hot spring. *Journal of Bioscience and Bioengineering* **106**, 211-214.

Aminin, A. L. N., Warganegara, F. M., Aditiawati, P., *et al*. (2008b). Culture-independent and culture-dependent approaches on microbial community analysis at Gedongsongo (GS-2) hot spring. *International Journal of Integrative Biology* **2**,145-152.

Aravind, L., Tatusov, R. L., Wolf, Y. I., *et al*. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genetics* **14**, 442-444.

Auchtung, T.A., Takacs-Vesbach, C. D., Cavanaugh, C. M. (2006). 16S rRNA phylogenetic investigation of the candidate division 'Korarchaeota'. *Applied and Environmental Microbiology* **72**, 5077-5082.

Auguet, J.C., Borrego, C.M., Baneras, L., *et al*. (2008). Fingerprinting the genetic diversity of the biotin carboxylase gene (*acc*C) in aquatic ecosystems as a potential marker for studies of carbon dioxide assimilation in the dark. *Environmental Microbiology* **10**, 2527-2536.

Aviel-Ronen, S., Qi Zhu, C., Coe, B. P., *et al*. (2006). Large fragment *Bst* DNA polymerase for whole genome amplification of DNA from formalin-fixed paraffin-embedded tissues. *BMC Genomics* **7**, 312.

Baker, G. C., & Cowan, D. A. (2004). 16S rRNA primers and the unbiased assessment of thermophile diversity. *Biochemical Society Transactions* **32**, 218-221.

Baker, G. C., Gaffar, S., Cowan, D. A., *et al*. (2001). Microbial community analysis of Indonesian hot-springs. *FEMS Microbiology Letters* **200**, 103-109.

Baker, G. C., Smith, J. J. & Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods* **55**, 541-555.

Bapteste, E., Brochier, C. & Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353-363.

Barns, S. M., Fundyga, R. E., Jeffries, M. W., *et al*. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences USA* **91**, 1609-1613.

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

Baross, J. A. & Deming, J. W. (1983). Growth of "black smoker" bacteria at temperatures of at least 250 °C. *Nature* **303**, 423-426.

Belkin, S., Wirsen, C. O. & Jannasch, H. W. (1986). A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent. *Applied and Environmental Microbiology* **51**, 1180-1185.

Benaglio, P. & Rivolta, C. (2010). Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region. *PLoS ONE* **5**, e13071.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P. *et al*. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59.

Béjà, O., Suzuki, M. T., Koonin, E. V., *et al*. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology* **2**, 516-529.

Berger, W. H. & Parker, F. L., (1970). Diversity of planktonic foraminifera in deep-sea sediments. *Science* **168**, 1345-1347.

Bernhardt, G., Lüdemann, H. D., Jaenicke, R., *et al*. (1984). Biomolecules are unstable under "black smoker" conditions. *Naturwissenschaften* **71**, 583-586.

Bewick, V., Cheek, L. & Ball, J. (2004). Statistics review 9: One-way analysis of variance. *Critical Care* **8**, 130-136.

Bills, G. F. & Polishook, J. D. (1994). Abundance and diversity of microfungi in leaf-litter of a lowland rain-forest in Costa Rica. *Mycologia* **86**, 187-198.

Blackwood, C. B., Paul, E. A., (2003). Eubacterial community structure and population size within the soil light fraction, rhizosphere, and heavy fraction of several agricultural systems. *Soil Biology and Biochemistry* **35**, 1245-1255.

Blanco, L., Bernad, A., Lazaro, J. M., *et al*. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *Journal of Biological Chemistry* **264**, 8935-8940.

Blank, C. E., Cady, S. L & Pace, N. R. (2002). Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Applied and Environmental Microbiology* **68**, 5123-5135.

Blumentals, I. I., Brown, S. H., Schicho, R.N., *et al*, (1990). The hyperthermophilic archaebacterium, *Pyrococcus furiosus*. Development of culturing protocols, perspectives on scaleup, and potential applications. *Annals of the New York Academy of Sciences* **589**, 301-314.

Bohannan, B. J. M. & Hughes, J. (2003). New approaches to analyzing microbial diversity data. *Current Opinion in Microbiology* **6**, 282-287.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., *et al*. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology* **6**, 245-252.

Brochier, C., Gribaldo, S., Zivanovic, Y., *et al*. (2005). Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biology* **6**, R42.

Brunk, C. F., Avaniss-Aghajani, E. & Brunk, C. A. (1996). A computer analysis of primer and probe hybridization potential with bacterial small-subunit rRNA sequences. *Applied and Environmental Microbiology* **62**, 872-879.

Brown, R. B. & Audet, J. (2008). Current techniques for single-cell lysis. *Journal of the Royal Society Interface* **5**, S131-S138.

Buckley, D. H. & Schmidt, T. M. (2001). Environmental factors influencing the distribution of Verrucomicrobia in soil. *FEMS Microbial Ecology* **35**, 105-111.

Burghardt, T., Junglas, B., Siedler, F., *et al*. (2009). The interaction of *Nanoarchaeum equitans* with *Ignicoccus hospitalis*: proteins in the contact site between two cells. *Biochemical Society transactions* **37**, 127-132.

Caldwell, S. L., Liu, Y., Ferrera, I., *et al*. (2010). *Thermocrinis minervae* sp. nov., a hydrogen- and sulfur-oxidizing, thermophilic member of the Aquificales from a Costa Rican terrestrial hot spring. *International Journal of Systematic and Evolutionary Microbiology* **60**, 338-343.

Cao, Y., G. N., Cherr, A. L., Córdova-Kreylos, T. W.-M., *et al*., (2006). Relationships between sediment microbial communities and pollutants in two California salt marshes. *Microbial Ecology* **52**, 619-633.

Casanueva, A., Galada, N., Baker, G., *et al*. (2008). Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* **12**, 651-656.

Case, R. J., Boucher, Y., Dahllöf, I., *et al*. (2007). Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology* **73**, 278-288.

Chandler, D. P., Brockman, F. J., Bailey, T. J., *et al*. (1998). Phylogenetic diversity of archaea and bacteria in a deep subsurface paleosol. *Microbial Ecology* **36**, 37-50.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**,783-791.

Chao, A. (1984). Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 265-270.

Chao, A., Ma, M.-C. & Yang. M. C. K. (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrics* **43**, 783-791.

Claus, D. & Berkeley, R. C. W. (**1986)**. Genus *Bacillus* Cohn 1872. In: Sneath, P. H. A., Mair, N. S., Sharpe, M. E., *et al*. (eds). *Bergey's Manual of Systematic Bacteriology*, *Williams* & *Wilkins*, Baltimore **2**, 1105-1140.

Clifford, H. T., & Stephenson, W. (1975). An introduction to numerical classification. *Academic Press*, London.

Cobb, B. D. & Clarkson, J. M. (1994). A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. *Nucleic Acids Research* **22**, 3801-3805.

Cohan, F. M. (2002). Sexual isolation and speciation in bacteria. *Genetica* **116**, 359-370.

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

Cole, S. T. & Saint Girons, I. (1994). Bacterial genomics. *FEMS Microbiology Reviews* **14**, 139-160.

Colwell, R. K., & Coddington, J. A., (1994). Estimating terrestrial biodiversity through extrapolation: *Philosophical Transactions of the Royal Society of London B* **345**, 101-118.

Cowan, D. A. (2004). The upper temperature for life - where do we draw the line? *Trends in Microbiology* **12**, 58-60.

Córdova-Kreylos, A. L., Cao, Y., Green, P. G., *et al*. (2006). Diversity, composition, and geographical distribution of microbial communities in California salt marsh sediments. *Applied and Environmental Microbiology* **72**, 3357-3366.

Crist, T. O. & Veech, J. A. (2006). Additive partitioning of rarefaction curves and species-area relationships: Unifying α-, β- and γ- diversity with sample size and habitat area. *Ecology and Organismal Biology* **9**, 923-932.

Cui, H. -L., Tohty, D., Zhou, P. -J., Liu, S. -J. (2006). *Haloterrigena longa* sp. nov. and *Haloterrigena limicola* sp. nov., extremely halophilic archaea isolated from a salt lake. *International Journal of Systematic and Evolutionary Microbiology* **56**, 1837-1840.

Darwin, C. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life, 1st edn., *John Murray*, London.

Dean, F. B., Hosono, S., Fang, L., *et al*. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences USA* **99**, 5261-5266.

de la Torre, J. R., Walker, C. B., Ingalls, A. E., *et al*. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environmental Microbiology* **10**, 810-818.

DeLong, E. F., Schleper, C., Feldman, R. & R.V., S. (1999). Application of genomics for understanding the evolution of hyperthermophilic and nonthermophilic Crenarchaeota. *Biological Bulletin* **196**, 363-366.

DeLong, E. F., Preston, C. M., Mincer, T. *et al*. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496-503.

Desai, M. M. & Fisher, D. S. (2007). Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759-1798.

D'Hondt, S., Rutherford, S., & Spivack., A. J. (2002). Metabolic activity of the subsurface life in deep-sea sediments. *Science* **295**, 2067-2070.

D'Hondt, S, Jørgensen, B. B., Miller, D. J. *et al*. (2004). Distributions of microbial activities in deep subseafloor sediments, *Science* **306**, 2216-2221.

Di Giulio, M. (2006). *Nanoarchaeum equitans* is a living fossil. *Journal of Theoretical Biology* **242**, 257-260.

Di Giulio, M. (2008). The split genes of *Nanoarchaeum equitans* are an ancestral character. *Gene* **421**, 20-26.

Dieffenbach, C. W. & Dveksler, G. S. (1995). In: PCR primer. *A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor*, NY.

Dolan, P. and Denver,D. (2008). TileQC: a system for tile-based quality control of Solexa data. *BMC Bioinformatics*, **9**, 250**.**

Don, R. H., Cox, P. T., Wainwright, B. J., *et al*. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**, 4008.

Dunbar, J., Takala, S., Barns, S.M., *et al*. (1999). Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Applied and Environmental Microbiology* **65**, 1662-1669.

Dupré, J. & O'Malley, M. A. (2007). Metagenomics and biological ontology. *Studies in History and Philososphy of Biological and Biomedial Sciences* **38**, 834-846.

Dworkin, M. & Falkow, S. (2006). Prokaryotes. In: Dworkin, M., Falkow, S., Rosenberg, E., *et al*. (eds). A handbook on the biology of Bacteria. Symbiotic associations. *Springer*.

Eck R.V. and Dayhoff, M.O. (1966). Atlas of protein sequence and structure. National Biomedical Research Foundation, *Silver Springs*, Maryland.

Eder, W. & Huber, R. (2002). New isolates and physiological properties of the Aquificales and description of *Thermocrinis albus* sp. nov. *Extremophiles* **6**, 309-318.

Edwards R. A., Rodriguez-Brito, B., Wegley, L., *et al*. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**, 57.

Elkins, J. G., Podar, M., Graham, D. E., *et al*. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences USA* **10**5, 8102-8107.

Ellis, D. G., Bizzoco, R. W. & Kelley, S. T. (2008). Halophilic archaea determined from geothermal steam vent aerosols. *Environmental Microbiology* **10**, 1582-1590.

Elshahed, M. S., Najar, F. Z., Roe, B. A., *et al*. (2004). Survey of archaeal diversity reveals an abundance of halophilic Archaea in a low-salt, sulfide- and sulfur rich spring. *Applied and Environmental Microbiology* **70**, 2230-2239.

Engel, A.S., **Bennett, P. C., & Stern, L.A. (2001).** Biogeochemical diversity of microbial mats from Lower Kane Cave, Wyoming, USA. In: Weber, J., Jamroz, E., Drozd, J., *et al*. (eds).

Biogeochemical processes and cycling of elements in the environment. *15th International Symposium of Environmental Biogeochemistry*, Wroclaw, Poland, 135-136.

Fardeau, M. L., Barsotti, V., Cayol, J. L., *et al*. (2010). *Caldinitratiruptor microaerophilus*, gen. nov., sp. nov. isolated from a French hot spring (Chaudes-Aigues, Massif Central): a novel cultivated facultative microaerophilic anaerobic thermophile pertaining to the *Symbiobacterium* branch within the Firmicutes. *Extremophiles* **14**, 241-247.

Farrelly, V., Rainey, F. A. & Stackebrandt, E. (1995). Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Applied and Environmental Microbiolology* **61**, 2798-2801.

Fernandez-Turiel, J. L., Garcia-Valles, M., D. Gimeno-Torrente, D., *et al*. (2005). The hot spring and geyser sinters of El Tatio, Northern Chile. *Sedimentary Geology* **180**, 125-147.

Ferrera, I., S. Longhorn, A. B. Banta, Y., *et al*. (2007). Diversity of 16S rRNA gene, ITS region and *aclB* gene of the Aquificales. *Extremophiles* **11**, 57-64.

Forterre, P., Bouthier de la Tour, C., Philippe, H., *et al*. (2000). Reverse gyrase from hyperthermophiles, probable transfer of a thermoadaptation trait from Archaea to Bacteria. *Trends in Genetics* **16**, 152-154.

Forterre, P. (2001). Genomics and early cellular evolution. The origin of the DNA world. *Comptes Rendus de l'Académie des Sciences III **324**,* 1067-1076.

Ferris, M. J., Muyzer, G. & Ward, D. M. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Applied and Environmental Microbiology* **62**, 340-346.

Fox, G. E., Magrum, L. J., Balch, W. E., *et al*. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences* USA **74**, 4537-4541.

Frangeul, L., Nelson, K. E., Buchrieser, C., *et al*. (1999). Cloning and assembly strategies in microbial genome projects. *Microbiology* **145**, 2625-2634.

Fraser, C. M., Gocayne, J. D. White, O., *et al*. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403.

Fromin, N., J., Hamelin, S., Tarnawski, D., *et al*. (2002). Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environmental Microbiology* **4**, 634-643.

Fuchs, T., Huber, H., Teiner, K., *et al*. (1995). *Metallosphaera prunae*, sp. nov., a novel metal-mobilizing thermoacidophilic archaeum, isolated from a uranium mine in Germany. *Systematic and Applied Microbiology* **18**, 560-566.

Galada, N. (2005). Exploring diversity and ecology of nanoarchaea in hydrothermal biotopes. MSc thesis. *University of the Western Cape*, SA.

Gao, B. & Gupta, R. S. (2007). Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* **8**, 86.

Garrity, G. M. & Holt, J. G. (2001). Phylum BVI. Chloroflexi phy. nov. In: Boone, D. R. & Castenholz, R. W. (eds). The Archaea and the deeply branching and phototrophic bacteria. Bergey's Manual of Systematic Bacteriology, 2nd edn., *Springer-Verlag*, NY, 427-446.

Garrity, G. M., Bell, J. A. & Lilburn, T. (2005). The revised road map to the manual. In Brenner, D. J., Krieg, N. R. & Staley, J. T. (eds). Bergey's Manual of Systematic Bacteriology, 2nd edn., *Springer*, NY, 159-187.

Gerbilskii, N. L. & Petrunkevitch, A. (1955). Intraspecific biological groups of acipenserines and their reproduction in the low regions of rivers with biological flow. *Systematic Zoology* **4**, 86-92.

Ghosh, D., Bal, B., Kashyap, V. K., *et al*. (2003). Molecular phylogenetic exploration of bacterial diversity in a Bakreshwar (India) hot spring and culture of *Shewanella*-related thermophiles. *Applied and Environmental Microbiology* **69**, 4332-4336.

Gich, F. B., Amer, E., Figueras, J. B., *et al*. (2000). Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis (ARDRA). *International Microbiology* **3**, 103-106.

Gilbert, W., & Maxam, A. (1973). The nucleotide sequence of the *lac* operator. *Proceedings of the National Academy of Sciences USA* **70**, 3581-3584.

Gini, C. (1912). Variabilit`a e Mutabilit`a, Bologna: Tipografia di Paolo Cuppini.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L, *et al*. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60-63.

Givan, A. (1992). Flow cytometry: first principles. *Wiley*, NY, 15-73.

Glass, J. I., Assad-Garcia, N., Alperovich, N, *et al*. (2006). Essential genes of a minimal bacterium. *Proceedings of National Academy of Sciences USA* **103**, 425-430.

Glansdorff, N., Xu, Y. & Labedan, B. (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biology Direct* **3**, 29.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379-391.

Gregory, T. R. (2003). Is small indel bias a determinant of genome size? *Trends in Genetics* **19**, 485-488.

Gribaldo, S., & Brochier-Armanet, C. (2006). The origin and evolution of Archaea: A state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1007-1022.

Grzymski, J. J., Carter, B. J., DeLong, E. F., *et al*. (2006). Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. *Applied and Environmental Microbiology* **72**, 1532-1541.

Hallam, S. J., Putnam, N., Preston, C. M., *et al*. (2004). Reverse methanogenesis: Testing the hypothesis with environmental genomics. *Science* **305**, 1457-1462.

Hamamura, N., Olson, S. H., Ward, D. M., *et al*. (2005). Diversity and functional analysis of bacterial communities associated with natural hydrocarbon seeps in acidic soils at Rainbow Springs, Yellowstone National Park. *Applied and Environmental Microbiology* **71**, 5943-5950.

Hamilton, G. (2006). The gene weavers. *Nature Reviews Microbiology* **441**, 683-685.

Handelsman, J., Rondon, M. R., Brady, S. F., *et al*. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology* **5**, R245-R249.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**, 669-685.

Hansen, M. C., Tolker-Nielsen, T., Givskov, M., *et al*. (1998). Biased 16S rRNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiology Ecology* **26**, 141-149.

Hatzenpichler, R., Lebecleva, E. V., Spieck, E., *et al*. (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proceedings of the National Academy of Sciences USA* **105**, 2134-2139.

Haubold, B. & Rainey, P. B. (1996). Genetic and ecotypic structure of a fluorescent Pseudomonas population. *Molecular Ecology* **5**, 747-761.

Heidelberg, K. B., Gilbert, J. A., Joint, I. (2010). Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microbial Biotechnology* **3**, 531-543.

Hershberger, K. L., Barns, S. M., Reysenbach, A. L., *et al*. (1996). Wide diversity of Crenarchaeota. *Nature* **384**, 420.

Heuer, H. & Smalla, K. (1997). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) for studying soil microbial communities. In: van Elsas, J. D., Wellington, E. M. H. & Trevors, J. T. (eds), *Modern Soil Microbiology, Marcel Dekker*, NY, 353-373.

Hewson, I., & Fuhrman, J. A. (2004). Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Applied and Environmental Microbiology* **70**, 3425-3433.

Heyndrickx, M., Vauterin, L., Vandamme, P., *et al*. (1996). Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. *Journal of Microbiological Methods* **26**, 247-259.

Heywood, V. H. (1995). Global biodiversity assessment. *Cambridge University Press*, Cambridge.

Hicks, M. & Kelly, R. M. (1999). Thermophilic microorganisms. In: Flickinger, M. C. & Drew, S. W. (eds). Encyclopedia of bioprocess technology: fermentation, biocatalysis, and bioseparation. *Wiley Interscience*, NY, 2536-2552.

Hohn, M. J., Hedlund, B. P. & Huber, H. (2002). Detection of 16S rRNA sequences representing the novel phylum "Nanoarchaeota": Indication for a wide distribution in high temperature biotopes. *Systematic and Applied Microbiology* **25**, 551-554.

Holden, C. (2005). Life in the air. *Science* **307**, 1558.

Huber, G. & Stetter, K. O. (1991). *Sulfolobus metallicus*, sp. nov., a novel strictly chemolithoautotrophic thermophilic archaeal species of metal mobilizers. *Systematic and Applied Microbiology* **14**, 372-378.

Huber H, Burggraf S, Mayer T, *et al*. (2000). *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species, *Ignicoccus islandicus* sp nov and *Ignicoccus pacificus* sp nov. and *Ignicoccus pacificus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **50**, 2093-2100.

Huber, H., Hohn, M. J., Rachel, R., *et al*. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63-67.

Huber, H., Hohn, M. J., Stetter, K. O., *et al*. (2003). The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Research in Microbiology* **154**, 165-171.

Huber, H., Huber, R., Stetter, K. O. (2006). Thermoproteales. In: Dworkin, M., Falkow, S., Rosenberg, E., *et al*. (eds). *The prokaryotes*, *Springer*, NY, 10-22.

Huber, H. & Stetter, K. O. (1998). Hyperthermophiles and their possible potential in biotechnology. *Journal of Biotechnology* **64**, 39-52.

Huber, R., Eder, W., Heldwein, S., *et al*. (1998). *Thermocrinis ruber* gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone National Park. *Applied and Environmental Microbiology* **64**, 3576-3583.

Huber, R., Langworthy, T. A., Konig, H., T *et al*. (1986). *Thermotoga maritima* sp. nov.represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C. *Archives of Microbiology* **144**, 324-333.

Huber, R. & Stetter, K. O. (1992). The order Thermotogales. *In:* Balows, A., Truper, H. G., Dworkin, M., *et al*. (eds). The prokaryotes, 2nd edn., *Springer*, Berlin Heidelberg, NY, 3809-3815.

Huber, R., Wilharm, T., Huber, D., *et al*. (1992). *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Systematic and Applied Microbiology* **15**, 340-351.

Hugenholtz, P., Pitulle, C., Hershberger, C., *et al*. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriololgy* **180**, 366-376.

Hughes, J., Hellman, J., Ricketts, T., *et al*. (2001). Counting the uncountable: Statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology* **67**, 4399-4406.

Hughes, J. B., & Hellmann, J. J. (2005). The application of rarefaction techniques to molecular inventories of microbial diversity. *Methods in Enzymology* **297**, 292-308.

Huse, S., Huber, J., Morrison, H., Sogin, M., Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.

Huss, M. (2010). Introduction into the analysis of high-throughput-sequencing based epigenome data. *Briefing in Bioinformatics* **11**, 512-523.

Hutchison III, C. A., Peterson, S. N., Gill, S. R., *et al*. (1999). Global transposon mutagenesis and the minimal Mycoplasma genome. *Science* **286**, 2165-2169.

Hutchison III, C. A. & Venter, J. C. (2006). Single-cell genomics. *Nature Biotechnology* **24**, 657-658.

Inagaki, F., Takai, K., Komatsu, T., *et al*. (2001). Archaeology of Archaea: geomicrobiological record of Pleistocene thermal events concealed in a deep-sea subseafloor environment. *Extremophiles* **5**, 385-392.

Inagaki, F., Suzuki, M., Takai, K., *et al*. (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. *Applied Environmental Microbiology* **69**, 7224-7235.

Inagaki, F., Nunoura, T., Nakagawa, S., *et al*. (2006). Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proceedings of the National Academy of Sciences USA* **103**, 2815-2820.

Irimia, D., Tompkins, R. G. & Toner, M. (2004). Single-cell chemical lysis in picoliter-scale closed volumes using a microfabricated device. *Analytical Chemistry* **76**, 6137-6143.

Ishii, K. & Fukui, M. (2001). Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Applied and Environmental Microbiology* **67**, 3753-3755.

Itoh, T., Suzuki, T. -I. & Nakase, T. (2002). *Vulcanisaeta distributa* gen. nov., sp. nov., and *Vulcanisaeta souniana* sp. nov., novel hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1097-1104.

Jahn, U., Gallenberger, M., Paper, W., *et al*. (2008). *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: New insights into a unique, intimate association of two archaea. *Journal of bacteriology* **190**, 1743-1750.

Jannasch, H. W., Huber, R., Belkin, S., *et al*. (1988). *Thermotoga neapolitana* sp. nov. of the extremely thermophilic eubacterial genus *Thermotoga. Archives of Microbiology* **150**,103-104.

Jones, B. & Renaut, R. W. (1997). Formation of silica oncoids around geysers and hot springs at El Tatio, Northern Chile. *Sedimentology* **44**, 287-304.

Jørgensen, B. B. (1982). Mineralization of organic matter in the sea bed-the role of sulphate reduction. *Nature* **296**, 643-645.

Junier, P., Molina, V., Dorador, C. *et al*. (2010). Phylogenetic and functional marker genes to study ammonia-oxidizing microorganisms (AOM) in the environment. *Applied Microbiology and Biotechnology* **85**, 425-440.

Jurgens, G., Lindstrom, K. & Saano, A. (1997). Novel group within the kingdom Crenarchaeota from boreal forest soil. *Applied and Environmental Microbiology* **63**, 803-805.

Kanokratana, P., Chanapan, S., Pootanakit, K., *et al*. (2004). Diversity and abundance of Bacteria and Archaea in the Bor Khlueng Hot Spring in Thailand. *Journal of Basic Microbiology* **44**, 430-444.

Karner, M. B., DeLong, E. F. & Karl, D. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507-510.

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

Kashefi, K. (2004). Response to Cowan: The upper temperature for life - where do we draw the line? *Trends in Microbiology* **12**, 60-62.

Kassen, R. & Rainey, P. B. (2004). The ecology and genetics of microbial diversity. *Annual review of microbiology* **58**, 207-231.

Kemp, P. F., & Aller, J. Y. (2003). Bacterial diversity in aquatic and other environments: What 16S rRNA libraries can tell us. *FEMS Microbiological Ecology* **47**, 161-177.

Kemp, P. F. & Aller, J. Y. (2004). Estimating prokaryotic diversity: When are 16S rRNA libraries large enough? *Limnology and Oceanography*: *Methods* **2**, 114-125.

Keohavong, P. & Thilly, W.G., (1989). Fidelity of DNA polymerases in DNA amplification. *Proceedings of National Academy of Sciences USA*, 9253-9257.

Kim, K. W. & Lee, S. B. (2003). Growth of the hyperthermophilic marine archaeon *Aeropyrum pernix* in a defined medium. *Journal of Bioscience and Bioengineering* **95**, 618-622.

Klaus, J. S., Janse, I., Heikoop, J. M., *et al*. (2007). Coral microbial communities, zooxanthellae and mucus along gradients of seawater depth and coastal pollution. *Environmental Microbiology* **9**, 1291-1305.

Könneke, M., Bernhard, A. E., de la Torre, J. R., *et al*. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543-546.

Kovach, W. L. (2004). Multivariate Statistical Package. Version 3.1 for Windows. Pentraeth, Wales, UK.

Kowalchuk, G. A., Stephen, J. R., De Boer, W., *et al*. (1997). Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class Proteobacteria in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Applied and Environmental Microbiology* **63**, 1489-1497.

Krebs, C. J., (1978). Species diversity. In: Krebs, C. J. (ed). Ecology: The experimental analysis of distribution and abundance, *Harper & Row*, NY, 449-487.

Krebs, C. J. (1989). Ecological Methodology. *Harper and Row*, NY.

Krebs, C. J. (1999). Ecological Methodology, 2nd edn., *Benjamin Cummings*, *University of British Columbia*, Vancouver.

Kroes, I., Lepp, P. W. & Relman, D. A. (1999). Bacterial diversity within the human subgingival crevice. *Proceedings of the National Academy of Sciences USA* **96**, 14547-14552.

Kruger, J., Singh, K., O'Neill, A., *et al*. (2002). Development of a microfluidic device for fluorescence activated cell sorting. *Journal of Micromechanichs and Microengineering* **12**, 486-494.

Kublanov, I. V., Bidjieva, S. K., Mardanov, A. V., *et al*. (2009a). *Desulfurococcus kamchatkensis* sp. nov., a novel hyperthermophilic protein-degrading archaeon isolated from a Kamchatka hot spring. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 1743-1747.

Kublanov, I. V., Perevalova, A. A., Slobodkina, G. B., *et al*. (2009b). Biodiversity of thermophilic prokaryotes with hydrolytic activities in hot springs of Uzon Caldera, Kamchatka (Russia). *Applied Environmental Microbiology* **75**, 286-291.

Kumar, S., Tamura, K., Nei, M. (2004). MEGA3: an integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

Kurapova, A. I., Zenova, G. M., Orleanskii, V. K, *et al*. (2008). Mesophilic and thermotolerant actinomycetes in strongly heated soils. *Moscow University Soil Science Bulletin* **63**, 142-147.

Kuwabara, T., Minaba, M., Ogi, N., *et al*. (2007). *Thermococcus celericrescens* sp. nov., a fast-growing and cell-fusing hyperthermophilic archaeon from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* **57**, 437-443.

Kvist, T., Mengewein, A., Manzei, S., *et al*. (2005). Diversity of thermophilic and non-thermophilic Crenarchaeota at 80 °C. *FEMS Microbiology Letters* **244**, 61-68.

Kyrpides, N., K. & Olsen, G. (1999). Archaeal and bacterial hyperthermophiles: Horizontal gene exchange or common ancestry? *Trends in Genetics* **15**, 298-299.

Lage, J. M., Leamon, J. H., Pejovic, T., *et al*. (2003). Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome research* **13**, 294-307.

Landrum, J. T., Bennett, P. C., Engel, A. S., *et al*. (2009). Partitioning geochemistry of arsenic and antimony, El Tatio geyser field, Chile. *Applied Geochemistry* **24**, 664-676.

Lartillot, N., Brinkmann, H. & Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology* **7,** S4.

Lau, M. C. Y., Aitchison, J. C., & Pointing, S. B. (2009). Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles life under extreme conditions* **13**, 139-149.

Lazarevic, V., Whiteson, K., Huse, S., *et al*., (2009). Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *Journal of Microbiology Methods* **79**, 266-271.

Lee, N., Nielsen, P. H., Andreasen, K. H., *et al*. (1999). Combination of fluorescent *in situ* hybridization and microautoradiography - a new tool for structure-function analyses in microbial ecology. *Applied and Environmental Microbiology* **65**, 1289-1297.

Lexa, M., Horak, J. & Brzobohaty, B. (2001). Virtual PCR. *Bioinformatics* **17**, 192-193.

Lexerød, N. L., & Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management* **222**, 17-28.

Li, H., Gyllenstein, U. B., Cui, X., *et al*. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**, 414-417.

Liang, Y., Yesuf, J., Schmitt, S., *et al*. (2009). Study of cellulases from a newly isolated thermophilic and cellulolytic *Brevibacillus* sp. strain JXL. *Journal of Industrial Microbiology and Biotechnology* **36**, 961-970.

Logsdon, J. M., & Faguy, D. M. (1999). *Thermotoga* heats up lateral gene transfer. *Current Biology* **9**, R747-R751.

Lopez, P., Forterre, P. & Philippe, H. (1999). The root of the tree of life in the light of the covarion model. *Journal of Molecular Evolution* **49**, 496-508.

Lynch, M. & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-73.

MacDonnell, M. T. & Colwell, R. R. (1985). Phylogeny of the Vibrionaceae and recommendation for two new genera, *Listonella* and *Shewanella*. *Systematic and Applied Bacteriology* **6**, 171-182.

MacGregor, B. J., Moser, D. P., Alm, E. W., *et al*. (1997). Crenarchaeota in Lake Michigan sediment. *Applied and Environmental Microbiology* **63**, 1178-1181.

Magalhaes, C., Bano, N., Wiebe, W. J., *et al*. (2008). Dynamics of nitrous oxide reductase genes (*nos*Z) in intertidal rocky biofilms and sediments of the Douro river estuary (Portugal), and their relation to N-biogeochemistry. *Microbial Ecology* **55**, 259-269.

Magurran, A.E., (1988). Ecological diversity and its measurement. *Princeton University Press*, NJ.

Magurran, A. E. (2004). Measuring biological diversity. *Blackwell Science*, *Oxford*.

Makarova, K. S. & Koonin, E. V. (2005). Evolutionary and functional genomics of the Archaea. *Current Opinion in Microbiology* **8**, 586-594.

Margalef, R. (1957). Information theory in ecology. *Generation Systems* **3**, 36-71.

Marsh, T. L. (1999). Terminal-restriction fragment length polymorphism (T-RFLP): An emerging method for characterizing diversity among homologous populations of amplicons. *Current Opinion in Microbiology* **2**, 323-327.

Marsh, R. M. & Norris, P. R. (1985). The isolation of some thermophilic, autotrophic iron- and sulfur-oxidising bacteria. *Federation of European Microbiological Societies Microbiology Letters* **17**, 311-315.

Marybeth, A. P., Kristina, D. R., Keith, R. S., *et al*. (2001). Hyperthermophilic enzymes part A: Continuous cultivation of hyperthermophiles. *Methods in Enzymology* **330**, 31-40.

Maxam. A. M., Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of National Academy of Sciences USA* **74**, 560-564.

Maynard, S. (1998). Evolutionary Genetics. *Oxford University Press*, Oxford.

McCaig, A. E., Glover, L. A., & Prosser, J. I., (1999). Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Applied and Environmental Microbiology* **65**, 1721-1730.

McCliment, E. A., Voglesonger, K. M., O'Day, P. A., *et al*. (2006). Colonization of nascent, deep-sea hydrothermal vents by a novel archaeal and nanoarchaeal assemblage. *Environmental Microbiology* **8**, 114-125.

McKernan, K., Peckham, H. E., Costa, G. L., *et al*. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two base encoding. *Genome Research* **19**, 1527-1541.

Meerman, J. (2004). Rapid ecological assessment Columbia river forest reserve past hurricane Iris. Ya`axché conservation trust and Toledoinstitute for development and environment.

Merrill, L. & Halverson, L. J. (2002). Seasonal variation in microbial communities and organic malodor indicator compound concentrations in various types of swine manure storage systems. *Journal of Environmental Quality* **31**, 2074-2085.

Miroshnichenko, M. L., Tourova, T. P., Kolganova, T. V., *et al*. (2008). *Ammonifex thiophilus* sp. nov., a novel hyperthermophilic anaerobic bacterium from Kamchatka hot spring. *International Journal of Systematic and Evolutionary Microbiology* **58**, 2935-2938.

Mori, K., Maruyama, A., Urabe, T., *et al*. (2008). *Archaeoglobus infectus* sp. nov., a novel thermophilic, chemolithoheterotrophic archaeon isolated from a deep-sea rock collected at Suiyo Seamount, Izu-Bonin Arc, western Pacific Ocean. *International Journal of Systematic and Evolutionary Microbiology* **58**, 810-816.

Mori, K., Yamaguchi, K., Sakiyama, Y., *et al*. (2009). *Caldisericum exile* gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. nov. and Caldisericia classis nov. *International Journal of Systematic and Evolutionary Microbiology* **59**, 2894-2898.

Moter, A. & Göbel, U. B. (2000). Fluorescence *in situ* hybridization (FISH) for direct visualization of microorganisms. *Journal of Microbiological methods* **41**, 85-112.

Moyer, C. L., Tiedje, J. M., Dobbs, F. C., *et al*. (1998). Diversity of deep-sea hydrothermal vent Archaea from Loihi Seamount, Hawaii. *Deep Sea Research Part II* **45**, 303-317.

Müller, A. K., Westergaard, K., Christensen, S., *et al*. (2001). The effect of long-term mercury pollution on the soil microbial community. *FEMS Microbiological Ecology* **36**, 11-19.

Mushegian, A. & Koonin, E. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of National Academy of Sciences USA* **93**, 10268-10273.

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. *Applied and Environmental Microbiology* **3**, 695-700.

Muyzer, G. & de Waal, E. C. (1994). Determination of the genetic diversity of microbial community using DGGE analysis of PCR amplified 16S rRNA. *NATO ASi Series* **G35**, 207-214.

Muyzer, G., Teske, A., Wirsen, C. O., *et al*. (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rRNA fragments. *Archives of Microbiology* **164**, 165-172.

Muyzer, G., Hottentrager, S., Teske, A., *et al*. (1996). Denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA — a new molecular approach to analyse the genetic diversity of mixed microbial communities. In A Akkermans, et al (ed). *Molecular Microbial Ecology Manual, Kluwer Academic Publishers*, Nowell, MA, 1-23.

Muyzer, G., Brinkhoff, T. N., Nübel, U., *et al*.. (1997). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In: Akkermans A. D. L., van Elsas, J. D. & de Bruijn, F. J. (eds). *Molecular Microbial Ecology Manual, Kluwer Academic Publishers*, Dordrecht, The Netherlands **3.4.4**, 1-27.

Muyzer, G. & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology, *Antonie Van Leeuwenhoek* **73**, 127-141.

Myers, R. M., Sheffield, V. C. & Cox, D. R. (1989). Mutation detection by PCR, GC-clamps, and denaturing gradient gel electrophoresis. In: Erlich, H. A. (ed). PCR technology: Principles and applications for DNA amplification, *Stockton Press*, NY, 71-88.

Nakagawa, T. & Fukui, M. (2003). Molecular characterization of community structures and sulfur metabolism within microbial streamers in Japanese hot springs. *Applied and Environmental Microbiology* **69**, 7004-7057.

Naum, M. Brown, E. W. & Mason-Gamer, R. J. (2008). Is 16S rRNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae? *Journal of Molecular Evolution* **66**, 630-642.

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., *et al*. (2001). Taxonomic study of aerobic thermophilic bacilli: Descriptions of *Geobacillus subterraneus* gen nov. sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatu*s, *Bacillus thermoleovorans Bacillus kaustophilus*, *Bacillus thermoglucosidasius*, *Bacillus thermodenitrifican*s to *Geobacillus* as *Geobacillus stearothermophilus*, *Geobacillus thermocatenulatus*, *Geobacillus thermoleovorans Geobacillus kaustophilus*, *Geobacillus thermoglucosidasius*, *Geobacillus thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* **51**, 433-46.

Nelson, K. E. (2003). The future of microbial genomics. *Environmental Microbiology* **5**, 1223-1225.

Nelson, K. E., Clayton, R. A., Gill, S. R., *et al*. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.

Niederberger, T. D., Götz, D. K., McDonald, I. R., *et al*. (2006). *Ignisphaera aggregans* gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *International Journal of Systematic and Evolutionary Microbiology* **56**, 965-971.

Nold, S. C., Kopczynski, E. D., Ward, D. M. (1996). Cultivation of aerobic chemoorganotrophic proteobacteria and gram-positive bacteria from a hot spring microbial mat. *Applied and Environmental Microbiology* **62**, 3917-3921.

Nübel, U., Garcia–Pichel, F., Khul, M., *et al*. (1999a). Quantifying microbial diversity morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Applied and Environmental Microbiology* **65**, 422-430.

Nübel, U., Garcia-Pichel, F., Kühl, M., and Muyzer, G. (1999b) Spatial scale and the diversity of benthic cyanobacteria and diatoms in a salina. *Hydrobiologia* **401**, 381-391.

O'Connor, M., Peifer, M., Bender, W. (1989). Construction of large DNA segmets in *Escherichia coli*. *Science* **244**, 1307-1312.

O'Donnell, T. & Head, I. M. (1997). Rapid enumeration of specific bacterial population: Applications in wastewater treatment. *Molecular Microbial Ecology*, *University of Newcastle upon Tyne*, 18-20.

Ochsenreiter, T., Selezi, D., Quaiser, A., *et al*. (2003). Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environmental Microbiology* **5**, 787-797.

Ogg, C. D., & Patel, **B. K. C. (2009).** *Caloramator australicus* sp. nov., a thermophilic anaerobic bacterium from the Great Artesian Basin of Australia aquifer. *International Journal of Systematic and Evolutionary Microbiology* **59**, 95-101.

Okubo, A., Sugiyama, S-i. (2009). Comparison of molecular fingerprinting methods for analysisof soil microbial community structure. *Ecological Research* **24**, 1399-1405.

Oremland, R.S. & Stolz, J.F. (2003). The ecology of arsenic. *Science* **300**, 939-944.

Oren, A. (1994). The ecology of extremely halophilic Archaea. *FEMS Microbiology Reviews* **13**, 415-440.

Oren, A. (2002). Halophilic microorganisms and their environments. In: Seckbach, J. (ed). Cellular origin: Life in extreme habitats and astrobiology. *Kluwer Academic Publishers*, Dordrecht, The Netherlands, 545-547.

Øvreås, L., Forney, L., Daae, F. L., *et al*. (1997). Distribution of bacterioplankton in meromictic Lake Sælenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Applied and Environmental Microbiology* **63**, 3367-3373.

Ouzounis, C. A., Kunin, V., Darzentas, N., *et al*. (2006). A minimal estimate for the gene content of the last universal common ancestor-exobiology from a terrestrial perspective. *Research in Microbiology* **157**, 57-68.

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

Paper, W., Jahn, U., Hohn, M. J., *et al*. (2007). *Ignicoccus hospitalis* sp. nov., the host of '*Nanoarchaeum equitans*'. *International Journal of Systematic and Evolutionary Microbiology* **57**, 803-808.

Paulson, T. G., Galipeau, P. C., *et al*. (1999). Loss of heterozygosity analysis using whole genome amplification, cell sorting, and fluorescence-based PCR. *Genome Research* **9**, 482-491.

Perevalova, A. A., Kolganova, T. V., Birkeland, N.-K., *et al*., (2008). Distribution of Crenarchaeota representatives in terrestrial hot springs of Russia and Iceland. *Applied and Environmental Microbiology* **74**, 7620-7628.

Philippe, H. & Laurent, J., (1998). How good are deep phylogenetic trees? *Current Opinion in Genetics and Development* **8**, 616-623.

Phoenix, V. R., Bennett, P. C., Summers, E. A., *et al*. (2006). Chilean high-altitude hot spring sinters: A model system for UV screening mechanisms by early precambrian cyanobacteria. *Geobiology* **4**, 15-28.

Pielou, E. C. (1969). An introduction to mathematical ecology. *John Wiley*, NY.

Petrosino, J. F., Highlander, S., Luna, R. A., G *et al*. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry* **55**, 856-866.

Poli, A., Romano, I., Cordella, P., *et al*. (2009). *Anoxybacillus thermarum* sp. nov., a novel thermophilic bacterium isolated from thermal mud in Euganean hot springs, Abano Terme, Italy. *Extremophiles* **13**, 867-74.

Pommerville, J. C. (2010). Alcamo's fundamentals of microbiology. *Jones and Bartlett*, Sudbury, Massachusetts.

Pop, M. & Salzberg, S. L. (2008). Bioinformatics challenges of new sequencing technology. *Trends in Genetics* **24**, 142-149.

Portillo, M. C. & Gonzalez, J. M. (2009). Members of the Candidate Division OP10 are spread in a variety of environments. *World Journal of Microbiology and Biotechnology* **25**, 347-353

Portillo, M. C., Sririn, V., Kanoksilapatham, W., *et al*. (2009). Differential microbial communities in hot spring mats from Western Thailand. *Extremophiles life under extreme conditions* **13**, 321-31

Potvin, M. & Lovejoy, C. (2009). PCR-Based Diversity estimates of artificial and environmental 18S rRNA gene libraries. *Journal of Eukaryotic Microbiology* **56**, 174-181.

Prieur, D., Erauso, G. & Jeanthon, C. (1995). Hyperthermophilic life at deep-sea hydrothermal vents. *Planetary and Space Science* **43**, 115-122.

Purdy, K. J., Cresswell-Maynard, T. D., Nedwell, D. B., *et al*. (2004). Isolation of haloarchaea that grow at low salinities. *Environmental Microbiology* **6**, 591-595.

QingYun, Y. & YuHe, Y. (2011). Metagenome-based analysis: A promising direction for planktonecological studies. *Science China Life Sciences* **54**, 75-78.

Qiu, X., Wu, L., Huang, H., *et al*. (2001). Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology* **67**, 880-887.

Raghunathan, A., Ferguson, H. R., Bornarth, C. J., *et al*. (2005). Genomic DNA amplification from a single bacterium. *Applied and Environmental Microbiology* **71**, 3342-3347.

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

Ranjard, L., Poly, F. & Nazaret, S. (2000). Monitoring complex bacterial communities using culture-independent molecular techniques. *Application to soil environment Research Microbiology* **151**, 167-177.

Reigstad, L. J., Jorgensen, S. L., Schleper, C. (2010). Diversity and abundance of *Korarchaeota* in terrestrial hot springs of Iceland and Kamchatka. *ISME Journal* **4**, 346-356.

Reysenbach, A. L., Ehringer, M. & K., H. (2000). Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: Another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles* **4**, 61-67.

Reysenbach, A. L., Giver, L. J., Wickham, G. S., *et al*. (1992). Differential amplification of ribosomal-RNA genes by polymerase chain-reaction. *Applied and Environmental Microbiology* **58**, 3417-3418.

Reysenbach, A.-L. & Pace, N. R. (1995). Reliable amplification of hyperthermophilic archaeal 16S rRNA genes by the polymerase chain reaction. In: Robb, F. T. & Place, A. R. (eds). Archaea: A laboratory manual: Thermophiles, *Cold Spring Harbour Laboratory Press, ColdSpring Harbour,* NY, 101-107.

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

Robe, P., Nalin, R., Capellano, C., *et al*. (2003). Extraction of DNA from soil. *European Journal of Soil Biology* **39**, 183-190.

Robertson, C., Harris, J., Spear, J., *et al*. (2005). Phylogenetic diversity and ecology of environmental Archaea. *Current Opinion in Microbiology* **8**, 638-42.

Rochelle, P. A., Cragg, B. A., Fry, J. C., *et al*. (1994). Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. *FEMS Microbiology Ecology* **15**, 25-226.

Rondon, M. R., August, P. R., Bettermann, A. D., *et al*. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* **66**, 2541-2547.

Rusch, A., Walpersdorf, E., de Beer, D., *et al*. (2005). Microbial communities near the oxic/anoxic interface in the hydrothermal system of Vulcano Island, Italy. *Chemical Geology* **224**, 169-182.

Sager, P. E. & Hasler, A. D. (1969). Species diversity in lacustrine phytoplankton. (I) The components of the index of diversity from Shannon's formula. *The American Naturalist* **103**, 51-59.

Saiki, R. K., Gelfand, D. H., Stoffel, S., *et al*. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Saiki, R. K., Scharf, S., Faloona, F., *et al*. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Saitou N & Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Salles, J. F., van Veen, J. A. & van Elsas, J. D. (2004). Multivariate analyses of *Burkholderia* species in soil: Effect of crop and land use history. *Applied and Environmental Microbiology* **70**, 4012-4020.

Sanger, F. & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 441-448.

Santegoeds, C. M., Nold, S. C. & Ward, D. M. (1996). Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. *Applied and Environmental Microbiology* **62**, 3922-3928.

Sapp, M., Wichels, A., Wiltshire, K. H., *et al*. (2007). Bacterial community dynamics during the winter-spring transition in the North Sea. *FEMS Microbiology Ecology* **59**, 622-637.

Schleper, C., Jurgens, G. & Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nature Reviews Microbiology* **3**, 479-488.

Schloss, P. D., Westcott, S.L., Ryabin, T. *et al*. (2009). Introducing mothur: open source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

Schloter, M., Lebuhn, M., Heulin, T., *et al*. (2000). Ecology and evolution of bacterial microdiversity. *FEMS Microbiology Reviews* **24**, 647-660.

Schmid, M. C, Hooper, A. B., Klotz, M. G., *et al*. (2008). Environmental detection of octahaem cytochrome c hydroxylamine/hydrazine oxidoreductase genes of aerobic and anaerobic ammoniumoxidizing bacteria. *Environmental Microbiology* **10**, 3140-3149.

Schönhuber, W., Bourhis, G.-L., Tremblay, J., *et al*. (2001). Utilization of tmRNA sequences for bacterial identification. *BioMed Central Microbiology* **20**, 471-2180.

Schramm, A. & Amann, R. (1999). Nucleic acid-based techniques for analyzing the diversity, structure, and dynamics of microbial communities in wastewater treatment. *I*n: Rehm et al (eds). Biotechnology: A multi volume references 88 comprehensive treatise. *Environmental processes I* **11a**, 85-108.

Schrenk, M. O., Kelley, D. S., Delaney, J. R., *et al*. (2003). Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Applied and Environmental Microbiology* **69**, 3580-3592.

Schütte, U. M. E., Abdo, Z., Bent, S. J., *et al*. (2008). Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbiology and Biotechnology* **80**, 365-380.

Seber, G. A. F. (1973). The estimation of animal abundance and related parameters. Griffin, London, England.

Sharma, A., Pandey, A., Shouche, Y. S., *et al*. (2009). Characterization and identification of *Geobacillus* spp. isolated from Soldhar hot spring site of Garhwal Himalaya, India. *Journal of Basic Microbiology* **49**, 187-194.

Shannon, C. E. A (1948). Mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423.

Shannon, C.E. & Weaver, W. (1949). The mathematical theory of communication. The University of Illinois Press, Urbana.

Shi, B. & Xia, X. (2003). Changes in growth parameters of *Pseudomonas pseudoalcaligenes* after ten months culturing at increasing temperature. *FEMS Microbiology Ecology* **45**, 127-134.

Shima, S., Yanagi, M., & Saiki, H. (1994). The phylogenetic position of *Hydrogenobacter acidophilus* based on 16S rRNA sequence analysis. *FEMS Microbiology Letters* **119**, 119-122.

Shimakita, T., Yamamoto, H., Naramura, T., *et al*. (2007). Rapid Count of Microbial Cells in Dialysate. *Therapeutic Apheresis and Dialysis* **11**, 363-369.

Siebers, B., Klenk, H. -P. & Hensel, R. (1998). PPi-dependent phosphofructokinase from *Thermoproteus tenax*, an archaeal descendant of an ancient line in phosphofructokinase evolution. *Journal of Bacteriology* **180**, 2137-2143.

Silver, S. & Phung. L.T. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Applied and Environmental Microbiology* **71**, 599-608.

Simpson, G. G. (1949). The meaning of evolution. *Yale University Press*.

Smalla, K., Oros-Sichler, M., Milling, A., *et al*. (2007). Bacterial diversity of soils assessed by DGGE, T-RFLP and SSCP fingerprints of PCR-amplified 16S rRNA gene fragments: Do the different methods provide similar results? *Journal of Microbiological Methods* **69**, 470-479.

Smith, B. & Wilson, J.B. (1996). A consumer's guide to evenness indices. *Oikos* **76**, 70-82.

Sneath, P. H. A., Mair, N. S., Sharpe, M. E., *et al*. (eds) (1986). Bergey's Manual of Systematic Bacteriology, *Williams* & *Wilkins*, Baltimore, **2**.

Sompong, U., Castenholz, R. W., Anuntalabhochai, S., *et al*., (2006). Genetical diversity of *Mastigocladus laminosus* in Ranong Hot Spring, southern part of Thailand, Chiang. *Mai Journal of Science* **33**, 363-70.

Song, Z. Q., Zhi, X. Y, Li, W. J., *et al*. (2009). Actinobacterial diversity in hot springs in Tengchong (China), Kamchatka (Russia) and Nevada (USA). *Geomicrobiology Journal* **26**, 256-263.

Spellerberg, I. F. & Fedor, P. J. (2003). A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' index. *Global Ecology and Biogeography* **12**, 177-179.

Sridevi, N. & Prabhune, A. A. (2009). *Brevibacillus* sp: A novel thermophilic source for the production of bile salt hydrolase. *Applied Biochemistry and Biotechnology* **157**, 254-262.

Stach, J. E. M., Bathe, S., Clapp, J. P., *et al*. (2001). PCR-SSCP comparison of 16S rRNA sequence diversity in soil DNA obtained using different isolation and purification. *FEMS Microbiology Ecology* **36**, 139-151.

Stackebrandt, E. & Lang, E. (1998). Molecular detection and identification of prokaryotes. *OECD workshop molecular methods for safe drinking water*, 1-10.

Stahl, D. A., Key, R., Flesher, B., *et al*. (1992). The phylogeny of marine and freshwater caulobacters reflects their habitat. *Journal of Bacteriology* **174**, 2193-2198.

Stakenborg, T., Vicca, J., Butaye, P., *et al*. (2005). Evaluation of amplified rDNA restriction analysis (ARDRA) for the identification of mycoplasma species. *BioMed Central Infectious Diseases* **5**, 46.

Staley, J. T., Gunsalus, R.P., Lory, S., *et al*. (2007). Microbial life: Microbial evolution and diversity, 2nd edn., Part V, Chapter 17. Taxonomy of Bacteria and Archaea, 494-496.

Stetter, K. O. & Zillig, W. (1985). *Thermoplasma* and the thermophilic sulfur-dependent archaebacteria. *Academic Press*, NY,

Stetter, K. O. (1986). Bakterielles Leben an der oberen Temperaturgrenze. *Forum Mikrobiologie* **9**, 15-19.

Stetter, K. O. (1989). Extremely thermophilic chemolithoautotrophic archaebacteria: Autotrophic Bacteria. In: Schlegel, H. G. & Bowien, B. (eds.). *Science Tech Publishers Madison, WIand Springer Verlag Berlin*, Heidelberg, NY, London, Paris, Tokyo, 167-176.

Stetter, K. O. (1992). Life at the upper temperature border. In: Tran Than Van, J., Tran Than Van, K., Mounolou, J. C. *et al*. (eds). *Frontiers of life*, 195-219.

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

Stetter, K. O. (1998). Hyperthermophiles: Isolation classification and properties. In: Horikoshi, K. & Grant, W. D. (eds). Extremophiles: Microbial life in Extreme Environments. *Wiley-Liss*, 1-24.

Stetter, K. O. (1999). Smallest cell size within hyperthermophilic Archaea ("Archaebacteria"). In: NRC Steering Group Astrobiology Space Student Board. Size limits of very small microorganisms. Proceedings of a Workshop, Washington, DC. *National Academy of Sciences Press*, 68-73.

Stetter, K. O. (2006a). History of discovery of the first hyperthermophiles. *Extremophiles: life under extreme conditions* **10**, 357-362.

Stetter, K. O. (2006b). Hyperthermophiles in the history of life. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1837-1843.

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

Stetter, K. O., Fiala, G., Huber, G., *et al*. (1989). Hyperthermophilic bacterial communities within terrestrial and marine hydrothermal areas. Recent advances in microbial ecology. In: Hattori, T., Ishida, Y., Maruyama, Y. *et al*. (eds). Proceedings of the 5th International Symposium of Microbial Ecology (ISME-5). *Japan Scientific Society Press*, Tokyo, Japan, 51-54.

Stöhr, R., Waberski, A., Völker, H., *et al*. (2001). *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter* and proposal of the reclassification of *Hydrogenobacter acidophilus* as *Hydrogenobaculum acidophilum* gen. nov., comb. nov., in the phylum '*Hydrogenobacter*/*Aquifex*'. *International Journal of Systematic and Evolutionary Microbiology* **51**, 1853-*1862.*

Strachan, T., & Read, A. P. (1999). Human Molecular Genetics, 3rd edn., *Garland Science, Taylor and Francis Group*, London & NY

Suzuki, M. & Giovannoni, S. J. (1996). Bias caused by template annealing in the amplification mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology* **62**, 625-630.

Suzuki, M., Cui, Z. J., Ishii, M., *et al*. (2001). Nitrate respiratory metabolism in an obligately autotrophic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus* TK-6. *Archives of Microbiology* **175**, 75-78.

Suzuki, Y., Matsui, H., Tsujimoto, Y., *et al*. (2009). Enzymatic degradation of fibroin fiber by a fibroinolytic enzyme of *Brevibacillus thermoruber* YAS-1. *Journal of Bioscience and Bioengineering* **108**, 211-215.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution* **24**, 1596-1599.

Takacs, C. D., Ehringer, M., Favre, R., *et al*. (2001). Phylogenetic characterization of the blue filamentous bacterial community from an Icelandic geothermal spring, *FEMS Microbiology Ecology* **35**, 123-128.

Takahata,Y., Nishijima, M., Hoaki, T., *et al*. (2001). *Thermotoga petrophila* sp. nov. and *Thermotoga naphtophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. *International Journal of Systematic and Evolutionary Microbiology* **51**,1901-1909.

Takai, K. & Horikoshi, K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285-1297.

Tassi, F., Martinez, C., Vaselli, O., *et al*. (2005). Light hydrocarbons as redox and temperature indicators in the geothermal field of El Tatio (Northern Chile). *Applied Geochemistry* **20**, 2049-2062.

Terahara, T., Hoshino, T., Tsuneda, S., *et al*. (2004). Monitoring the microbial population dynamics at the start-up stage of wastewater treatment reactor by terminal restriction fragment length polymorphism analysis based on 16S rRNA and rRNA gene sequences. *Journal of Bioscience and Bioengineering* **98**, 425-428.

Teske, A. & Sørensen, K. B. (2008). Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME Journal* **2**, 3-18.

Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Tin, S., Bizzoco, R. W. & Kelley, S. T. (2011). Role of the terrestrial subsurface in shaping geothermal spring microbial communities. *Environmental Microbiology Reports* **3**, 491-499.

Traag, B. A., Driks, A., Stragier, P., *et al*. (2010). Do mycobacteria produce endospores? *Proceedings of the National Academy of Sciences USA* **107**, 878-881.

Trent, J. D., Chastain, R. A. & Yayanos, A. A. (1984). Possible artefactual basis of apparent bacterial growth at 250 °C. *Nature* **307**, 737-740.

Treves, D. S., Manning, S. & Adams, J. (1998). Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* **15**, 789-797.

Tringe, S. G. & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature reviews Genetics* **6**, 805-814.

Tringe, S. G., von Mering, C., Kobayashi, A., *et al*. (2005). Comparative metagenomics of microbial communities. *Science* **308**, 554-557.

Tyson, G. W., Chapman, J., Hugenholtz, P., *et al*. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

Uzarraga, R., Auria, R., Davidson, S., *et al*. (2011). New cultural approaches for microaerophilic hyperthermophiles. *Current Microbiology* **62**, 346-50.

Valentine, D. L. (2002). Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. *Antonie van Leeuwenhoek* **81**, 271-282.

Vaneechoutte, M. & Heyndrickx, M. (2001). Application and analysis of ARDRA patterns in bacterial identification, taxonomy and phylogeny. In: Dijkshoorn, L., Towner, K. J. & Struelens, M. (eds). New approaches for the generation and analysis of microbial typing data. *Elsevier*, The Netherlands, 211-247.

van Niel, E. W. J., Budde, M. A. W., de Haas, G. G., *et al*. (2002). Distinctive properties of high hydrogen producing extreme thermophiles, *Caldicellusiruptor saccharolyticus* and *Thermotoga elfii*. *International Journal of Hydrogen Energy* **27**, 1391-1398.

Van Ooteghem, S. A., Beer, S. K. & Yue, P. C. (2001). Hydrogen production by the thermophilic bacterium *Thermotoga neapolitana*. In: Proceedings of the 2001 DOE hydrogen program review, document no. NREL/CP-570-30535. Golden, CO: National Renewable Energy Laboratory.

Venter, J. C., Remington, K., Heidelberg, J. F., *et al*. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

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

Walter, N. A., Bottomly, D., Laderas, T., (2009). High throughput sequencing in mice: A platform comparison identifies a preponderance of cryptic SNPs. *BMC Genomics* **10**, 379

Wang, G. C.-Y., Wang, Y. (1997). Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Applied and Environmental Microbiology* **63**, 4645-4650.

Ward, N. & Fraser, C. M. (2005). How genomics has affected the concept of microbiology. *Current Opinion in Microbiology* **8**, 564-571.

Warren, R. L. Sutton, G. G., Jones, S. J. M., *et al*. (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* **23**, 500-501.

Waters, E., Hohn, M. J., Ahel, I., *et al*. (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proceedings of the National Academy of Sciences USA* **100**, 12984-12988.

White, R. H. (1984). Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 °C. *Nature* **310**, 430-432.

Whitaker, R. J., Grogan, D. W. & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976-978.

Widmer, F., Fliеβbach, A., Laczko, E., *et al*. (2001). Assessing soil biological characteristics: A comparison of bulk soil community DNA-, PLFA-and BiologTM-analyses. *Soil Biology and Biochemistry* **33**, 1029-1036.

Winker, S. & Woese, C. R. (1991). A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Systematic and Applied Microbiology* **13**, 161-165.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews* **51**, 221-71.

Woese, C. R. & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5088-5090.

Wright, B .E. (2000). A Biochemical mechanism for nonrandom mutations and evolution. *Journal of Bacteriology* **182**, 2993-3001.

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

Yannarell, A. C. & Triplett, E.W. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *Applied and Environmental Microbiology* **71**, 227-239.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp 18 and pUC19 vectors. *Gene* **33**, 103-119.

Zenova, G. M., Kurapova, A. I., Lysenko, A.M., Zvyagintsev, D. G. (2009). The structural-functional organization of thermotolerant complexes of actinomycetes in desert and volcanic soils *Eurasian Soil Science* **42**, 531-535.

Zenova, G. M., Manucharova, N. A., &. Zvyagintsev, D. G. (2011). Extremophilic and extremotolerant actinomycetes in different soil types. *Eurasian Soil Science* **44**, 417-436.

Zhang, H.-B., Wu, C. (2001). BAC as tools for genome sequencing. *Plant Physiology and Biochemistry* **39**, 195-209.

Zhang, K., Martiny, A. C., Reppas, N. B., *et al*. (2006). Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology* **24**, 680-686.

Zhao, H., Kassama, Y., Young, M., *et al*. (2004). Differentiation of *Micromonospora* isolates from a coastal sediment in Wales on the basis of Fourier transform infrared spectroscopy 16S rRNA sequence analysis, and the amplified fragment length polymorphism technique. *Applied and Environmental Microbiology* **70**, 6619-6627.

Zheng, D., Alm, E. W., Stahl, D. A. & Raskin, L. (1996). Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Applied and Environmental Microbiology* **62**, 4504-4513.

Zimmer, C. & Wähnert, U. (1986). Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic materials. *Progressive Biophysical Molecular Biology* **47**, 31-112.

Zuckerkandl, E., Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**, 357-366.