Genome sequence and functional comparison of *Thermus* NMX2 A.1

by

Nokuthula Tlalajoe

May 2013

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

UFS
UV

Genome sequence and functional comparison of *Thermus* NMX2 A.1

by

Nokuthula Tlalajoe

B.Sc. Hons. (UFS)

Submitted in fulfillment of the requirements for the degree

MAGISTER SCIENTIAE

In the Faculty of Natural and Agricultural Sciences

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

Bloemfontein

South Africa

May 2013

Supervisor: Prof. E. van Heerden

Co-Supervisors: Prof. D. Litthauer (Internal examiner)

Dr. E. Brzuszkiewicz (External examiner)

I hereby dedicate this dissertation to my late grandmother Elizabeth Limakatso Tlalajoe; for molding me into a confident, independent and strong individual at a very young age

"Imagination is more important than knowledge."-Einstein

Acknowledgments

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

God: The main reason I am still here, because of your word Isaiah 41 vs. 9-10

Prof. E. van Heerden: I am not even sure where to start, what to mention and what to leave out, because everything you did the role you played, your presence, all of it made a great impact. Thank you for the whole package of support, the exposure, for stretching me to realise my potential for the fun activities while working on this project and most of all thank you for the financial support in carrying out my career

Prof. D. Litthuer: WOW…..the “coolest” Prof. ever, I am honoured to have been supervised by you. My friends started seeing me as super smart just because I was under your supervision. Thank you for your patience, for sitting with me for hours getting me on board with all the Bioinformatics tools and also thank you for the nominations for the NRF grants

Dr. E. Brzuszkiewicz: For helping me understand the essentials of joining contigs, explaining everything in depth and making sure I follow, also for replying to the non-ending emails seeking help; thank you

Family members: Thank you for being there for me, even though you never understood as to who many times do I need to graduate in order for me to be done with school, well I guess those are the perks of being the first one in the family to achieve beyond a degree

Mr. Petso Mokhatla: Thank you for your kindness, patience, support and mostly for your love

Lecturers and colleagues: Inputs and advice on how to go about experiments, encouragements and sharing of wisdom

Fellow students and friends: Thank you for making the lab feel welcoming, tea time were the best times ever getting together and talking about anything and everything not to mention the jokes that brought smiles onto our faces and great laughter

Declaration

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

Nokuthula Tlalajoe (2006048512) Date: May 2013

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A

A	Absorbance
AIX	Ampicillin/IPTG/X-Gal

B

BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BiDi BLAST	Bi-directional Basic Local Alignment Search Tool

C

cm	centimeters
COG's	Clusters of Orthologous Groups of proteins

D

DNA	Deoxyribonucleic acid
°C	Degree Celsius
dNTP	di-Nucleotide triphosphate
DAPI	4', 6-diamidino-2-phenylindole

E

EDTA	Ethylene diaminetetraacetic acid
EMB	Eosin-methylene blue

G

g	Acceleration due to gravity
g/L	grams per litre
gDNA	Genomic DNA

H

HB	Heterotrophic broth
HGT	Horizontal Gene Transfer
h	hour

I

IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMG/ER	Integrated Microbial Genome/ Expert Review

K

KOG's	Eukaryotic Orthologous Groups of proteins
KEGG	Kyto Encyclopedia of Genes Genomes
km	kilo meters

L

LB	Luria-Bertani
L	Litre

M

MOPS	3-(N-Morpholino)-ethanesulfonic acid
Mr	Relative molecular mass
Min	minutes
mM	milli Molar
M	Molar
µL	microlitre
µM	micro Molar
Mg	Magnesium
mL	millilitre

MEGA	Molecular Evolutionary Genetics Analysis

N

NB	Nutrient broth
NCBI	National Centre for Biotechnology Information
nm	nanometers
ND	NanoDrop
ng	nano grams
nr database	non-reduntant database

O

OD	Optical density
ORFs	Open Reading Frames

P

PCR	Polymerase chain reaction

R

RPM	Revolution per minute
RNA	Ribosomal nucleic acid
RFLP	Restriction Fragment Length Polymorphism
rDNA	ribosomal Deoxyribonucleic acid
RDP	Ribosomal Database Project

S

SDS	Sodium Dodecyl Sulphate

T

TAE	Tris-Acetic acid-EDTA
TYG	Tryptone, yeast extract, glucose

Tris-HCl	Tris-hydrochloric acid
Taq	*Thermus aquaticus*

U

UV	Ultra violet
U	Unit

V

VRBA	Violet red bile agar
v/v	Volume per volume
V_{max}	Maximum velocity
V	voltage

W

w/v	Weight per volume

X

X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranosidehosphate
X g	Times gravity

CHAPTER 1

Literature review

1. Introduction

1.1 Thermophiles

Significant diversity has been observed in all living organisms across the existing ecological niches, including extreme environments which consisted of varies living organisms, such as acidophiles vs. alkaliphiles, piezophiles, halophiles and thermophiles vs. psychrophiles (Woese *et al.,* 1990). Prokaryotes have the ability to adapt themselves to a wide range of physicochemical surroundings, in which most may be considered as not being “normal” from human point of view.

Temperature (psychrophiles: -5°C to 20°C, mesophiles: 15°C to 45°C, thermophiles: 45°C to 80°C, and hyperthermophiles: ≥80°C; Figure 1.1) has been one of the established physical parameters used to separate what we consider as mesophilic organisms from thermophilic organisms. Thermophilic organisms are usually found in environments that are stable above 45°C and are mostly found in hydrothermal areas all over the world as well as sub surfaces such as deep mines (Cava *et al.,* 2009; Li *et al.,* 2005).

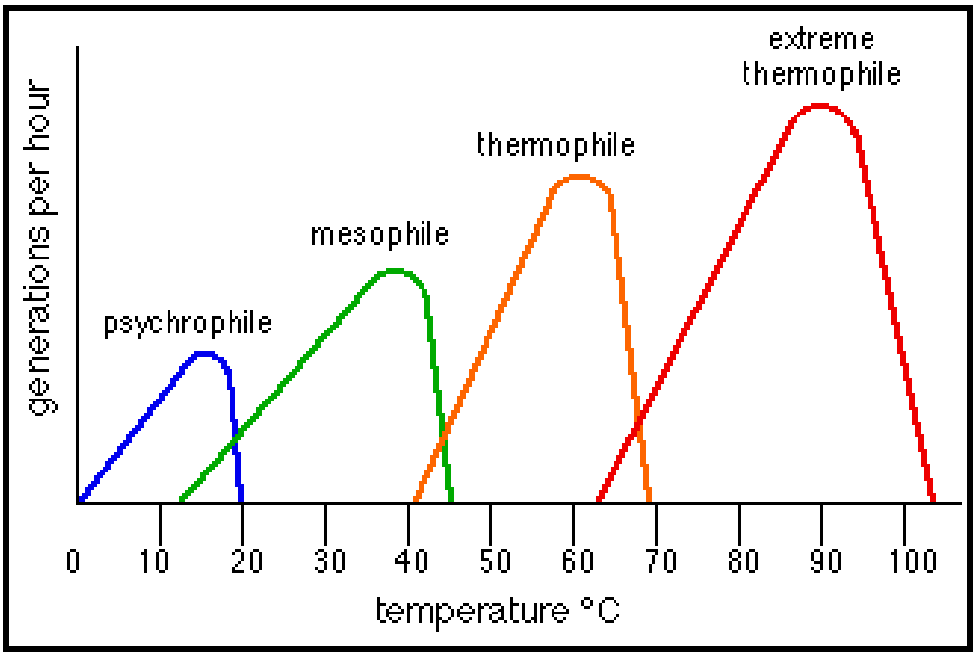


Figure 1.1: Four categories of microbes based on established temperatures classified under physical parameters. Copyright © The McGraw-Hill Companies, Inc.

Thermophilicity can be viewed from more than one angle, as there is more than one way in which organisms adapt to thermophilic environments. Natural occurrences where sporadic heating in specific environments occurs such as those produced by decomposition of sun-heated water surfaces for instance shallow lakes, present appropriate environments for the evolution of microorganisms that can grow optimally between 50°C and 60°C. However, these microorganisms may also have the ability to grow at temperatures below 50°C and are, therefore, known to be “moderate thermophiles” (Cava *et al.,* 2009).

In contrast, other microorganisms have the ability to grow optimally above 70°C and are unable to grow at temperatures below 50°C. These organisms are known as the “extreme thermophiles”. They are very few and definitely limited to the prokaryotic world. In addition, another biological restricted group known as the “hyperthermophiles” can grow optimally at temperatures above 85°C and only a few phylogenetic groups of the Bacteria domain can be considered hyperthermophiles (Cava *et al.,* 2009).

Literature review

Horizontal gene transfer (HGT) also known as lateral gene transfer refers to the exchange of genes between distantly related strains or even species. In nature, there are several examples of HGT between species of bacteria as well as in eukaryotes. Comparative analyses of the molecular data that are exploding from genome sequencing projects indicate that, throughout the life span; HGT has been a crucial driving force (Figure 1.2) behind the evolutionary changes of cellular life and the emergence of the three domains of life (Archaea, Bacteria, and Eucarya; Brown, 2003).

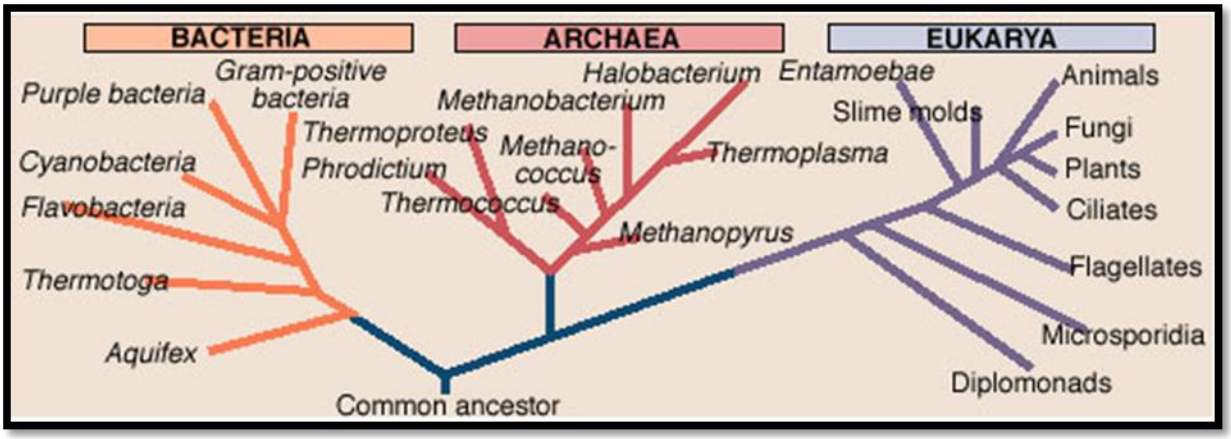


Figure 1.2: Schematic representation on major bacterial evolutionary lineages including all three domains.

The reason for believing in the occurrence of ancient HGT is simple in the sense that genes are not found where they are expected to be (Brown, 2003). Some bacterial genes are more closely related to versions in the archaea meanwhile some eukaryotic genes seem to have originated in bacteria rather than archaea (Brown, 2003). Nonetheless, lateral gene transfer appears to be the leading force driving that accelerated evolution in prokaryotes, granting bacteria the unique ability to rapidly adapt to various environmental changes for instance the extreme temperatures (César *et al.,* 2011).

It also enhances the understanding of natural competence apparatus within well studied species such as *T. thermophilus* and *T. scotoductus* (Gounder *et al.,* 2011). The ability of HGT in being able to be specific within physiologies of natural transformation models

also plays a great role. Internal and external environmental variables are known to be of great influence that delimits particular gene-exchange communities on basis of shared factors such as GC-contents, genome size and oxygen and carbon sources (César *et al.,* 2011).

1.2 Importance of thermophiles

Thermophilicity is an interesting subject for basic biology due to their biological properties. Extreme thermophiles contain enzymes and macromolecular complexes which have been selected as target models in structural biology because they are easier to crystallize than thier mesophilic counterparts (Jenney & Adams, 2008). The crystallization of high-resolution structures performed on the 70S ribosome from the *Thermus* sp. was observed with large complexes. This observation was concluded from the thermostability and structural rigidity characteristics during the adaptation process to thermophilic environments (Yusupov *et al*., 1989). This greater ability to crystallize has been the most important reason for the first structural genomic programs to choose subject models such as extreme thermophiles or hyperthermophiles (Jenney & Adams, 2008).

Thermophiles are also of great interest due to the adaptability of their macromolecules to function at high temperatures. Detail attention was paid to their enzymes, of which optimal activity was exhibited around the optimal growth temperature of the organism. Enzymes synthesized by thermophiles and hyperthermophiles are identified as thermozymes. These enzymes are resistant to irreversible inactivation at high temperatures. Thermozymes display an increased resistance to denaturing chemical reagents this encouraged the chemical industries to study their specific applications in biotransformation processes or even components of their final products (Li *et al.,* 2005).

Brock and Freeze (1969) discovered a thermophilic organism named *Thermus aquaticus*. This was a major turning point in the potential applications of thermozymes described for the use of the DNA polymerase in the polymerase chain reaction (PCR).

Literature review

The production of thermozymes is generally carried out on surrogate mesophilic microorganisms, from which the protein can be easily purified by its differential thermal stability with regard to most of the proteins from the mesophilic host (Cava *et al.,* 2009, da Costa *et al.,* 2006).

In addition to the interest in extreme thermophiles and hyperthermophiles mentioned above, only a few of them can be regarded as at least promising models for laboratory use. The reasons behind the scarcity of thermorphiles as laboratory models may vary, but one that tends to be common among them is the high temperature at which they grow. The high temperature results into difficulties to grow these organisms, because of dehydration of the inoculation medium, requirement of special solidifying components, thermo-resistant equipment, or glass plates (Cava *et al.,* 2009, da Costa *et al.,* 2006, Childers *et al.,* 1992). Other problems that are not so evident are related to the metabolism of the chosen laboratory models, in most cases requiring anaerobic growth conditions or low yield chemolitotrophic ways of gaining energy (Vieille, 2001).

However, a few strains of *T. thermophilus* appear to be suitable for genetic models as well as excellent candidates for their specific applications as hosts for the selection of thermostable mutants by direct mutation (Degryse *et al.,* 1978, Cava *et al.,* 2009). In this review the genus *Thermus* will be of discussion laying out aspects of interest that involves the variety of *Thermus* sp., their metabolism and the comparisons amongst them.

2. The genus *Thermus*

The genus *Thermus* was formerly proposed by Brock and Freeze (1969) with the description of *T. aquaticus* (Zhang et al., 2010). Ever since, hundreds of strains from the genus *Thermus* have been isolated from neutral to slightly basic thermal environments all around the world, generally in terrestrial hydrothermal areas where the temperature of the water range from 50°C to 70°C. Few are isolated from shallow marine hot springs, abyssal geothermal areas, self-heating piles of organic matter, and even

artificial thermal environments such as industrial heating systems (Zhang et al., 2010, Hjorleifsdottir et al., 2001).

Scientific classification of this genus was then agreed upon as follow, the kingdom is Bacteria; phylum was found to be related to *Deinococcus–Thermus* which explains the common special characteristics of being able to withstand extreme environments, the class is *Deinococci* followed by the order and family known to be the Thermales and Thermaceae, respectively (Figure 1.3; Griffiths and Gupta, 2004).

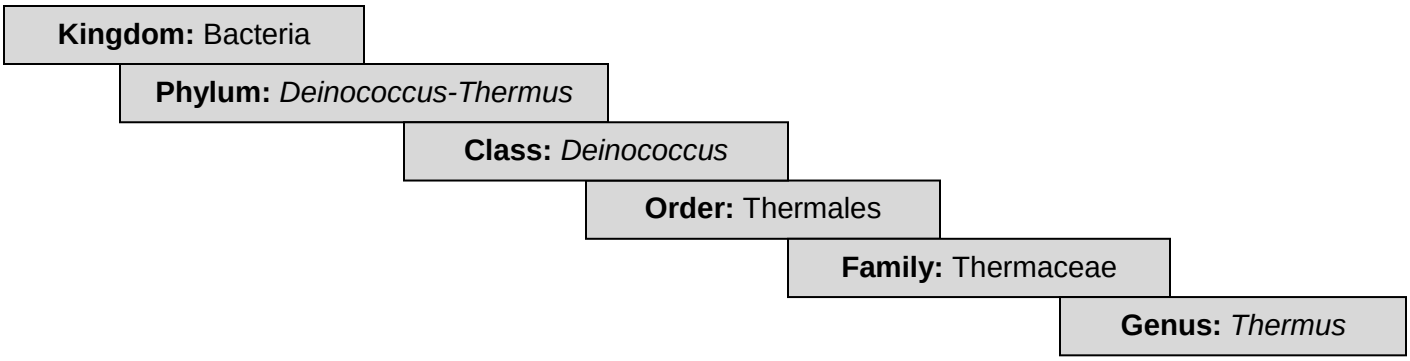


Figure 2.3: Taxonomic hierarchy of genus *Thermus* (Taken from Brock and Freeze 1969).

2.1 Taxonomy and phylogenetic characteristics

The genus *Thermus* comprises eight validly described species, namely, *T. aquaticus*, *T. thermophilus*, *T. filiformis*, *T. brockianus*, *T. oshimai*, *T. scotoductus*, *T. antranikianii* and *T. igniterrae*. However, there are many additional *Thermus* spp. that have been isolated but not taxonomically characterized. Among those is the species *T. caldophilus* GK24, which was originally isolated from the Kawamata hot springs, Tochigiken, Japan (Kim *et al.,* 2006).

Comparison of the 16S rDNA sequences from the strains of each of the eight validated *Thermus* spp. indicate the similarities to be in the range of 91.2%-96.4% (Figure 1.4) *T. oshimai* appeared to be the most dissimilar species of the genus *Thermus* as reflected by the similarity values based on the 16S rDNA sequences. The other seven species of

the genus *Thermus* showed similarities in the range of 94%-96% on their 16S rDNA sequences (da Costa *et al.,* 2006).

On a more closer basis, when comparison is done within each species the 16S rDNA sequences similarity values are in the range of 98.9%-99.7% for *T. aquaticus*, 99.9%-100% for *T. brockianus,* 99.2%-99.9% for *T. filiformis*, 99.8%-100% for *T. oshimai*, 98.7%-99.9% for *T. scotoductus*, 99.4%-100% for *T. thermophilus*, and 99.9%-100% for *T. igniterrae* and *T. antranikianii* (Chung *et al.,* 2000). The above values are obtained from comparing all published *Thermus* strains as well as a large number of unpublished *Thermus* strains (da Costa *et al.,* 2006).

Difficulties in differentiating new species from established species on organisms in this group are being experienced, due to the phenotypic characteristics of most species overlapping with each other (Table 1.1). The method used to identify diversity within groups of strains considered to be part of the same species is sometimes extensive and easily verified by the variation in the fatty acid composition (Table 1.2, da Costa *et al.,* 2006). For example, the species *T. thermophilus* has extremely variable fatty acid composition although some strains, namely HB27, HB8 and AT-62, share extremely high DNA-DNA hybridization values and 16S rDNA sequences similarity (Coimbro, 1989). This species was therefore easily distinguished from the strains of related species by its ability to grow at temperatures above 80°C and in media containing 3% NaCl. This was made possible by the presence of genes leading to the synthesis of the compatible solute mannosylglycerate which are necessary for growth of the organism when grown in media containing more than 1% NaCl (da Costa *et al.,* 2006, Zhang *et al.,* 2010).

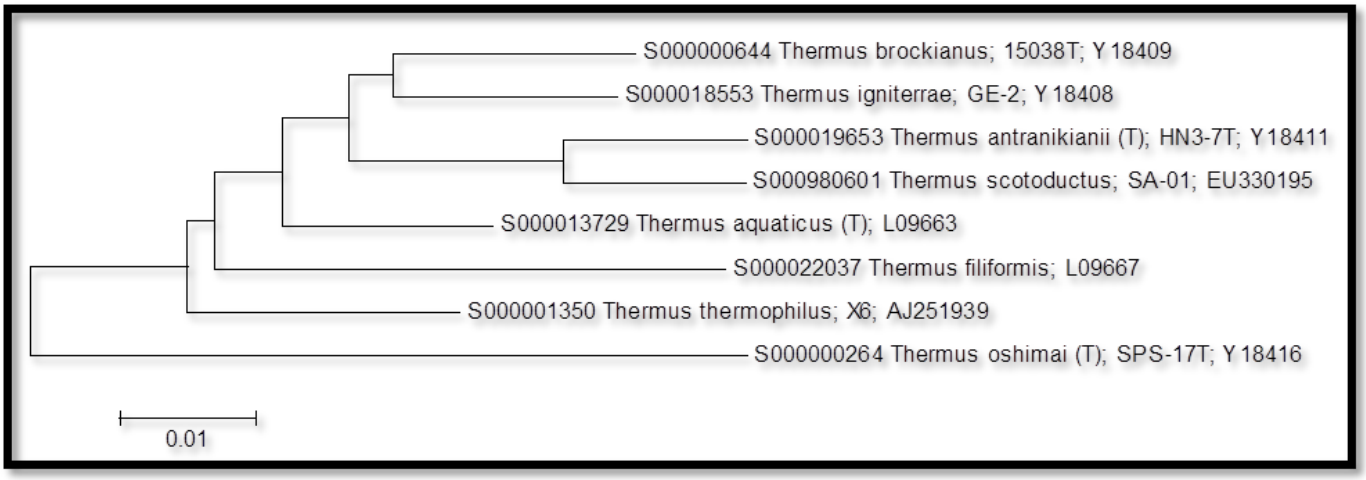


Figure 1.4: Phylogenetic dendogram based on 16S rDNA gene sequences comparisons of validly described type of species of the phylum *Deinococcus-Thermus*. Construction of dendogram was done from evolutionary distances using the neighbour-joining method. The scale bar represents 0.01 inferred nucleotide changes per 100 nucleotides (Taken from da Costa et al., 2006).

Table 1.1: Biochemical characteristics that differentiate the type of strains from the genus *Thermus* (Taken from da Costa *et al.,* 2006)

	T. aquaticus	*T. thermophiles*	*T. filiformis*	*T. scotoductus*	*T. brockianus*	*T. oshimai*	*T. igniterrae*	*T. antranikianii*
Characteristic	(YT-1^T)	(HB8^T)	(Wai33-A1^T)	(ITI-252^T)	(YSO38^T)	(SPS-17^T)	(RF-4^T)	(HN3-7^T)
Pigmentation	Yellow	Yellow	Yellow	White	Yellow	Yellow	Yellow	Yellow
Presence of:								
α-Galactosidase	-	+	+	+	+	+	-	+
β-Galactosidase	-	+	+	-	+	+	+	+
Degratation of ρ-nitrophenyl substrates:								
β-Glucopyranoside	+	+	+	-	+	+	+	+
Degradation of:								
Arbutin	w	+	+	w	+	+	+	+
Esculin	-	+	+	w	+	+	+	+
Hydrolysis of:								
Elastin	+	-	-	-	-	+	+	-
Starch	+	+	+	+	-	+	+	+
Fibrin	+	-	-	+	+	+	+	-
Gelatin	+	+	+	+	-	+	+	-
Casein	+	+	+	+	-	+	+	+

Tween 80	-	+	-	-	+	+	+	+
Reduction of nitrate	-	-	-	+	+	+	+	+
Growth at/in:								
80°C	-	+	-	-	-	-	-	+
82°C	-	+	-	-	-	-	-	-
2% NaCl	-	+	-	-	-	w	-	-
3% NaCl	-	+	-	-	-	-	-	-
4% NaCl	-	+	-	-	-	-	-	-

**Abbreviation and symbols: (-) negative; (^{T}) type strain; (w) weak growth; 13[DM2]: 0*i*,; 14:0*i*,; 14:0, tetradecanoic acid, 15:0*i*,; 15:0*a*,; 15:0,; 16:0*i*,; 15:0*i*,; 3-0H,; 17:0*i*,; 17:0*a*,; 17:0,; 16:0*i* 3-OH,; 16:0 3-OH,; 18:0*i*,; 17:0*i* 3-OH.; 17:0 3-OH; and 19:0.

[a]The method used did not allow the quantification of long-chain diols.

[b]Undetected fatty acids.

[c]Unknown fatty acid or alcohol with equivalent chain length (ECL) of 16.090

Table 1.2: Fatty acid composition of the strains of the species of the genus *Thermus* after growth at 70°C [a] (Taken from da Costa *et al.,* 2006)

Fatty acid	*T. aquaticus* (YT-1^T)	*T. thermophiles* (HB8^T)	*T. filiformis* (Wai33- A1^T)	*T. scotoductus* (ITI-252^T)	*T. brockianus* (YSO38^T)	*T. oshimai* (SPS-17^T)	*T. igniterrae* (RF-4^T)	*T. antranikianii* (HN3-7^T)
iso-C13:0	-[b]	-	-	-	0.7	0.7	1.1	-
iso-C14:0	1.0	0.7	0.9	-	1.3	-	-	-
C14:0	1.4	-	-	-	0.7	-	-	-
iso-C15:0	19.3	32.4	4.0	17.9	31.8	36.2	50.7	10.8
anteiso-C15:0	2.1	4.3	17.9	13.8	2.5	2.9	2.9	1.7
C15:0	-	-	-	1.0	0.8	3.1	1.3	1.9
iso-C16:0	13.4	5.3	8.7	1.6	11.0	2.5	1.0	9.6
C16:0	16.2	10.0	4.1	8.6	12.3	8.8	9.0	11.9
Un[c]	-	0.7	4.0	2.4	-	-	-	-
iso-C15:03-OH	3.2	-	-	-	-	-	-	-
iso-C17:0	24.9	41.4	6.3	30.3	35.2	38.3	31.1	51.0
anteiso-C17:0	2.6	5.1	35.5	22.1	2.8	3.2	1.9	6.2
C17:0	-	-	-	1.3	-	-	-	-
iso-C16:03-OH	2.6	-	0.9	-	-	-	-	-
C16:03-OH	2.3	-	-	-	-	-	-	-
iso-C18:0	0.6	-	1.0	-	0.6	-	-	1.4
iso-C17:03-OH	7.6	-	2.3	-	-	-	-	-
C17:03-OH	0.8	-	8.6	-	-	-	-	-
C19:0	-	-	1.0	0.6	-	-	-	-

**Abbreviation and symbols: (-) negative; (^{T}) type strain; (w) weak growth; 13[DM2]: 0*i*,; 14:0*i*,; 14:0, tetradecanoic acid, 15:0*i*,; 15:0*a*,; 15:0,; 16:0*i*,; 15:0*i*,; 3-0H,; 17:0*i*,; 17:0*a*,; 17:0,; 16:0*i* 3-OH,; 16:0 3-OH,; 18:0*i*,; 17:0*i* 3-OH.; 17:0 3-OH; and 19:0.

[a]The method used did not allow the quantification of long-chain diols.

[b]Undetected fatty acids.

[c]Unknown fatty acid or alcohol with equivalent chain length (ECL) of 16.090

2.2. Morphological characteristics

The strains of the genus *Thermus* are non-sporulating and form rods or slender bacillar-shaped cells which likely form septated filaments in exponential cultures on rich medium, and subsequently separate by binary fission when it reaches the stationary phase (Cava *et al.,* 2009, da Costa *et al.,* 2006, Kim *et al.,* 2006). Members of the genus *Thermus* are heterotrophic, thermophilic, Gram negative bacteria that grow aerobically with high growth rates and good yields on complex medium at an optimum growth temperature of 62°C to 75°C, and they do not require specific amino acids or vitamins (Zhang *et al.,* 2010).

When grown anaerobically some strains of the genus *Thermus* use nitrogen oxides or metals to grow through anaerobic respiration in which the nitrate will act as an electron acceptor (Pati *et al.,* 2011, da Costa *et al.,* 2006). Overall, most strains show orange– or yellow-pigmented colonies. This is because of the presence of relevant fraction of carotenoids in their membranes (Cava *et al.,* 2009). Strains that are isolated from man-made environments are black and tend to lack pigmentation, even though yellow-pigmented strains can also be isolated in small amounts from these environments (da Costa *et al.,* 2006). It was also observed that when strains of the genus *Thermus* are grown in rich media under low stirring conditions they tend to form fragile multicellular structures also known as “rotund bodies”. No motile forms of the genus *Thermus* have been identified and described (Cava *et al.,* 2009, Kim *et al.,* 2006).

Transmission electron microscopy shows that the strains of the genus *Thermus* have an envelope that appears as a complex pattern of four layers consisting of a cytoplasmic membrane with simple outline, a cell wall with an inner, electron-dense thin layer indicating the peptidoglycan connected to an outer corrugated layer by irregularly spaced invaginations (Figure 1.5, Cava *et al.,* 2009, da Costa *et al.,* 2006). The “rotund bodies” are occasionally seen in many strains of the genus *Thermus* by phase-contrast and electron microscopy. These unusual morphological structures consist of a number

of cells bound by a common external layer of the envelope enclosing a large space between the cells (Casto *et al.,* 1995, da Costa *et al.,* 2006).

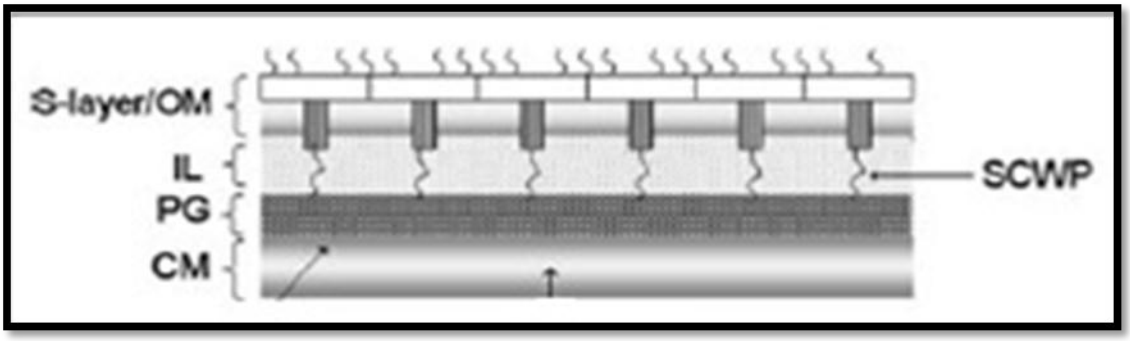


Figure 1.5: Scheme of a cell envelope structure in *T. thermophilus*, showing the different layers present in the cell envelope. CM cytoplasmic membrane, PG peptidoglycan, IL intermediate amorphous layer containing the secondary cell wall polymers (SCWP) covalently associated to the peptidoglycan and bound through the SLH domain to the S-layer; S-layer/ OM the S-layer outer-membrane complex (Taken from Cava *et al.,* 2009).

It is also reported that the murein composition and peptide cross-bridges of the genus *Thermus* are basically for a Gram positive bacterium, but due to the presence of murein content, as well as the degree of cross-linkages, the glycan chain length present features that are mostly observed in Gram negative organisms that have been studied this far, which then explains the Gram negative staining of the genus (Cava *et al.,* 2009).

2.3 Metabolism

A majority of thermophiles are well adapted in degrading various organic acids, proteinaceous substrates, as well as growing on a variety of simple and complex carbohydrates as sources of carbon and energy. It is known that most hyperthermophiles grow optimally on α and β linked carbohydrates present in plant materials such as starch, xylan, pectin, and cellulose or in animal cells or even in microorganisms (glycogen). These carbohydrates are preferably used as di- or

polysaccharides due to the thermostability of the monosaccharide species (Stetter, 1996 &1999; Cava *et al.,* 2009).

The substrates are made available by numerous thermozymes such as (exo) proteases, lipases, glucosidases, pollulanases and galactosidases which are known to be encoded in the genome of *T. thermophilus* (Cava *et al.,* 2009). However, to date none of the strains of the genus *Thermus* appears to have the ability to ferment. The strains of the species have a respiratory metabolism and most of the strains are strictly aerobic. Subsequently, some strains are capable of growing under anaerobic conditions using nitrate as the final electron acceptor while other strains also reduce nitrite (da Costa *et al.,* 2006).

2.3.1 Aerobic respiration

Despite that oxygen has a low solubility at high temperatures, most strains of the genus *Thermus* have been isolated aerobically, where oxygen serves as the only electron acceptor that supports growth in many strains. The respiratory enzyme of *Thermus* sp. have been the subject of several biochemical and structural studies, allowing its aerobic electron transport chain to be one of the best characterized (Figure 1.6, Cava *et al.,* 2009).

The reducing equivalents from the catabolic metabolism are fed into the electron transport chain (eTC) via the proton- translocating Type I NADH dehydrogenase (NDH-1) that is similar to the mitochondrial complex I. In addition, electrons either enter the eTC through succinate dehydrogenase or complex II, and putatively through a monomeric quinine-dependent type-II NADH. A nicotinamide nucleotide transhydrogenase can act as the respiratory chain-linked to the proton pump. From which the two terminal cytochrome oxidases described support the evidence that a periplasmic cytochrome also serves as a electron acceptor. In contrast, the other one could receive electrons directly from the recently described complex III (Cava *et al.,* 2009).

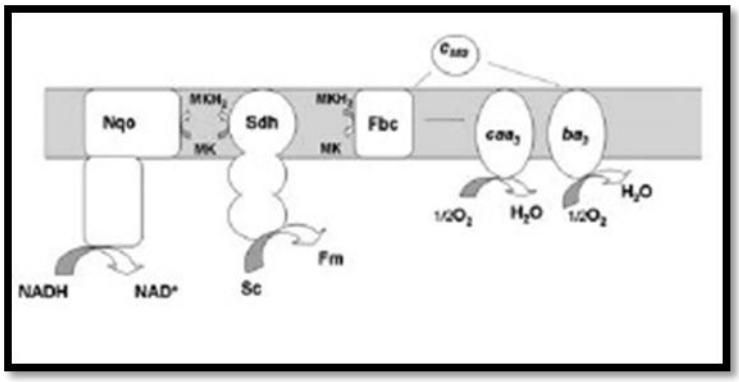


Figure 1.6: Aerobic respiration in *T. thermophilus*. Diagram of the aerobic eTC of *T. thermophilus* in which the respiratory complexes I (Nqo), II (Sdh) and III (Fbc) are indicated along the periplasmic cytochrome c552 and the final caa3 and ba3 oxidases. Menaquinone-8 (MK) is the main component of the quinine pool. Small arrows indicate the electrons pathways, (Taken from Cava *et al.,* 2009).

2.3.1.1 Carbon fixation

Carbon dioxide influences morphology and stimulates the growth of many microorganisms. The conversion of inorganic carbon (carbon dioxide) to organic compounds by living organisms is defined as carbon fixation. The utilization of carbon by heterotrophic bacteria (*Thermus* sp.) is attributed largely to oxidation of organic carbon coupled to respiration (Global Biogemichal Cycles, 2013).

Amongst the carbon fixation cycles, the Calvin cycle seems to be the most commonly known (Figure 1.7). It is a cycle with a series of biochemical reactions which carries out carbon fixation using the light-independent reactions. The enzymes found in the Calvin cycle are functionally equivalent to many enzymes used in other metabolic pathways of which the key enzyme in this cycle is called the RuBisCO (Campbell *et al*., 2006).

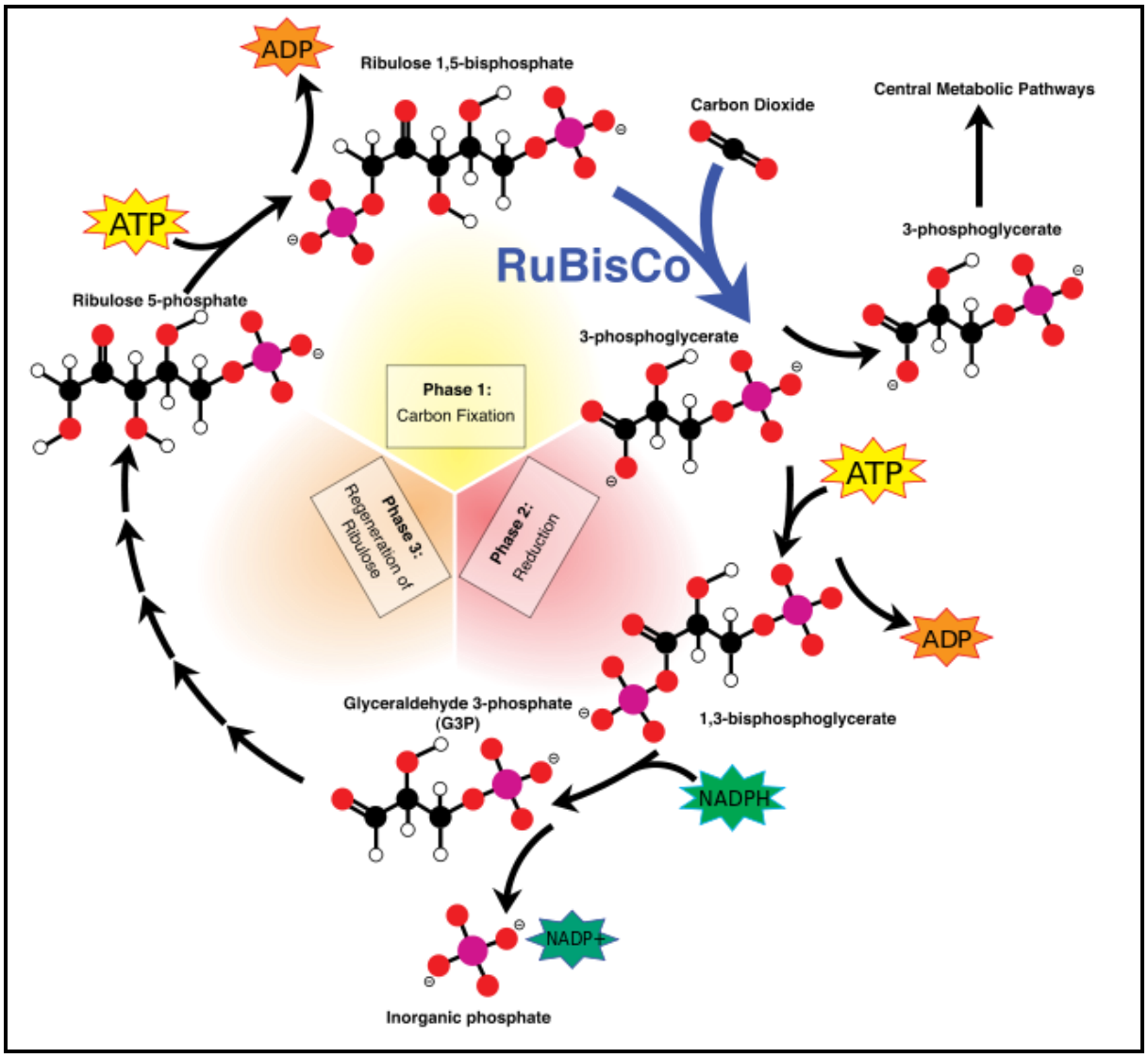


Figure 1.7: An overview of the Calvin cycle and carbon fixation. The RuBisCO enzyme catalyses the carboxylation of ribulose-1,5-bisphosphate a 5-carbon compound, by carbon dioxide (a total of 6 carbons) in a two-step reaction. The primary product of the reaction is a six-carbon intermediate it is so unstable that it immediately splits in half, forming two molecules of glycerate- 3-phosphate. The phosphoglycerate kinase catalyses the phosphorylation of glycerate-3-phosphate by ATP forming, 1,3-bisphosphoglycerate and ADP as products. The enzyme glycerate-3-phosphate dehydrogenase catalyses the reduction of 1,3-bisphosphoglycerate by NADPH. Glyceraldehyde 3-phosphate is produced, and the NADPH itself was oxidized and becomes $NADP^+$. Again, two NADPH are utilized per CO2 fixed (Taken from Campbell *et al*., 2006).

2.3.2 Anaerobic respiration

This process is also known as denitrification which occurs in several anaerobic and facultative anaerobic bacteria by the reduction of nitrate to nitrite, NO, N_2O and nitrogen (N_2) in four consecutive steps. Facultative anaerobic strains of *T. thermophilus* are able to grow anaerobically by partial or complete denitrification. Initially, nitrate is reduced to nitrite that accumulates in the medium as the final product. This process is then codified by a DNA fragment that could be transferred by conjugation to an aerobic strain, allowing the receptor to grow anaerobically (Cava *et al.,* 2009).

The DNA fragment is named nitrate respiratory conjugative element (NCE). It codes for the expected terminal nitrate reductase (Nar), the new type of respiratory NDH (Nrc) and for the regulatory elements involved in their expression. A cryptic replicative origin is located immediately downstream of the Nar operon, as well as the unidentified origin for conjugative transference seems to be downstream and not far from the Nar operon. This results into the NCE to constitute a self-mobilizable 'nitrate respiratory island' which is widely conserved in denitrifying strains of *T. thermophilus* isolated across the world (Figure 1.8, Cava *et al.,* 2009).

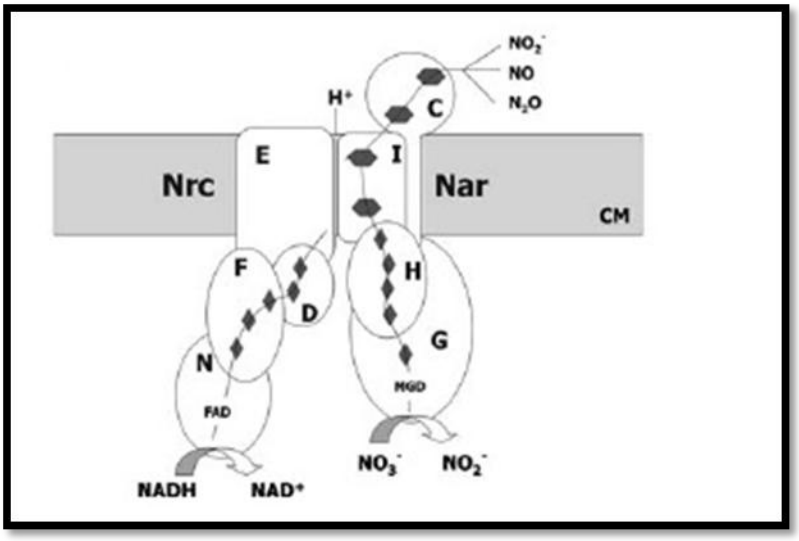


Figure 1.8: NarC and NarI constitute a respiratory super complex where NarC rules the deviation of the electrons towards the Nir, Nor and Nos reductases depending on the availability of nitrogen oxides. In the absence of nitrate, electrons will flow from the external heme b of NarI to the heme c groups of NarC to finally reach the denitrification reductases (Taken from Cava *et al.,* 2009).

2.3.2.1 Nitrogen metabolism

The presence of a nitrogen source within the growth media of anaerobic respiration is an intermediate in typical denitrification pathways of which dinitrogen gas (N_2) serves as the terminal product. Microorganism's possesing reduced denitrification pathways that are able to terminate with N_2O have been reported on their genomics that revealed phenotypes which results from mutations in the N_2O reductase (nos) genes or even in the complete absence of the nos genes. Geothermal environments has a wide variety of thermophiles that can respire nitrate or nitrite including members of several phyla of bacteria; such as Thermales consisting of some of the best-studied nitrate- and nitrite-reducing thermophiles (Figure 1.9-1.11, Moreno-vivia *et al.,* 2004; Bonete *et al.,* 2007; Hedlund *et al.,* 2011).

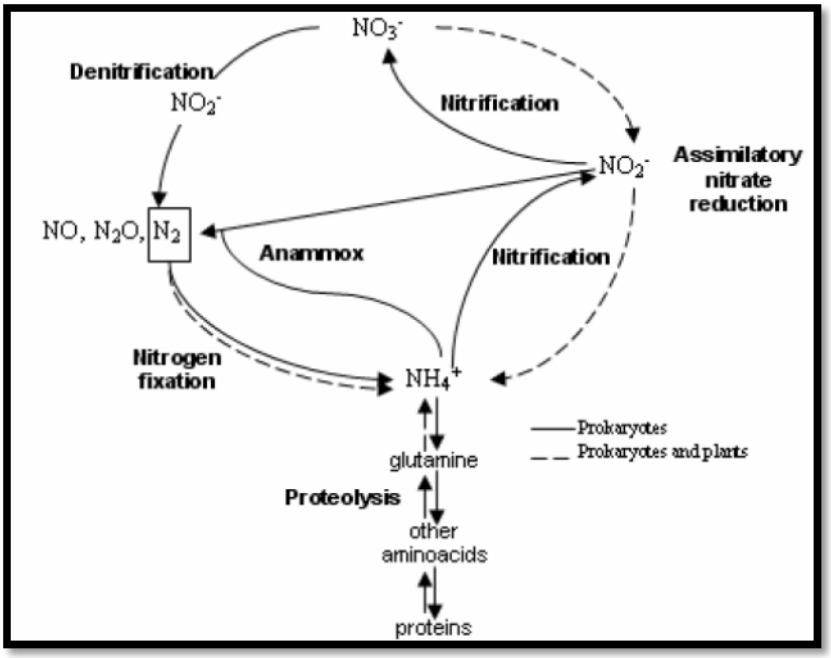


Figure 1.9: Schematic representation of a nitrogen cycle. NO_3^- is used as an electron acceptor for growth under anaerobic conditions, in which the assimilatory NO_3^- reduction plays an important role; while it acts as an electron acceptor. Denitrification functions in elimination of excess reductant power allowing nitrogen to be released as a gas (Taken from Moreno-vivia *et al.,* 2004).

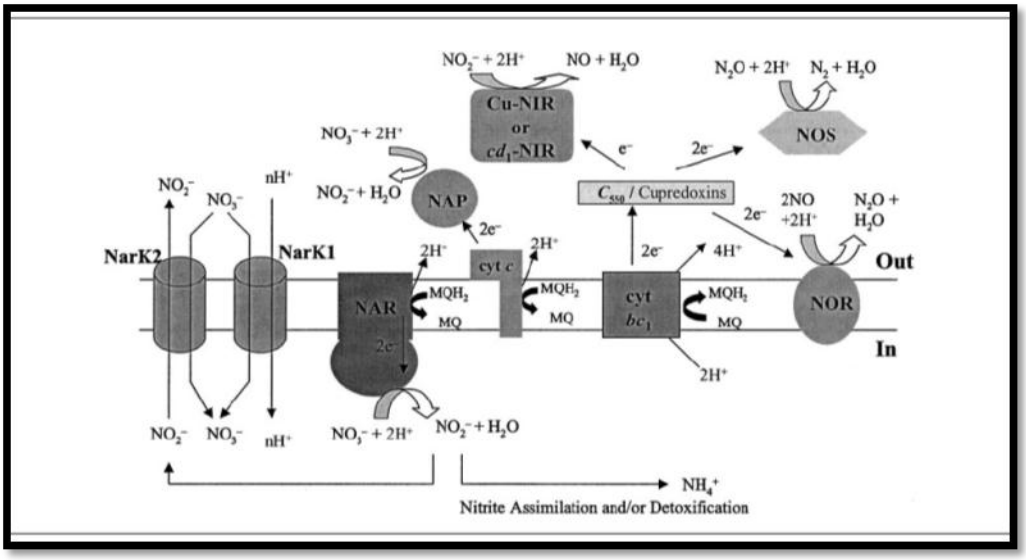


Figure 1.10: Schematic representation of the denitrification pathway in bacteria. The periplasmic nitrate reductase (NAP), the membrane-bound nitrate (NAR), periplasmic nitrite reductases (Cu-NIR or cd_1-NIR), membrane-bound nitric oxide reductase (NOR) and the periplasmic nitrous oxide reductase (NOS) are represented without indicating their subunit composition and cofactors (Taken from Moreno-vivia *et al.,* 2004).

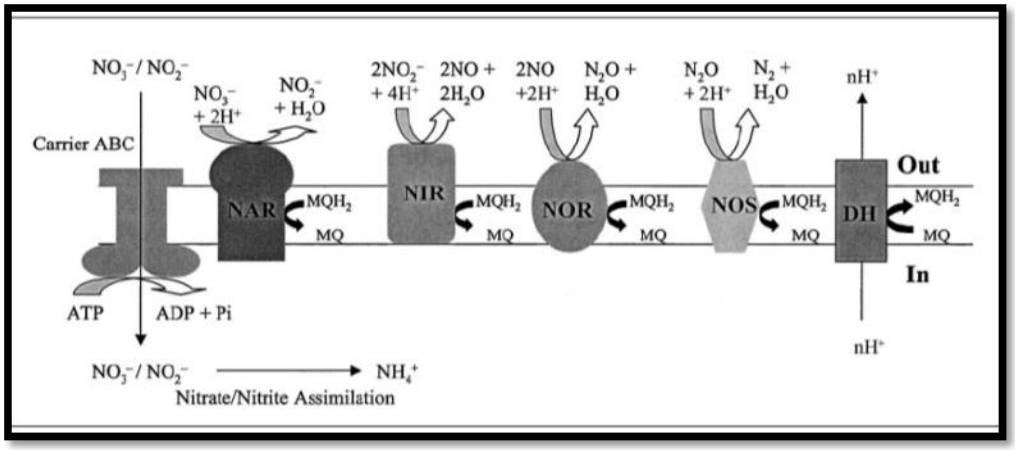


Figure 1.11: Schematic representation of the denitrification pathway in archaea. Location of membranes and dependence on menaquinol as the electron donor are assumed for the four reductases (nitrate reductase, NAR; nitrite reductase, NIR; nitric oxide reductase, NOR; as well as nitrious oxide reductase, NOS). The mentioned reductases are presented without indicating their subunit composition and cofactors. It is assumed that the putative ABC-type nitrate/nitrite transporter present in some archaea serves for assimilatory purposes instead of the respiratory nitrate reduction, this assumption could result from of the active site of NAR being located on the outside of the membrane (Taken from Moreno-vivia *et al.,* 2004).

3. Conclusions

Thermophiles spread to all domains of life, suggesting their broad nature across the entire ecological niche. This was then made possible by the presence of the adaptation of prokaryotes through horizontal gene transfer which dominated within the evolutionary distribution.

Moreover with all the limitations and advantages mentioned concerning the genus *Thermus*, broad applications are accompanied by this genus. Biotechnological applications are greatly influenced and this has a large contribution to the economy. Also, thermozymes are crucial even though only few are commercially available; their role is vast across a broad range of industries such as the application of PCR.

Studies done on the genus *Thermus* allowed great understanding on various aspects; which most importantly included the competency of the genes being able to play roles of carrying out certain functions better, in comparison to other related species. In addition, the metabolism of this genus is also noted to be an interesting subject to divert great attention to, since these species hardly require any external factors such as amino acids or even vitamins to completely carrying out their metabolic pathways.

It is also noted that the nitrate operon which allows denitrification respiration drew attention and was extensively looked into, in which some of the strains of the genus *Thermus* showed ability. Specific genes within this operon are identified to play crucial roles in enabling the reduction of nitrate, allowing the organism to enable growth in condition whereby there is no oxygen or there are low levels of it.

4. References

Bonete, M. J., Camacho, M., Martínez-Espinosa, R. M., Esclapez, J., Bautista, V., Pire, C., Zafrilla, B., Diaz, S., Pérez-Pomares, F. and Llorca, F. (2007) In the light of haloarchaea metabolism. *Appl. Microbiol*, 170-183.

Brown, J. R. (2003) Ancient Horizontal Gene Transfer. *Genetics*, Volume 4

Campbell, N. A.; Brad, W.; Robin J. H. (2006). Biology: Exploring Life *Pearson Prentice Hall.*

Casto, R., Lasa, I. G. and Biolog, C. D. (1995) Horizontal Transference of S-Layer Genes within *Thermus thermophilus*. *Microbiol*, **177**(19)**,** 5460-5466.

Cava, F., Hidalgo, A. and Berenguer, J. (2009) *Thermus thermophilus* as biological model. *Extremophiles*, **13,** 213-231.

César, C. E., Alvarez, L., Bricio, C., van Heerden, E., Litthauer, D. and Berenguer, J. (2011) Unconventional lateral gene transfer in extreme thermophilic bacteria. *Int. Microbiol,* **14,** 187-199.

Childers, S. E., Vargas, M. and Noll, K. M. (1992) Improved Methods for Cultivation of the Extremely Thermophilic Bacterium *Thermotoga neapolitana*. *Appl. and Enviro. Microbial*, **58**(12)**,** 3949-3953.

Chung, A. P., Rainey, F. A., Valente, M., Nobre, M. F. and da Costa, M. (2000) *Antranikianii* sp. nov., two new species from Iceland. *Inter. J. of Sys. and Enviro. Microbiol,* 209-217.

Coimbro, D., Tenreiro, S., Nobre, M. F., Hoste, B., Gillis, M., Kristjarisson, J. K. and da Costa, M. S. (1989) DNA:DNA hybridization and chemotaxonomic studies of *Thermus scotoductus. Res. Microbiol,* **146,** 315-324.

da Costa, M. S., Rainey, F. A. and Nobre, M. F. (2006) The Prokaryotes. 797-812.

Degryse, E., Glansdorff, N. and Pirard, A. (1978) Niu'nbinlng 9. *Design*, **196,**189-196.

Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., Reva, O., Kumwenda, B., Srivastava, M., Bricio, C., Berenguer, J., van Heerden, E. and

Litthauer, D. (2011) Sequence of the hyperplastic genome of the naturally competent *Thermus* SA-01. *BMC Genomics*, **12,** 577.

Griffiths, E. and Gupta, R. S. (2004) Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum. *J. of Bact,* **186(10),** 3097-3107.

Hedlund, B., McDonald, A. and Lam, J. (2011) Potential role of *T. thermophilus* and *T. oshimai* in high rates of nitrous oxide (N_2O) production in 80˚C hot springs in the US Great Basin. *Geobiol*, 471-480.

Hjorleifsdottir, S., Skirnisdottir, S. and Hreggridsson, G. (2001) Species Composition of Cultivated and Non-cultivated Bacteria from Short Filaments in an Icelandic Hot Spring 88˚C. *Microbial Eco*, **42**(2)**,** 117-125.

Jenny, F. E. and Adams, M. W. (2008) The impact of extremophiles on structural genomics (and vice versa). *Extremophiles: life under extreme conditions*, **12**(1), 39-50.

Jones, W. J., Nagle, D. P. and Whitman, W. B. (1987) Methanogens and the Diversity of *Archaebacteria*, **51**(1).

Kim, D., Park, B. H., Jung, B., Kim, M., Hong, S. and Lee, D. (2006) Identification and molecular modeling of a family 5 endocellulase from *Thermus caldophilus* GK24, a cellulolytic strain of *Thermus thermophilus. Int. J.of Mol.Sci*, **72,** 571-589.

Li, W. F., Zhou, X. X. and Lu, P. (2005) Structural features of thermozymes. *Biotechnology advances*, **23**(4)**,** 278-281.

Li, Y. and Fang, J. (2010). Biochemical and Biophysical Research Communications Distance-dependent statistical potentials for discriminating thermophilic and mesophilic proteins. *Biochem and Biophysi Research Communi*, **396**(3)**,** 736-741.

Mardis, E, R. (2008) The impact on next-generation sequencing technology on genetics. *Trends in Genetics*, vol. 24, 3.

Moreno-vivia, C., Rolda, M. D. and Moreno-vivia, C. (2004) Nitrate reduction and the nitrogen cycle in Archaea. *Microbiol,* 3527-3546.

Pati, A., Zhang, X. and Lapidus, A. (2011) *Oceanithermus profundus* type strain (506T). Sciences-New-York, 210-220.

Stetter, K. O. (1996) Hyperthermophilic procaryotes. *FEMS Microbiol Rev,* **18,** 149–158

Stetter, K. O. (1999) Extremophiles and their adaptation to hot environments. *FEBS Lett,* **452,** 22–25.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Toward a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya*. Proceedings of the National Academy of Sci USA,* **87,** 4576–4579.

Yusupov, M. M., Tischenko, S. V., Trakhanov, S. D., Ryazantsev, S. N. and Garber, M. B. (1998) A new crystalline form of 30 S ribosomal subunits from *Thermus thermophilus*. *FEBS Letters*, **238**(1)**,** 113-115.

Zhang, X., Ying, Y., Ye, Y, Xu, X., Zhu, X. and Wu, M. (2010) *Thermus arciformis* sp. nov., a thermophilic species from a geothermal area. *Inter. J. of Sys and Evol Microbiol,* **60,** 834-839.

CHAPTER 2

1. Introduction into study

Thermus scotoductus SA-01 was isolated by Kieft and co-workers (3.2 km) from the deep gold mine of South Africa in the year 1999 and characterized it as a facultative anaerobe capable of coupling the oxidation of organic substrates to reduction of a broad range of electron acceptors which included nitrate (Kieft *et al*., 1999; Balkwill *et al*., 2004). Moreover, *Thermus* sp. NMX2 A.1 strain was isolated from a thermal spring in New Mexico, USA (Williams and Sharp, 1995) and it had demonstrated similar metabolic activities to *T. scotoductus* SA-01 and phylogenetic identity was determined as a *T. scotoductus* (Balkwill *et al*., 2004).

Gounder and co-workers published a paper in BMC Genomics whereby they reported the complete genome sequence of *T. scotoductus* SA-01 and compared it to the complete genome sequences of *T. thermophilus* strains HB8 and HB27 (Gounder *et al*., 2011). In a nutshell what could be deduced is that *T. scotoductus* SA-01 differed a lot from the *T. thermophilus* strains such included the comparison of the Nar operon.

A statement from Bricio and co-worker supports the lack of similarities at the Nar operon amongst *T. scotoductus* SA-01, *T. thermophilus* strains HB8 and HB27; when they pointed out that both sequenced *T. thermophilus* strain HB8 and HB27 are not able to grow in the absence of oxygen, however, there were many isolates of *Thermus* sp. identified as *T. thermophilus* that had shown to grow anaerobically but not *T. thermophilus* strain HB8 and HB27 that is why they carried out a modification on them (Bricio *et al*., 2011).

The objectives of this study were to sequence and annotate the genome of *Thermus* sp. NMX2 A.1 strain and compare it to *T. scotoductus* SA-01 with a detail focus on the Nar operon genes. Carry out the experimental metabolic activity of the closely related strains using potassium nitrate as the final electron acceptor determining the metal-

reducing abilities; and to determine the phylogenetic positions by characterizing them on their 16S rRNA then finally extend the search for new *Thermus* sp. using samples collected from various deep mines of South Africa.

2. Reference

Balkwill, D. L., Keift, T. L., Tsukuda, T., Kostandarithes, H. M., Onstott, T. C., Macnaughton, S., Bownas, J. and Fredrickson, J. K. (2004) Identification of iron-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles*, **8,** 347-44.

Bricio, C., Alvarez, L., GÓmez, M. J. and Berenguer, J. (2011) Partial and complete denitrification in *Thermus thermophilus*: lessons from genome drafts. *Biochem. Soc. Trans,* **39,** 249-253.

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

Kieft, T. L., Fredrickson, J. K., Onstott, T. C, Gorby, Y. A., Kostandarithes, H. M. and Bailey, T. J. (1999) Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Appl Enviro Microbiol*, **65,** 1214-1221.

Williams, R. A. D., Smith, K. E., Welch, S. G., Micallef, J. and Sharp, R. J (1995) DNA relatedness of *Thermus* strains, description of *Thermus brockianus* sp. nov., and proposal to reestablish *Thermus thermophilus* (*Oshima* and *Imahori*). *Int. J. Sys. Bacteriol,* **46,** 495–499.

CHAPTER 3

Genome sequence and annotation

1. Introduction

Over the past seven years, major fundamental shifts has been made from using first-generation automated Sanger sequencing for analysis of genomes, to much more advanced and newer methods referred to as (NGS) next-generation sequencing (Metzker, 2010). One of the major advantages of the NGS is the ability to produce an enormous amount of data at a relatively low cost. Moreover, the variety within the features makes it possible to co-exist in multiple platforms within the market place and in the laboratory (Wiley *et al*., 2009).

Tremendous transformations of the NGS technologies have led to application in various fields and in a variety of contexts. Such include *de novo* whole genome sequencing, profiling mRNAs and other small and non-coding RNAs, re-sequencing of genomes for variations, assessing DNA binding proteins and chromatin structure and detecting methylation patterns (Zhou *et al*., 2010).

The high-throughput NGS technologies include a number of processes that are mostly divided into template preparation, sequencing and data analysis (Zhou *et al*., 2010). This unique combination differentiates one technology from another, and determines the type of data produced from each platform (Metzker, 2010).

With so many applications and instrument platforms available on the market (Table 3.1), Roche/454 GS-FLX, Illumina Solexa Genome Analyzer, Applied Biosystems SOLID™ System, Polonator G.007, Ion Torrent, Helicos Heliscope™ and Pacific Biosciences SMRT, questions do arise as to which platform is the best choice for a given biological experiment (Zhou *et al*., 2010, Metzker, 2010).

Table 3.1: New Generation Sequencing technology available in the market (Adapted from Metzker, 2010)

Platform	Year	Comments	Run time	Gb per run
Roche/454FLX	2004	*de novo* assembly, Read length: 400 bp	8 hrs	0.5
Illumina Solexa Genome Analyzer	2006	Read length: 75 bp Currently most widely used	4 days	18
Applied Biosystems SOLiD™ System	2007	Read length: 50 bp Inherent error correction	7 days	30
Polonator G.007	2008	Read length: 26 bp Least expensive	5 days	12
Ion Torrent	2010	Read length: 200 bp Result of combining basic chemistry & semiconductor-based detection	<2 hrs	1
Helicos Heliscope™	Recently available	Read length: 32 bp High error rates	8	37
Pacific Biosciences SMRT	2010	Read length: 964 bp Potential for reads > 1 kb	N/A	N/A

The market currently has three dominating commercial platforms being the Roche/454 GS-FLX Genome Sequencer system, the Illumina Solexa Genome Analyzer and the Applied Biosystems SOLID™ System (Zhou *et al*., 2010). The Roche/454 GS-FLX Genome Sequencer system pyrosequencing technology, which was developed by 454 Life Science is based on the sequencing-by-synthesis detection of an inorganic pyrophosphate (PPi) released during the enzymatic incorporation of a deoxynucleotide triphosphate at the 3' end of a DNA primer. It revolutionized DNA sequencing through the ability to generate large amounts of data quickly (Wiley *et al*., 2009, Roche Applied Sciences, 454 Life Sciences).

Genome sequence and annotation

The application of 454 Sequence™ technologies on the Genome Sequencer FLX™ System entails four primary steps to convert the DNA from a genome into sequence data, which are represented in (Figure 3.1).

In this technology, the DNA sample is first sheared into small pieces. Two short adaptors, labeled A-adaptor and B-adaptor are then ligated to the ends of the small pieces of DNA. These adaptors provide priming sites for amplification and sequencing, as well as a special key sequence. The B-adaptor in addition contains a 5'-biotin tag that enables the immobilization of library fragments onto streptavidin-coated magnetic beads (Zhou *et al*., 2010).

The double-stranded products bound to the beads are then denatured to release the complementary non-biotinylated strands containing both an A- adaptor and a B-adaptor sequence. These denatured strands form the single-stranded template DNA library. For subsequent DNA amplification, the Genome Sequencer FLX system employs emulsion-based clonal amplification, called emPCR (Dressman, 2003).

The single stranded DNA library gets immobilized by hybridization onto primer-coated capture beads. The process is optimized to produce beads where a single library fragment is bound to each bead. The bead-bound library is emulsified along with the amplification reagents in a water-in-oil mixture. Each bead with a single library fragment is captured within its own emulsion micro reactor, where the independent clonal amplification takes place. After amplification, the micro reactors are broken, releasing the DNA-positive beads for further enrichment. For sequencing purposes, the DNA beads are layered onto a PicoTiter Plate device, depositing the beads into the wells, followed by enzyme and packing beads (Wiley *et al*., 2009, Zhou *et al*., 2010, Metzker, 2010).

The enzyme beads contain sulfurylase and luciferase, which act as key components of the detection part of the sequencing reaction, while the packing beads ensure that the DNA beads remain positioned in the wells during that sequencing reaction. The fluidics sub-system delivers sequencing reagents that contain buffers and nucleotides by

flowing them across the wells of the plate. Nucleotides are flowed sequentially in a specific order over the PicoTiter Plate device. When a nucleotide is complementary to the next base of the template strand, it is incorporated into the growing DNA strand by the polymerase.

The incorporation of a nucleotide releases a pyrophosphate moiety. The sulfurylase enzyme converts the pyrophosphate molecule into ATP using adenosine phosphosulfate. The hydrolyzed ATP results from the luciferase enzyme which uses luciferin to produce oxyluciferin and give off light. The light emission is detected by a CCD camera, which is coupled to the PicoTiter Plate device. The intensity of light from a particular well indicates the incorporation of nucleotides. Across multiple cycles, the pattern of detected incorporation events reveals the sequence of templates represented by individual beads, such results to the occurrence of the raw reads processed by the 454 platform (Zhou *et al*., 2010, Metzker, 2010, Wiley *et al*., 2009).

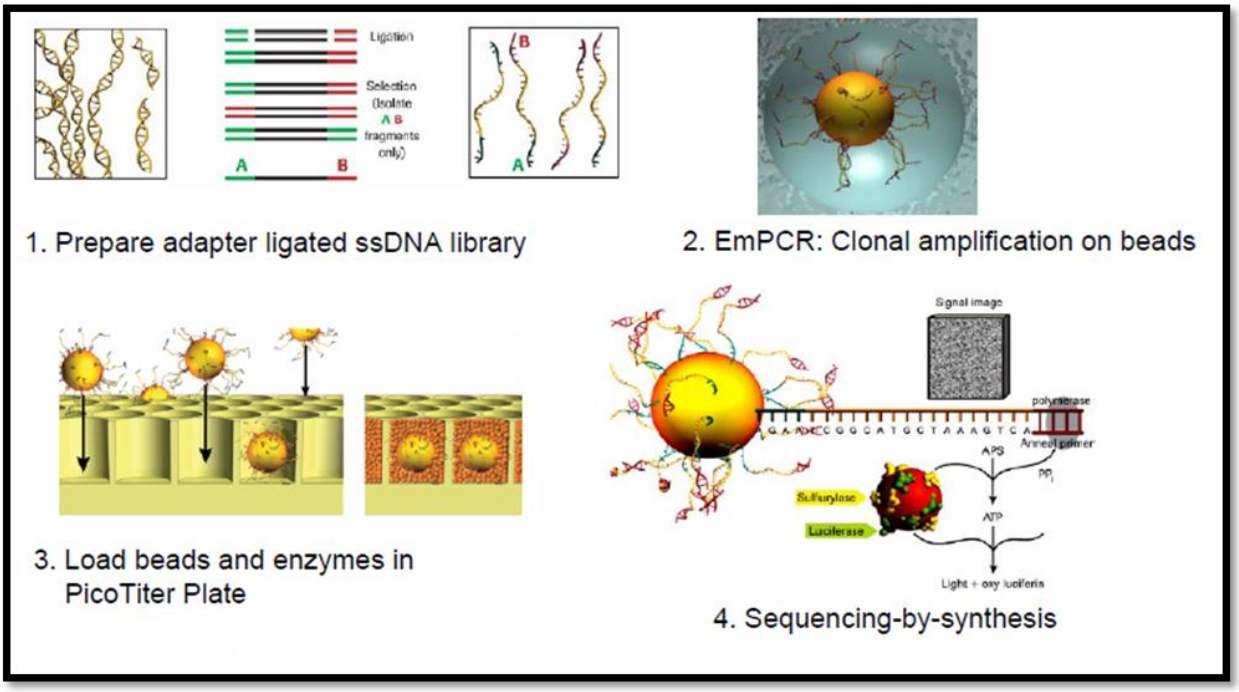


Figure 3.1: The complete sequencing workflow of the Genome Sequencer FLX System results from four primary steps when carrying out the conversion of genomic data into sequence data (Adapted from 454 Life Sciences). **1**. The generation of the template DNA library by shearing the DNA into small pieces; **2**. Emulsion-based clonal amplification of the resultant library; **3**. The beads are captured in Pico Titer wells on a fabricated substrate and then; **4**. Pyrosequenced where the single stranded PCR amplicon is hybridized to a sequencing primer and incubated with enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase as well as the substrates adenosine 5' phosphosulfate (APS) and luciferin (Taken from Wiley *et al*., 2009).

Genome sequence and annotation

1.1 The aim of the chapter

The aim of this chapter was to sequence and assemble the genome of *Thermus* sp. NMX2 A.1 strain using the Roche/454 GS-FLX Genome Sequencer system, and to use automated and curated annotation to assign functionality.

2. Materials and Methods

2.1 Strain used and growth incubations

The strain used was *Thermus* sp. NMX2 A.1 in ATCC medium 697 provided by Prof. Kieft at (New Mexico Tech).

Biomass for DNA isolation was obtained by aerobic culture in ATTC medium 697 [4 g/L yeast extract, 8 g/L peptone and 2 g/L NaCl; pH 7.5] at 68°C.

2.2 Genomic DNA extraction

Genomic DNA extraction was performed according to the method described by Labuschagne and Albertyn (2007). Briefly, 2 mL liquid culture was grown to an OD_{600nm} of 0.8, centrifuged for 1 min at 20 817 x *g*, using the FA45-30-11 rotor and the supernatant was removed by aspiration. The pellet was suspended in DNA isolation buffer (500 µL) [100 mM Tris-HCl (pH 8), 50 mM EDTA, 1% (w/v) SDS] followed by the addition of 200 µL glass beads, vortexing for 4 min and immediately placed on ice. Ammonium acetate (7 M, 275 µL, pH 7) solution was added to the suspension which was vortexed to homogenize the solution, and incubated for 5 min at 65°C followed by 5 min on ice. Chloroform (500 µL) was added; the mixture was vortexed for 30 seconds and centrifuged (14 000 x *g,* 5 min, 4°C).

The DNA was precipitated by adding 1 volume of isopropanol (approximately 750 µL) to the aqueous phase and centrifuged (14 000 x *g,* 5 min, 4°C) after 10 min at -20°C. Seventy percent (v/v) ethanol was used to wash the DNA pellet to remove salts followed by centrifugation (14 000 x *g,* 5 min 4°C) to collect the DNA. Subsequently, the pellet was dried under vacuum (Eppendorf concentrator 530), dissolved in (100 µL) distilled water (heated at 55°C for 10 min), vortexed and 1 µL RNase A (10 µg/mL, Fermentas) added to remove any RNA contamination

The quantity and purity of the genomic DNA extract was analyzed by using a NanoDrop spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE). The quality

and integrity of the genomic DNA was also evaluated by agarose gel electrophoresis and visualized on a 0.8% (w/v) agarose gel containing 2.5 mg/µL ethidium bromide. The agarose gel was prepared in TAE buffer containing 100 mM Tris-HCl (pH 8), 0.05 M ethylenediaminetetra acetic acid (EDTA) and 100 mM glacial acetic acid. Electrophoresis was performed at 6 V/cm for 30 min in TAE buffer and the genomic DNA was visualized using a ChemDoc XRS (Biorad Laboratories) gel documentation system.

2.3 PCR amplification of the 16S rRNA gene

The identity of the *Thermus* strain was confirmed by amplification and sequencing of the 16S rRNA gene from the genomic DNA using the universal bacterial primers 27F and 1492R (Table 3.2).

Table 3.2: Primer sequences for bacteria 16S rRNA gene amplification

Primers	Sequence	Reference
27F	5'- AGA GTT TGA TCM TGG CTC AG-3'	Lane, 1991
1492R	5'- GGT TAC CTT GTT ACG ACT T-3'	Lane, 1991

For bacterial 16S rRNA gene amplification the PCR reaction mixture contained 0.2 µM of each primer (Table 2.1), 1x buffer, dNTPs (0.2 µM), template DNA (200 ng) and 1.25 U of *Taq* DNA polymerase (New England BioLabs) in a final volume of 50 µL. Amplifications were run for 30 cycles in a thermal cycler pXe 0.2 Thermal Cycler (Thermo Electron) after an initial denaturation at 95ºC for 3 min. Each cycle was run at 94ºC for 1 min, 47.1ºC for 2 min and 70ºC for 2 min, and the final extension, 72ºC for 10 min. The PCR product (5 µL) was evaluated on a 0.8% (w/v) agarose gel and recovered by excising from an agarose gel with a sterile blade and extracted using the Biospin Gel Extraction kit (BioFlux) following the manufacturer's instructions and eluted with sterile double distilled water (50 µL).

2.4 Cloning of the 16S rRNA gene into pGEM-T Easy vector and sequencing

The purified PCR product was ligated into pGEM-T Easy vector (Figure 3.2) using 6 µL of the purified PCR product, pGEM-T Easy vector (50 ng), T4 DNA ligase (1 U, Fermentas), 1x T4 ligase buffer in a final volume of 20 µL, at room temperature for 1 h and further incubated at 4ºC overnight.

Five µL of the ligated product was transformed into Top 10 *Escherichia coli* competent cells (Lucigen). The transformation mix was placed on ice for 30 min, heat shocked for 40 sec by incubation at 42ºC and immediately placed on ice. Subsequently, 500 µL SOC media [LB medium containing 0.02 M Mg^{2+} and 0.02 M glucose] was added to the test tube containing the transformation mix and incubated at 37ºC for 1 h. Positive clones were identified through selection on LB-plates containing 30 µg/mL kanamycin followed by small scale plasmid isolation performed using the Plasmid DNA Extraction Kit (Biospin) according to the instructions of the manufacturer. The digested vector was evaluated as described in (section 2.2).

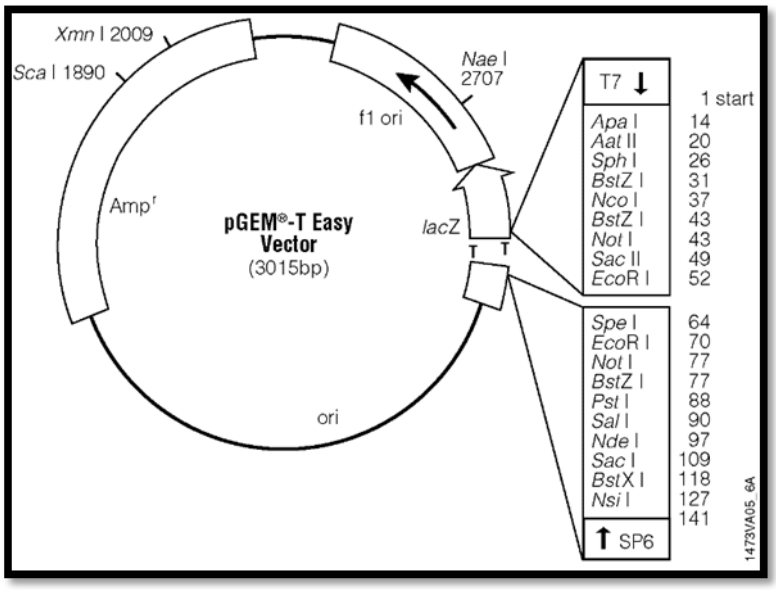


Figure 3.2: pGEM®-T Easy Vector System (Promega).

Genome sequence and annotation

A number of positive clones were selected for DNA sequencing to verify the identity of the bacterial strain and to check for possible contamination using the SP6 and T7 primers (Table 3.3).

Table 3.3: Primer sequences for pGEM-T Easy vector insert sequencing

Primers	Sequence
T7 Promoter	5'-TAATACGACTCACTATAGGG-3'
SP6 Promoter	5'-TACGATTTAGGTGACACTATAG-3'

The ABI Prism® BigDye v.3.1 Cycle (AB Applied Biosystem HITACHI, 3130x Genetic Analyzer) sequencing kit was used according to manufacturer's instructions, the composition for the sequencing reaction consisted of 5x dilution Buffer (2 µL), premix (0.5 µL), sequencing primer [3.2 *p*mol] (1 µL) and a DNA template (~ 200 ng) (6.5 µL).

The PCR products were purified using the EDTA/ethanol precipitation post-reaction clean up. The sequencing reaction volume was adjusted to 20 µL and transferred to a 1.5 mL eppendorf tube containing 125 mM EDTA (5 µL) and absolute ethanol (60 µL), vortexed for 5 sec and precipitated at room temperature for 15 min and centrifuged (20 000 x *g*, 10 min, 4°C). Complete aspiration of the supernatant was done and the pellet washed using 70% (v/v) ethanol (60 µL) and centrifugation (20 000 x *g*, 5 min, 4°C). Another complete aspiration of the supernatant was done, the remaining product was dried under vacuum (Eppendorf concentrator 530) for 5 min and sequenced with the ABI 377 Prism® BigDye v.3.1 Cycle (AB Applied Biosystem HITACHI, 3130x Genetic Analyzer) at the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State.

Searches of homologous sequences against the GenBank nr database were performed using the (Basic Local Alignment Search Tool) nBLAST alogarithm at NCBI (Altschul *et al.,* 1990), shown in Supplement A. Followed by multiple alignments of the sequences from both the forward (SP6) and reverse (T7) primers of the positive selected clones.

2.5 High throughput sequencing

After verifying the identity of the *Thermus* sp. NMX2 A.1 strain preparation of the DNA samples for high-throughput 454-pyrosequencing (GS20/FLX) were carried out. Two samples were sent for sequencing, one of the samples was re-suspended in 100 µL dH_2O and the other was dried under vacuum (Eppendorf concentrator 530).

Genomic DNA of *Thermus* sp. NMX2 A.1 strain was sequenced by GATC Biotech using a ½ FLX plate even though a good library was obtained, read lengths of only 233,4 base pairs and an assembly of only 60-fold coverage was achieved. A second run using a ¼ FLX plate was applied for improved read lengths, providing an approximate coverage of 87-fold with read lengths of 382 base pairs.

2.6 Retrieval of sequence raw data

The sff files (two data sets) were retrieved at (http://www.gatc-biotech.com/en/), for *de novo* assembly using the Roche 454 Newbler software. The data sets were assembled individually and combined. After assembly, the ace file was subjected to the conversion pipeline described at (http://genome.imb-jena.de/software/roche454ace2caf/). The resultant gap file could be edited using Gap v4.11.2.

2.7 Chromosomal alignment

Contigs obtained from the *de novo* assembly were aligned and ordered using the *T. scotoductus* SA-01 chromosome as template, with the application of MAUVE. When comparing genomes, the system used must account for all evolutionary phenomena to provide a complete outcome of genetic differences (Darling *et al*., 2004).

The advantages of the MAUVE method are that it identifies conserved genomic regions, rearrangements and inversions in the conserved regions, and the exact sequence breakpoints of such arrangements across multiple genomes. Also, this comparison method performs traditional multiple alignment of conserved regions to identify

nucleotide substitutions and small insertions and deletions (collectively known as indels; Darling *et al*., 2004).

2.8 Annotation

Two annotation pipelines available online were used. Contigs from the whole genome assembly were arranged from largest to smallest using Gap4. Contigs were then joined using a linker sequence (NNNNNCACACACTTAATTAATTAAGTGTGTGNNNNN). This spacer sequence introduces translational stop codons in all six frames; it gets placed between the contigs to form a pseudochromosome. The linker avoids the prediction of genes across the contig limits.

The pseudochromosome in FASTA format was then submitted to the Institute for Genome Research annotation pipeline at the University of Maryland (http://ae.igs.umaryland.edu/cgi/index.cgi) for *ab initio* gene finding GLIMMER (Gene Locator and Interpolated Markov ModelER) as well as Hidden Markov Models and BLAST similarity searches (Kosuge *et al.,* 2006; Ghahramani, 2001; Altschul *et al.,* 1990). The results of the annotation are provided as a MYSQL database which can be visualized using Manatee and internet browser. The annotation data is password protected and is available at (http://manatee.igs.umaryland.edu/tigr-scripts/chadoprokmanatee/shared/login.cgi).

For the second annotation run, contigs of the *Thermus* sp. NMX2 A.1 strain were aligned against the chromosome of *T. scotoductus* SA-01 using MAUVE, and then further submitted to the Joint Genome Institute of the US Department of Energy (http://img.jgi.doe.gov/cgi-bin/submit/main.cgi) Microbial Annotation pipeline.

The DOE-JGI Microbial Annotation Pipeline (DOE-JGI MAP) is an automated pipeline for the annotation of bacterial and archaeal genomes. Annotation includes both the identification of protein-coding and non-coding genes and repeats, as well as the prediction of the function of each gene and the assignment of a product name. The output of this pipeline is available through the Integrated Microbial Genomes/Expect Review (IMG/ER) system. This data warehouse integrates genomes and metagenome

datasets provided by the users of the system with a broad set of publicly available bacterial, eukaryotic and phage genomes; including engineered, environmental and host associated metagenome samples. Genes available in the IMG/ER system are characterized using several functional resources such as Clusters of Orthologous Groups of proteins (COG's), Kyto Encyclopedia of Genes and Genomes (KEGG), Eukaryotic Orthologous Groups of proteins (KOG's), PFAM, TIGRfam, MetaCyc, Interpro and Gene Ontology (https://img.jgi.doe.gov/).

2.9 Bi-directional BLAST

Bi-directional BLAST is a conventional procedure for the first approach to homolog detection. It detects, annotates, and analyzes candidate orthologous or paralogous sequences in a single go; and lately it is supplemented with domain architecture information, and should also corroborate by synteny analysis (de Almeida, 2010).

Organisms chosen for BiDi Blast by Antje Wollherr- personal communication, (Gounder *et al*., 2011) with *Thermus* sp. NMX2 A.1 strain were the related strains *T. scotoductus* SA-01, *T. thermophilus* HB27, *T. thermophilus* HB8, *T*. *thermophilus* JL-18, *Thermus* sp, RL, *Thermus* sp. CCB US3 UF1 and *T. thermophilus* SGO-5JP17. BiDi Blast results were visualized with MS Excel and a colour-key chosen for mapping the different levels of similarities.

3. Results and discussion

3.1 Verification of *Thermus* sp. NMX2 A.1 strain

Genomic DNA was successfully isolated from the *Thermus* sp. NMX2 A.1 strain as described (section 2.2). The obtained bands showed high integrity DNA and almost no shearing, indicated in (Figure 3.3). The A_{260}/A_{280} absorbance ratio of 1.8 indicated low protein or RNA contamination.

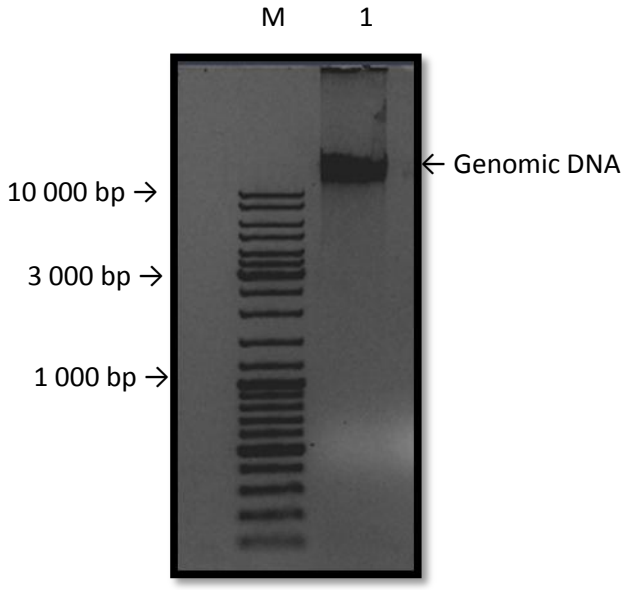


Figure 3.3: Extracted genomic DNA from *Thermus* sp.NMX2 A.1 strain visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; GeneRuler™ DNA ladder (Fermentas), lane 1; isolated genomic DNA >10kbp.

The amplification and cloning of the 16S rRNA gene was performed as described (section 2.3) and a DNA fragment of approximately 1 500 bp was obtained as expected for bacterial amplicons (Figure 3.4). The 16S rRNA gene present in the genome of the *Thermus* sp. NMX2 A.1 strain was subjected to BLAST analysis against the NCBI database to find the closest homologues and to confirm the identity of the organism. The closest homologue was that of the 16S rRNA gene of *Thermus* sp. NMX2 A.1 strain

(>gb|L09661.1) with a 99% coverage showing a 99% identity with an E-value of zero (Supplement A).

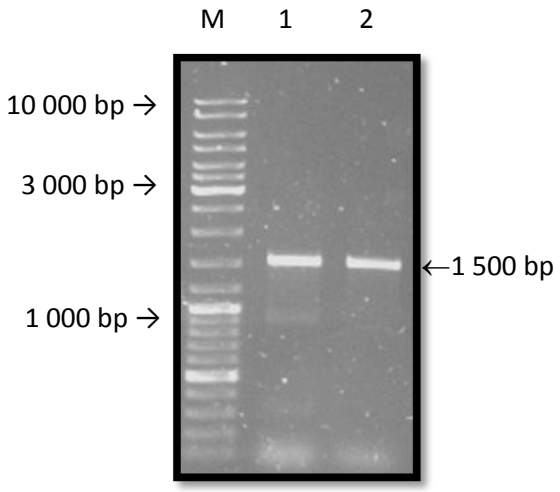


Figure 3.4: Amplification of 16S rRNA fragment from genomic DNA on an ethidium bromide-stained agarose gel 0.8% (w/v). Lane M; GeneRuler™ DNA ladder (Fermentas), lane 1 & 2; positive amplified band of 16S rRNA from the *Thermus* sp. NMX2 A.1 strain.

3.2 Assembly of sequencing data

The results for the *de novo* genome assembly are shown in (Table 3.4). The initial sequencing run performed by GATC gave shorter than expected read lengths due to a technical problem; an additional quarter plate run was performed. As the read lengths were not optimal, it was decided to run the assembly on the individual as well as the combined data sets.

The combined data produced the largest average contig size so it was decided to use this assembly. Experience with previous data sets had shown that this assembly would in all likelihood be the least error prone. The UFS assembly was very similar to that provided by GATC Biotech. The approximate genome size is 2,293 Mbp which is close to the genome size of *T. scotoductus* SA-01 consisting of 2,346 Mbp.

Table 3.4: Analysis of sequence data on quality aspects while comparing two sets of runs assembled with Newbler assembly software

		UFS Assembly Run 1	UFS Assembly Run 2	UFS Combined Assembly Run
Run Metrics				
	Total Number of Reads	618637	158939	777576
	Total Number of Bases	141066691	60295300	201361991
	Number Aligned Reads	611000, 98.77%	153987, 96.88%	765159, 98.40%
	Number Aligned Bases	139074590, 98.59%	59317381, 98.38%	198131940, 98.40%
	Inferred Read Error			0.91%, 1799320
	Number assembled	592514	147809	739131
	Number Partial			10730
	Number Singleton			2684
	Number Repeat			14053
Large Contig Metrics				
	Number of Contigs	135	141	133
	Number of Bases	2287344	2292662	2293277
	Average Contig Size	16943	16260	17242
	N50 Contig Size	43146	38422	43540
	Largest Contig Size	194322	157163	162095

3.2.1 Additional comparative genome features

A comparison is shown in (Table 3.5) amongst three fully sequenced and published *Thermus* strains namely *T. thermophilus* HB8 and HB27, *T. scotoductus* SA-01 to the newly sequenced strain *Thermus* sp. NMX2 A.1 strain. The estimated genome size is very similar to that of *T. scotoductus* SA-01.

Unfortunately, the available data does not allow the identification of a plasmid or mega plasmid in the *Thermus* sp. NMX2 A.1 strain. In addition, the hypothetical genes of the *Thermus* sp. NMX2 A.1 strain and other *Thermus* strains differ a lot; this occurred due to the focus of this experiment at large being on specific genes that carry out denitrification which will be discussed in the next chapter. The curating search and annotation was therefore restricted and the focus was extensively shifted to the genes of interest.

Table 3.5: Comparing the genome features of three *Thermus* strains (*T. thermophilus* HB8 and *T. thermophilus* HB27, *T. scotoductus* SA-01) to the newly sequenced *Thermus* sp. NMX2 A.1 strain (Adapted from Gounder *et al*., 2011)

Features	**HB27**	**HB8**	**SA-01**	**TNMX2 A.1**
Size in base pairs	1,894,877	1,849,742	2,346,803	-
Plasmid size in base pairs	232,605	256,992 + 9322	8,383	-
Total genome size	2,127,482	2,116,056	2,346,803	2,293,277 (estimated)
Hypothetical genes	64 (22)	107 (30,7)	168 (1)	565
Conserved hypothetical genes	240 (39)	241 (36.2)	367 (3)	409
Number of protein coding genes	1,982 (228)	1,970 (251, 14)	2,506 (12)	2, 468

3.3 Chromosomal alignment using the MAUVE method

A pairwise chromosomal alignment was done using the MAUVE method which is known to represent the first alignment system that integrates the analysis of large-scale evolutionary events with traditional multiple sequences (Darling *et al*., 2004, Figure 3.5).

The MAUVE alignment was also used to order the contigs of the *Thermus* sp. NMX2 A.1 strain using the chromosome of *T. scotoductus* SA-01 as the template. The observed outcomes indicate that there are many similarities between *T. scotoductus* SA-01 (upper section) and *Thermus* sp. NMX2 A.1 strain (lower section). The boundaries of the different coloured blocks indicate break points in the chromosomes while the vertical red lines indicate contig boundaries.

A high degree of synteny is observed and the sequence similarities are high. In addition, the red blocks above the depiction of the *T. scotoductus* SA-01 chromosome indicates gene islands which were acquired by horizontal gene transfer (Gounder *et al*., 2011) but are absent in *Thermus* sp. NMX2 A.1 strain contigs.

Finally, *Thermus* sp NMX2 A.1 strain also has genes situated throughout the entire assembled contigs (toward the extreme right of the *Thermus* sp. NMX2 A.1 strain depiction) which are not present within *T. scotoductus* SA-01 chromosome. These may be islands which were acquired by *Thermus* sp. NMX2 A.1 strain after separation of the two lines from the common ancestor; this will be discussed in more detail in (section 3.5) using the BiDi BLAST results to support this data.

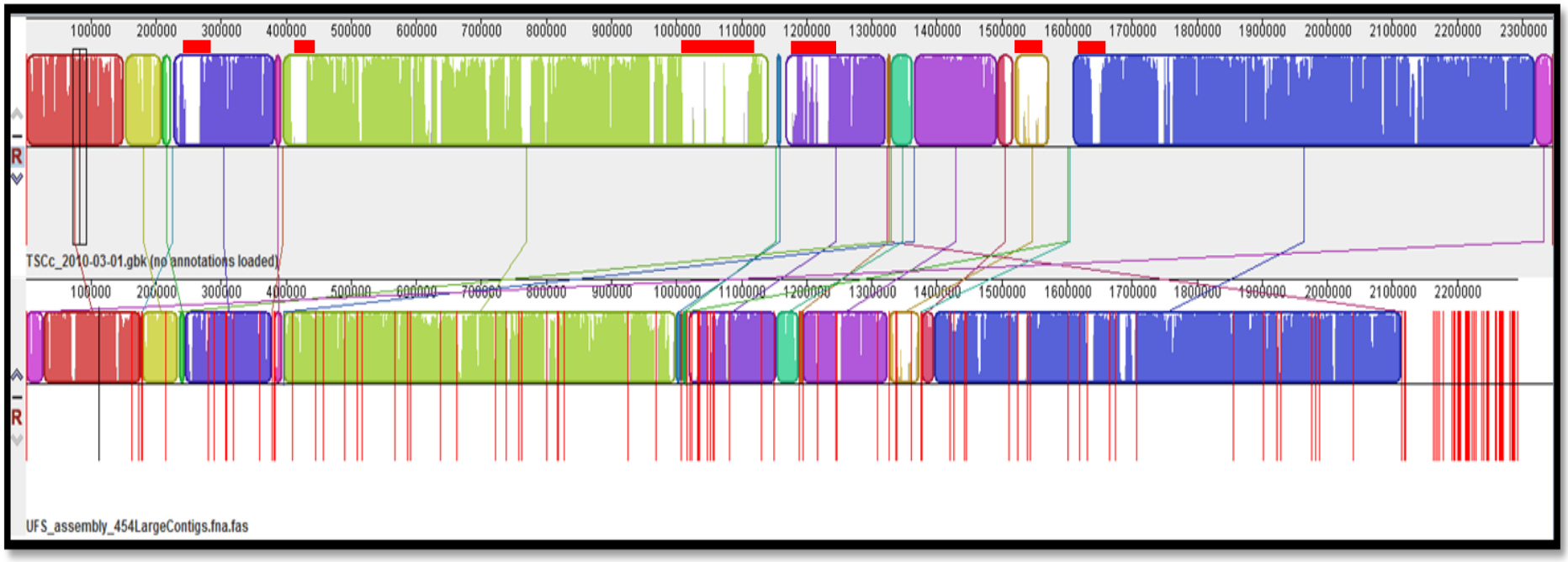


Figure 3.5: A schematic indication of a pair wise chromosomal alignment between two closely related *Thermus* strains. Using the *T. scotoductus* SA-01 chromosome and the contigs of *Thermus* sp. NMX2 A.1 strain.

3.4 Annotations

3.4.1 Annotation using the JCVI annotation at the University of Maryland

The results of the annotation were downloaded as a MYSQL database and are viewed in Manatee, which is a browser based program. A summary of the different role categories of proteins identified by the JCVI annotation pipeline are given in (Figure 3.6 a). A total number of 2 721 protein coding genes (ORF's) where the gene distribution within the different role categories are done according to the function they are assigned to by the JCVI pipeline.

The ORFs were submitted to a BLAST search against the non-redundant (nr) protein database. In addition, the BLAST analysis was also performed using the same pipeline. In Figure 3.6 (b) the observed results are shown in which the amino acid sequences of the ORF's were subjected to the pBLAST against the non-redundant protein database. As observed in the chart below, the majority of the ORFs (72.6%) had *T. scotoductus* SA-01 proteins as closest homologue. Apart from the 8.3% of ORFs which had no homologues in the database, almost all the remaining ORFs hit with proteins from the *Deinococcus-Thermus* phylum. The ORFs could be genes obtained by horizontal gene transfer but this is pure speculation at this stage.

A close analysis of the annotation data showed that the JCVI pipeline had misidentified a large number of ORFs possibly due to frame shifts as a result of sequencing or assembly errors, leading to an overestimation of the number of ORFs. It was, therefore, decided to submit the contigs to the Joint Genome Institute annotation pipeline.

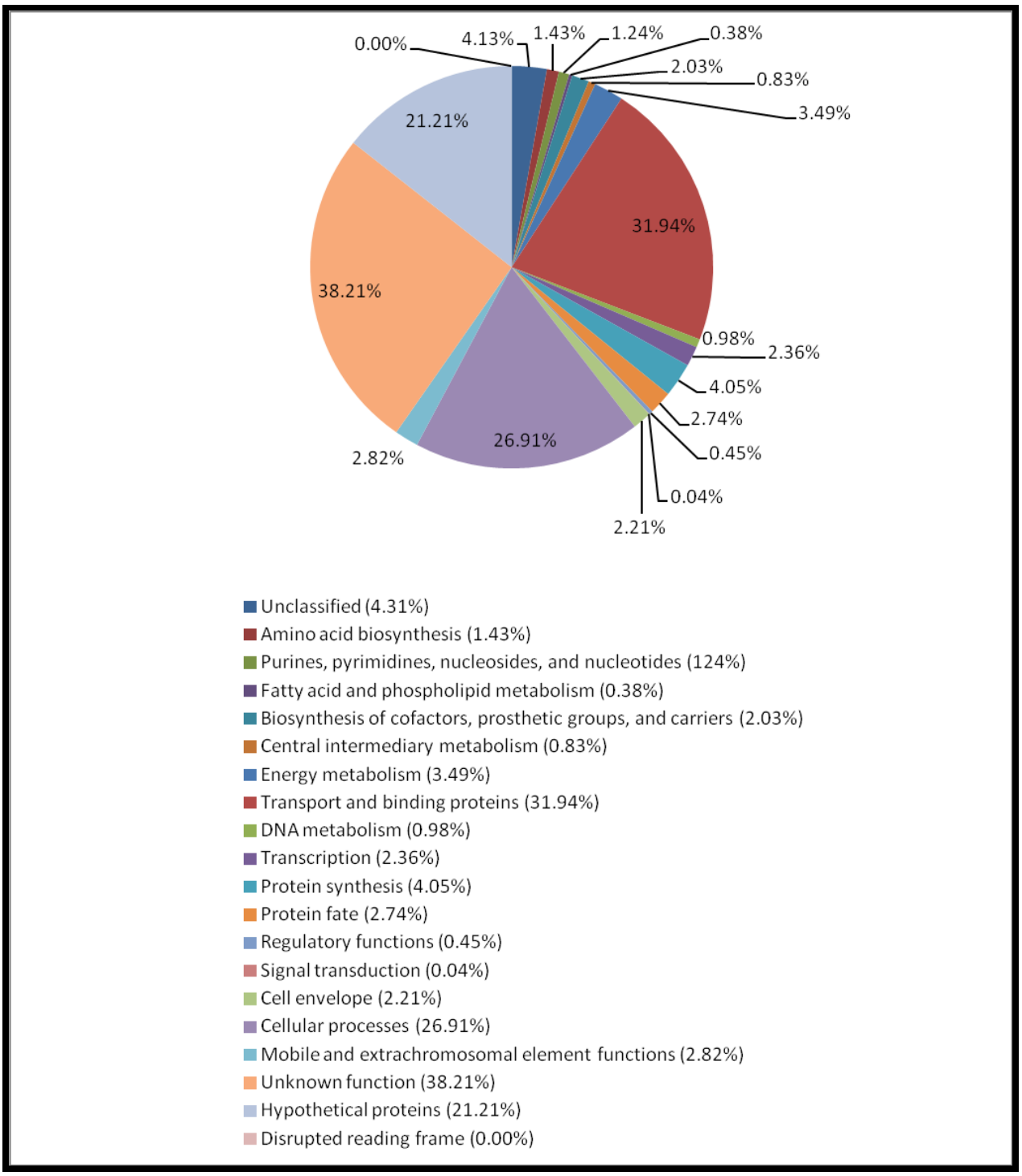


Figure 3.6 (a): Percentage distribution of genes in different role categories identified by the TIGR annotation.

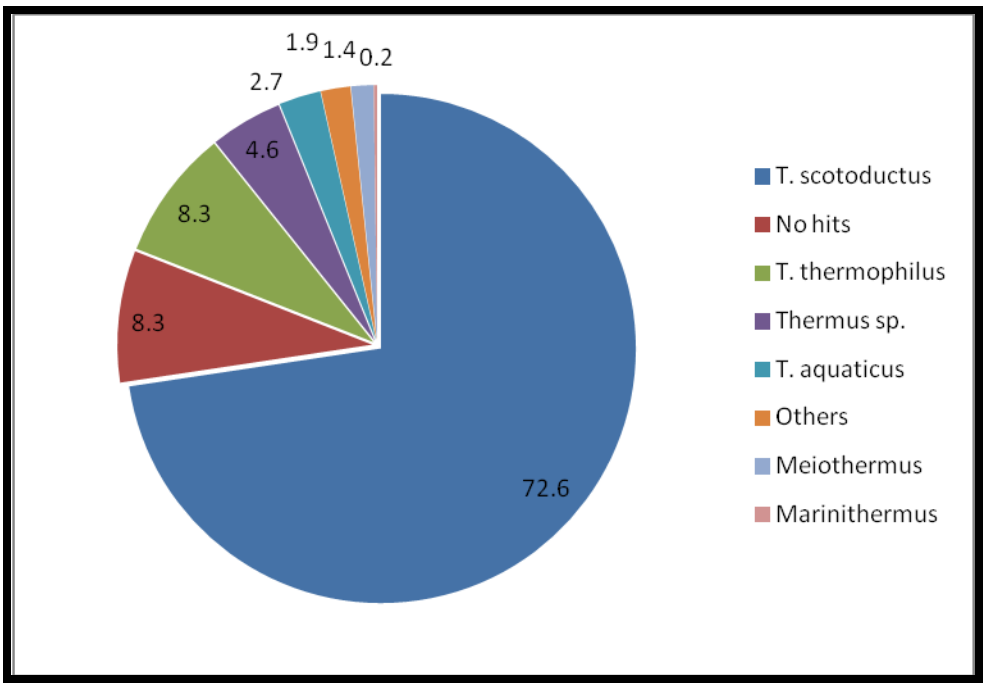


Figure 3.6 (b): Percentage distribution of species to which the amino acid sequences from the ORFs hit against when BLAST annotation was done using a pBLAST on a local nr BLAST server.

3.4.2 Annotation using Microbial Integrated Genome/ Expert Review (IMG/ER) from the DOE Joint Genome Institute

The contigs, which were aligned and ordered, using *T. scotoductus* SA-01 as template, were submitted to the annotation pipeline at the DOE-JGI. This was done to facilitate comparison between the two genomes; proteins from operons which could possibly be split due to assembly problems could therefore still be placed next to each other after the alignment.

The IMG/ER system was used to do comparative analysis amongst three *Thermus* strains namely *T. thermophilus* HB27, *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain. The comparison ranged from a broad approach of composition, total number of genes to the comparison of individual ORFs or metabolic pathway, summary is shown in (Table 3.6).

A closer look at *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain indicate a similar total number of bases in their respective genomes (2.4 Mbp and 2.3 Mbp) was obtained with a G+C percentage of (64.9% and 65.3%) for *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain, respectively. However, the *Thermus* sp. NMX2 A.1strain genome is still in draft form.

Table 3.6: Comparing the genome features of two *Thermus* strains (*T. thermophilus* HB27, *T. scotoductus* SA-01) to the newly sequenced *Thermus* sp. NMX2 A.1 strain using the IMG/ER interface from DOE-JGI

Compositions	***Thermus thermophilus* HB27**	***Thermus scotoductus* SA-01**	***Thermus* strain NMX2 A.1**
DNA, total number of bases	**2127482**	**2355186**	**2293277**
DNA coding number of bases	2017603	2218814	2152019
DNA G+C number of bases	69%	64.9%	65.3%
DNA scaffolds	**2**	**2**	**133**
CRISPR Count	2	1	3
Plasmid Count	1	1	-
Genes total number	**2273**	**2514**	**2522**
Protein coding genes	2210	2461	2468
RNA genes	63	53	54
rRNA genes	6	6	7
5S rRNA genes	2	2	1
16S rRNA genes	2	2	5
23S rRNA genes	2	2	1
tRNA genes	47	47	43
Other RNA genes	10	-	4

3.4.3 Comparison between the two online pipelines

After obtaining annotation results from both the online pipelines, comparison was done by looking for any differences that might have resulted. (Table 3.7) shows the summarized data; the two pipelines shows differences within the ORFs of the same strain, as it is indicated with the Manatee pipeline that the value is in fact an over estimation this was because of the many mistakes and overlapping of contigs observed when the data was manually checked.

Meanwhile, the same amount of total number of bases is found in both of the pipeline annotations, which gives a good estimate of 2.3 Mbp as the actual size. Further analyses were carried out using the IMG/ER annotation pipeline because it appears more robust with respect to annotation and the website offers a large number of tools for data analysis.

Table 3.7: Summarized comparison between the two online pipelines used for annotation

University of Maryland (Manatee) outcomes	DOE-JGI (IMG) outcomes
• ORFs = 2 721 (over estimation)	• ORFs = 2 468

3.5 Genome comparison

3.5.1 Bi-directional BLAST

Bi-directional BLAST results (Figure 3.7) indicate the comparison of the forward and reverse ORFs of *Thermus* sp. NMX2 A.1 strain against the selected related strains. A visual inspection of Figure 3.5 show that *T. scotoductus* SA-01 is highly similar to *Thermus* sp. NMX2 A.1 strain, as already illustrated by MAUVE analysis. Sections of the *Thermus* sp. NMX2 A.1 strain genome are common to all the strains compared (an example indicated by the red arrow) and these can be considered to represent the core genome of the *Thermus* sp. included in the analysis. A detailed analysis of the core genes, which comprise (35.2%) of the total, will not be made but rather those which usually consist of essential pathways such as energy metabolism or DNA replication.

What was also observed within in the Figure 3.5 was that, *Thermus* sp. NMX2 A.1 strain showed the highest similarity (81.6%) with *T. scotoductus* SA-01. Short sections of the genome (165 genes) are common to *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain only. Hypothetical or conserved hypothetical proteins represent (61 genes). The remaining genes are proteins with putative functions (9 genes), transporters (10 genes), CRISPR associated proteins (16 genes), enzymes associated with nitrate reduction (see discussion later) and a diverse mixture of other proteins.

Thermus sp. NMX2 A.1 strain contains unique genes (10.9%) which are not present in any of the other *Thermus* strains used in the analysis. Of these, (58%) were annotated as hypothetical or uncharacterized proteins. Of special interest there are (12) ORFs (TNMX_1839 to TNMX_1850) which form a putative transcriptional unit and which code for enzymes of the Calvin pathway. A group of (16) genes which are shared only with *T. scotoductus* SA-01, correspond with part of a gene island which were acquired by *T. scotoductus* SA-01.

Interestingly, analysis for putative horizontally transferred genes and islands using SeqWord Sniffer or the predictions of the JGI annotation pipeline failed to predict any

alien genes. This is unexpected as (8.3%) of the ORFs in the *Thermus* sp. NMX2 A.1 strain had no homologues in the nr database. It is possible that the prediction methods did not detect putative alien genes because the genome was only partially assembled or because of amelioration if the genes were acquired a long time ago (Lawrence & Ochman, 1997)

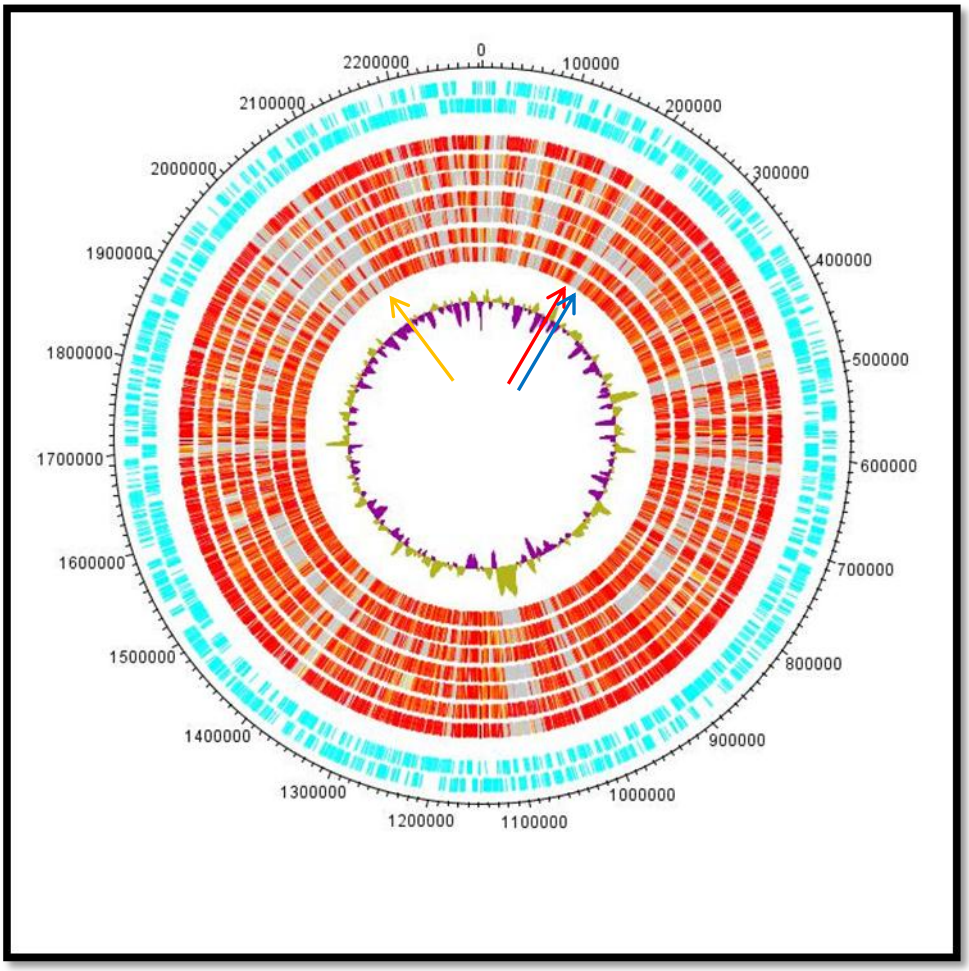


Figure 3.7: Seven-way comparison of genomes of choice used for the Bi-Di Blast analysis. The outer blue two rings represents the forward and reverse ORF's of *Thermus* sp. NMX2 A.1 strain followed by the strains which shows similarities within sequences when compared with *Thermus* sp. NMX2 A.1 strain (From outside to inside: *T. scotoductus* SA-01, *T. thermophilus* HB27, *T. thermophilus* HB8, *T. thermophilus* JL-18, *Thermus* sp, RL, *Thermus* sp. CCB US3 UF1 and *T. thermophilus* SGO-5JP17). The red lines indicate 90%-100% identity; whereas the grey lines indicate a low 0%-10% identity. The inner most ring indicates the GC content.

3.5.2 Natural competence

Some of the *Thermus* strains have been shown to possess natural competence and are able to take up intact DNA with high efficiency (Friedrich *et al*., 2003). This was also demonstrated for *T. scotoductus* SA-01 (Gounder *et al*., 2011). The presence of genes or gene islands which are unique to *Thermus* sp. NMX2 A.1 strain prompted analysis for competence proteins in the *Thermus* sp. NMX2 A.1 strain using BLAST. A comparison of the natural competency genes of both *T. scotoductus* SA-01 and *T. thermophilus* HB27 were compared to those of *Thermus* sp. NMX2 A.1 strain as shown in (Table 3.8).

Gounder and co-workers suggested that the possibility of the presence of PilA in *T. thermophilus* HB27 and not in *T. scotoductus* SA-01 could be the result of superior transformability within *T. thermophilus* HB27 (Gounder *et al*., 2011). Moreover, what is observed in the table below is that the PilA is in fact present within *Thermus* sp. NMX2 A.1 strain. In addition, this newly sequenced strain appears to have most of the genes required for natural competence.

Table 3.8: Curated natural competency genes after blasting the sequences present within *Thermus* sp. NMX2 A.1 strain

***Thermus* NMX2 A.1**	**Gene product names**	***T. scotoductus* SA-01**	***T.thermophilus* HB27**
TNMX_01921	type IV pilus assembly protein PilM	PilM: competence protein PilM	PilM: pilus-associated protein pilM
TNMX_01922	PilN: competence protein PilN	PilN: competence protein PilN	PilN: competence protein pilN
TNMX_01923	Tfp pilus assembly protein PilO	PilO: competence protein PilO	PilO: competence protein pilO
TNMX_01924	PilW: competence protein PilW	PilW: competence protein PilW	PilW: competence protein pilW
TNMX_01925	Type II secretory pathway, component PulD	general secretion pathway protein	PilQ: competence protein pilQ
TNMX_00189	Type II secretory pathway, ATPase PulE/Tfp pi lusassembly pathway, ATPase PilB	PilF: type IV pilus assembly protein PilF	PilF: ATP-binding motif-containing protein pilF
TNMX_00707	Type II secretory pathway, component PulF	PilC: type 4 fimbrial assembly protein PilC	PilC: pili assembly protein pilC
TNMX_00240	Type II secretory pathway, prepilin signalpep tidase PulO and related peptidases (2.1.1.-)	leader peptidase PilD	PilD: type 4 prepilin peptidase pilD
TNMX_00030	competence factor ComEA	competence factor ComEA	ComE: competence protein/comEA-related protein
TNMX_00031	DNA internalization-related competence protei nComEC/Rec2	DNA internalization-related competence protein ComEC/Rec2	ComE: competence factor comEC
TNMX_00783	Predicted amidophosphoribosyltransferases	ComF: competence protein ComF	competence protein F
TNMX_00832	hypothetical protein	no bi-blast hit	ComZ: comZ
TNMX_00643	DNA protecting protein DprA	DprA: competence protein DprA	DprA: competence protein dprA
TNMX_00416	competence/damage-inducible protein CinA C-te rminal domain	competence/damage-inducible protein	CinA: competence-damage protein cinA
TNMX_02266	Uncharacterized conserved protein	FimA: major pilin protein FimA	FimA: major pilin protein fimA
TNMX_01638	Predicted nucleic acid-binding protein, conta insPIN domain	PilT protein domain protein	no biblasthit
TNMX_00762	prepilin-type N-terminal cleavage/methylation domain	no bi-blast hit	PilA: pilA

Genome sequence and annotation

TNMX_ 00832 still remains a hypothetical protein with an expected value of 0.0 to its homolog. It can, therefore, be concluded that *Thermus* sp. NMX2 A.1 strain can in actual fact carry out natural competency due to all the genes present in *T. scotoductus* SA-01 and two more at which the prepilin-type N-terminal cleavage/methylation domain is regarded as the one with great potential in enhancing the transformability, the curated genes are shown in (Supplement A).

3.5.3 Calvin cycle

The Calvin cycle is a light independent (dark-phase of photosynthesis) process of fixing carbon. It utilizes light energy stored as ATP and NADPH to convert CO_2 and H_2O to organic compounds. The process of regulating the mechanism is made possible by functional enzymes within the Calvin cycle (Nowitzki *et al*., 1995). In relation to the genus *Thermus*, it was reported to comprise of eight recognized species by the time da Costa and co-workers published in the year 2006 (Zhang *et al*., 2010). Within those strains none of them were reported to have the Calvin cycle genes not even *T. scotoductus* SA-01.

After sequencing and annotating the newly sequenced *Thermus* sp. NMX2 A.1 strain, some of the Calvin cycle enzymes were found amongst the unique genes present in a putative transcriptional unit (Figure 3.8) of *Thermus* sp. NMX2 A.1 strain. Table 3.9 and Figure 3.9 indicates the enzymes that are found within the *Thermus* sp. NMX2 A.1 strain which correspond to the ones that play essential functions in carrying out the mechanism of the Calvin cycle.

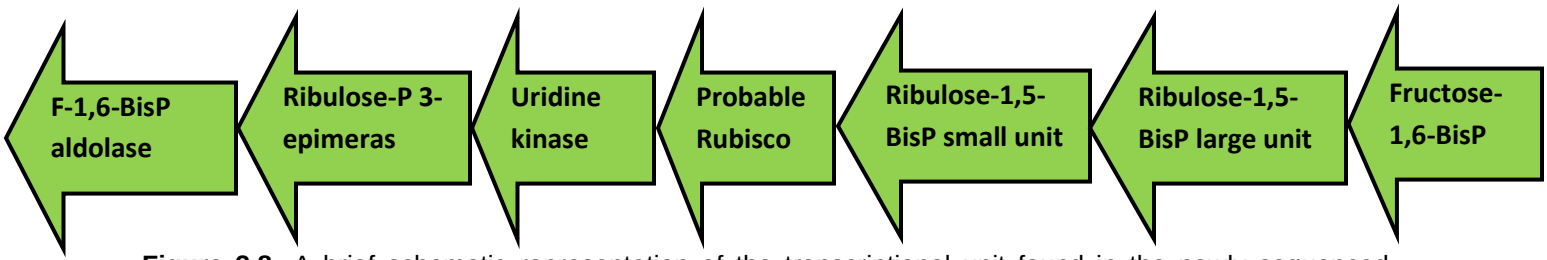


Figure 3.8: A brief schematic representation of the transcriptional unit found in the newly sequenced *Thermus* sp. NMX2 A.1 strain; with enzymes that function within the Calvin cycle mechanism.

Table 3.9: A comparison between enzymes that are present in the Calvin cycle; which *Thermus* sp. NMX2 A.1 strain also has amongst its unique genes

Calvin cycle enzymes	*Thermus* sp. NMX2 A.1 strain enzymes
EC 1.2.1.13 glyceraldehyde-3-phosphate dehydrogenase ($NADP^+$) (phosphorylating)	TNMX_01234 glyceraldehyde-3-phosphate dehydrogenase, type I
EC 2.2.1.1 transketolase	TNMX_01851 transketolase, bacterial and yeast
EC 2.7.1.19 phosphoribulokinase	TNMX_01838 uridine kinase
EC 2.7.2.3 phosphoglycerate kinase	TNMX_01801 2-phosphoglycerate kinase
EC 3.1.3.11 fructose-bisphosphatase	TNMX_01842 fructose-1,6-bisphosphatase, class II
EC 3.1.3.37 sedoheptulose-bisphosphatase	-
EC 4.1.1.39 ribulose-bisphosphate carboxylase	TNMX_01840 Ribulose bisphosphate carboxylase small subunit
EC 4.1.2.13 fructose-bisphosphate aldolase	TNMX_01836 fructose-1,6-bisphosphate aldolase, class II, various bacterial and amitochondriate protist
EC 5.1.3.1 ribulose-phosphate 3-epimerase	TNMX_01837 ribulose-phosphate 3-epimerase
EC 5.3.1.1 triose-phosphate isomerase	TNMX_01232 triose-phosphate isomerase
EC 5.3.1.6 ribose-5-phosphate isomerase	TNMX_01611 ribose-5-phosphate isomerase
EC 5.3.1.9 glucose-6-phosphate isomerase	TNMX_00122 glucose-6-phosphate isomerase

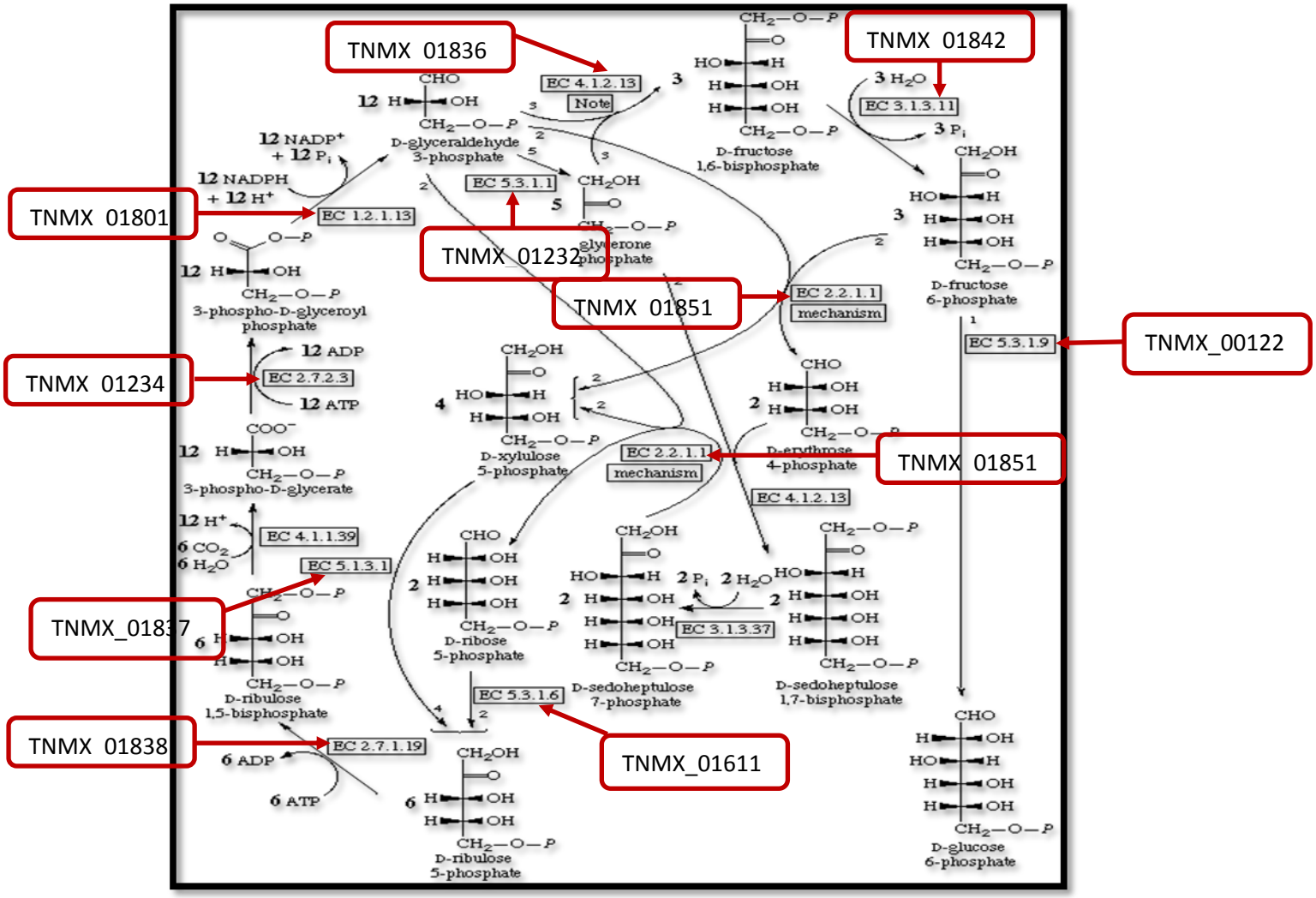


Figure 3.9: A schematic representation of the Calvin cycle enzymes corresponding with the ones also found within the *Thermus* sp. NMX2 A.1 strain, which are shown by the red blocks.

These set of enzymes that play a role in the mechanism of the Calvin cycle, can hypothetically state that the *Thermus* sp. NMX2 A.1 strain has the ability of using carbon dioxide as its only source of carbon; however, this has to be experimentally verified.

4. The Nar operon

Bricio and co-workers stated in their literature that the ability to grow by nitrate respiration is encoded by a cluster of genes named the nitrate respiration conjugative element (NCE). The presence of the genes gives the specific strain the ability to alternatively grow under partial denitrifcation in which reduction of nitrate to nitrite results; or it can grow completely where nitrite is reduced to a gas form as a final product (Bricio *et al*., 2011).

The sequence of the NCE consists of a seven major gene operons that encode a heterotetrameric Nar (nitrate reductase) (NarCGHI), its dedicated chaperone (NarJ) and two nitrate/nitrite transporters (NarK and NarT). A second operon was found to encode a four-subunit NADH dehydrogenase (NrcDEFN) (Bricio *et al*., 2011).

The presences of these genes are responsible for the reduction of a nitrate product by the specific strain subjected into such conditions. In the next chapter, the experimental procedures will be discussed on *Thermus* sp. NMX2 A.1 strain when subjected to anaerobic conditions.

In addition to the BiDi BLAST outcomes, the (Figure 3.10 and Table 3.10) below indicates the proteins present in the newly sequenced *Thermus* sp. NMX2 A.1 strain that are present within the Nar operon. It can, therefore, be predicted that it can grow in the absence of oxygen, with media supplemented with a nitrate source; which will in turn serve as a final electron acceptor.

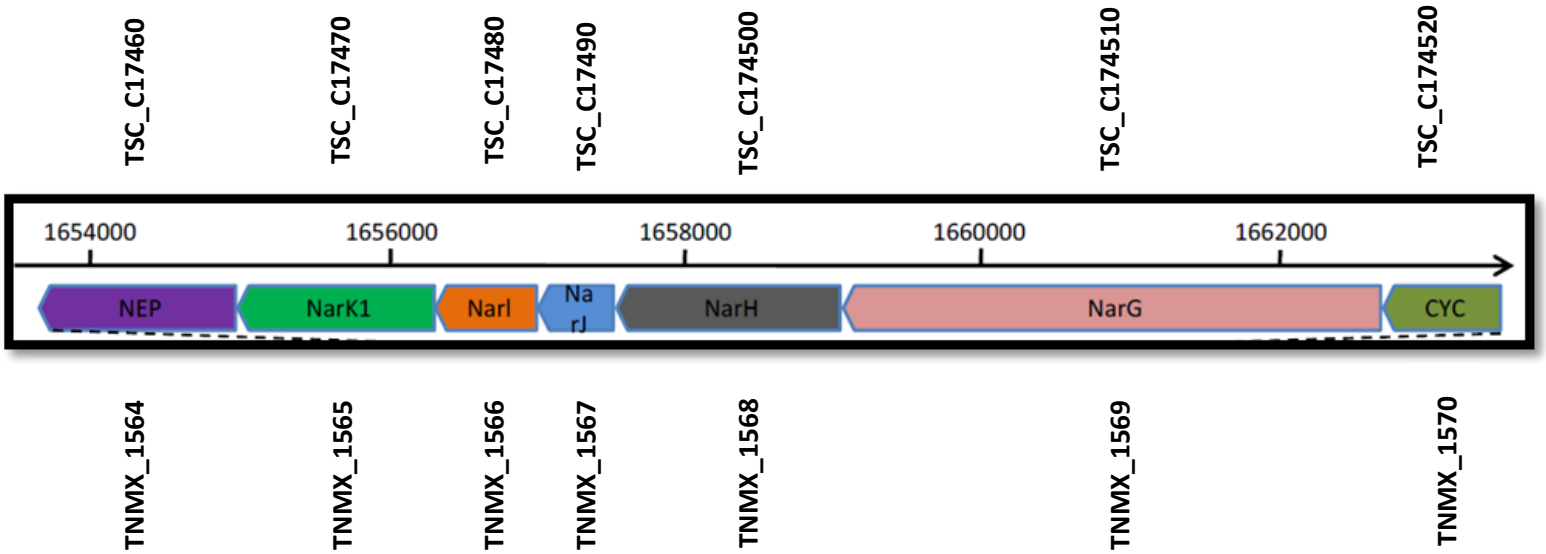


Figure 3.10: A schematic representation of the **NEP**=nitrite extrusion protein and Nitrate (Nar) operon proteins found in both *Thermus* sp. NMX2 A.1 and *T. scotoductus* SA-01.

Table 3.10: Nar operon genes compared from outcomes obtained from the Bi-directional BLAST between *Thermus* sp. NMX2 A.1 and *T. scotoductus* SA-01

***Thermus* sp. NMX2 A.1 strain**	***T. scotoductus* SA-01**
Nitrate/nitrite transporter	NarK1 protein
respiratory nitrate reductase, gamma subunit	NarI:respiratory nitrate reductase, subunit gamma
Nitrate reductase delta subunit	NarJ:nitrate reductase molybdenum cofactor assembly chaperone
nitrate reductase, beta subunit	NarH:nitrate reductase, subunit beta
respiratory nitrate reductase, alpha subunit	NarG:nitrate reductase, subunit alpha
Cytochrome c553	periplasmic cytochrome C
Bacterial transcriptional activator domain	Regulatory protein A
cAMP-binding protein catabolite geneactiva and regulator	Denitrification regulator T
Indolepyruvate ferrodoxin oxidorectase, alpha & beta sheet	Ferredoxin-1

5. Conclusions

A new generation of sequencing technologies; from Roche/454 GS-FLX, Illumina Solexa Genome Analyzer, Applied Biosystems SOLID™ System, Polonator G.007, Ion Torrent, Helicos Heliscope™ and Pacific Biosciences SMRT provided unprecedented opportunities for high-throughput functional genomic research. To date, tremendous transformations of the NGS technologies have led to application in various fields and in a variety of contexts, including whole-genome sequencing, targeted re-sequencing, discovery of transcription factor binding sites, and non-coding RNA expression profiling. Due to the improved cost effectiveness and their multiple applications, the NGS are widely found in the market and greatly sold on the basis of best suitable platform for a given biologic experiment.

After verifying the strain from which 16S rRNA PCR amplification and Sanger sequencing were applied. Consequently, with the application of Roche/454 GS-FLX platform, *Thermus* sp. NMX2 A.1 strain was sequenced; followed by processing the data in all relevant channels at which the raw data got assembled using the *de novo* assembly (Roche newbler software). Moreover, the use of two online pipelines namely the JCVI annotation at the University of Maryland as well as the Microbial Integrated Genome/Expert Review from the DOE Joint Genome Institute came to play an important role in functionality and curation of genes. Additional genome comparative systems used included MAUVE alignment and the BiDi BLAST.

The genome size of *Thermus* sp. NMX2 A.1 strain was observed to be 2.3 Mbp with 2 486 open reading frames, and the comparing systems used indicate great similarities between *Thermus* sp. NMX2 A.1 strain and *T. scotoductus* SA-01. Such was observed when the MAUVE chromosomal alignment showed good synteny between the two strains; also, the BiDi BLAST showed up to (81.6%) close relation. Also, from the MAUVE chromosomal alignment outcomes it was seen that *Thermus* sp. NXM2 A.1 strain has unique genes which *T. scotoductus* SA-01 lack; but, *T. scotoductus* SA-01 also has islands of which the newly sequenced strain does not possess. This could be

the result of which the *T. scotoductus* SA-01 was isolated from a deep gold mine (3.2 km) in South Africa and *Thermus* sp. NMX2 A.1 strain being isolated from thermal hot springs in New Mexico, USA.

Three functionalities were chosen for specific comparison analysis, competency, carbon fixation and the Nar operon. The Nar operon genes that were found in *T. scotoductus* SA-01 were also present in *Thermus* sp. NMX2 A.1 strain except the prepilin-type N-terminal cleavage/methylation domain.

However, in addition to the similarities observed between the two closely related strains as well as additional known *Thermus* strains, *Thermus* sp. NMX2 A.1 strain appear to have unique genes within its genome of which none of the other widely known *Thermus* strains have. *Thermus* sp. NMX2 A.1 strain account (10.9%) for these unique genes, they also appear to play a role in the Calvin cycle putative transcription unit. This implies that the newly sequenced *Thermus* sp. NMX2 A.1 strain has potential of using carbon solely as a source of carbon; however, this must be experimentally proven.

With the above stated similarities and differences observed when comparing the newly sequenced strain to *T. scotoductus* SA-01 and additional *Thermus* strains widely known. A great amount of factors or outcomes were ranging within the similarities and good synteny, and a close up between *Thermus* sp. NMX2 A.1 strain and *T. scotoductus* SA-01 indicate even more relation on genome level. So despite the two closely related strains being known for metal reduction, with the data accumulated it can be concluded that *Thermus* sp. NMX2 A.1 strain is closely related to *T. scotoductus* SA-01 on genome level and that it will be safe to regard this strain as a *T. scotoductus* NMX2 A.1 as hypothesized by Balkwill and co-workers (2004) that *Thermus* sp. NMX2 A.1 strain is actually a *T. scotoductus*.

6. Supplement A

6.1 Identification of *Thermus* sp. NMX2 A.1 strain using sequence data

Table 3.11 (a): The 16S rRNA sequence of *Thermus* sp. NMX2 A.1 strain used for identification

tcgattagagtttgatcctggctcagggtgaacgctggcggcgtgcctaagacatgcaagtcgagcggggcaggttta
tacctgttcagcggcagacgggtgagtaacgcgtgggtgacctacccggaagaggcggacaacctggggaaacccag
gctaatccgccatgtggtcctgtcctgtggggcaggactaaagggtggatagcccgcttccggatgggcccgcgtcc
catcagctagttggtggggtaaaggcccaccaaggcgacgacgggtagccggtctgagaggatggccggccacaggg
gcactgagacacgggccccactcctacgggaggcagcagttaggaatcttccgcaatggacggaagtctgacggagc
gacgccgcttggaggaggaagcccttcggggtgtaaactcctgaactggggacgaaagccctgtgtaggggatgac
ggtacccaggtaatagcgccggccaactccgtgccagcagccgcggtaatacggagggcgcgagcgttacccggatt
tactgggcgtaaagggcgtgtaggcggcctggggcgtcccatgtgaaaggccacggctcaaccgtggaggagcgtgg
gatacgctcaggctagagggtgggagagggtggtggaattcccggagtagcggtgaaatgcgcagataccgggagga
acgccgatggcgaaggcagccacctggtccacttctgacgctgaggcgcgaaagcgtggggagcaaaccggattaga
tacccgggtagtccacgccctaaacgatgcgcgctaggtctctgggtttatctgggggccgaagccaacgcgttaag
cgcgccgcctggggagtacggccgcaaggctgaaactcaaaggaattgacgggggcccgcacaagcgggtggagcat
gtggttttaattcgaagcaacgcgaagaaccttaccaggccttgacatgctggggaaacctggggtgaaagcctggg
gtgcccgtgagggagccccagcacaggtgctggcatggaccgtcgtcagctcgtgtcgtgagatgtttgggttaagt
cccgcaacgagcgcaacccctgcccttagtttgccagcgggttgggccgggcactataaggggactgcctgcgaaag
caggaggaaggcggggacgacgtctggtcatcatggcccttacggcctgggcgacacacgtgctacaatgcccacta
cagagcgaagcgacccagtgatggggagcgaatcgcaaaaaggtgggcgtagttcggattggggtctgcaacccgac
cccatgaagccggaatcgctagtaatcgcggatcagccatgccgcggtgaatacgttcccgggccttgtacacaccg
cccgtcacgccatgggagcgggttctacccgaagtcgccgggagccttagggcaggcgcaagagggtagggctcgtg
actggggcgaagtagtaacaattgtttaac

Table 3.11 (b): Obtained results after BLASTing the 16S rRNA sequence of the *Thermus* sp. NMX2 A.1 strain

> gb|L09661.1|TTHNMX2 *Thermus* sp. NMX2 A.1 16S ribosomal RNA (rnn) gene sequence
Length=1470

Score = 2582 bits (1398), Expect = 0.0
Identities = 1445/1466 (**99%**), Gaps = 10/1466 (**1%**)
Strand=Plus/Plus

Table 3.11 (c): Obtained results after BLAST analysis of the 16S rRNA sequence of the *Thermus* sp. NMX2 A.1 strain

Accession	Description	Max score	Total score	Query coverage	E-value	Max indent
L09661.1	*Thermus* sp. NMX2 A.1 16S ribosomal RNA (rnn) gene sequence	822	822	99%	0.0	99%
AF255590.1	*Thermus* sp. SRI-96 16S ribosomal RNA gene, partial sequence	817	817	99%	0.0	99%
DQ861411.1	Uncultured *Thermus* sp. clone c108 16S ribosomal RNA gene, partial sequence	811	811	99%	0.0	98%
DQ414817.1	*Thermus* sp. GD31 16S ribosomal RNA gene, partial sequence	811	811	99%	0.0	98%
NR_024867.1	*Thermus scotoductus* strain Se-1 16S ribosomal RNA, partial sequence >gb|AF032127.1| *Thermus scotoductus* 16S ribosomal RNA gene, complete sequence	808	808	99%	0.0	98%

6.2 Sequences used to curate the hypothetical proteins within the newly sequenced *Thermus* sp. NMX2 A.1 strain

TNMX_01922_ competence protein PilN

```
Score = 408 bits (1048), Expect = 8e-143, Method: Compositional matrix adjust.
 Identities = 203/207 (98%), Positives = 205/207 (99%), Gaps = 0/207 (0%)

Query  1    LIRLNLLPKNLRRRIEPGWWRLAAGAFALLTLSLLGFLHYTAYTELTLAKQERDALKAEV  60
            +IRLNLLPKNLRRRIEPGWWRLAAGAFALLTLS LGFLHYTAYTELTLAKQERDALKAEV
Sbjct  1    MIRLNLLPKNLRRRIEPGWWRLAAGAFALLTLSFLGFLHYTAYTELTLAKQERDALKAEV  60

Query  61   EALKPFIAEQNRLVQERKTLEALLAIREGLKKNFVPWSEYLAAFIRQIPQQGGRLPVALR  120
            EALKPFIAEQNRLVQERK LEALLAIREGLKKNFVPWSEYLAAFIRQIPQQGGRLPVALR
Sbjct  61   EALKPFIAEQNRLVQERKALEALLAIREGLKKNFVPWSEYLAAFIRQIPQQGGRLPVALR  120

Query  121  SVGTRAIPEDEANRMAQAGTYDGKKVQVEFTVQGEALNQNALVSFVRAFETSPRFGIEFQ  180
            SVGTRAIPEDEANRMAQAGTYDGKKVQVEFTVQGEALNQNALV+FVRAFETSPRFGIEFQ
Sbjct  121  SVGTRAIPEDEANRMAQAGTYDGKKVQVEFTVQGEALNQNALVNFVRAFETSPRFGIEFQ  180

Query  181  GASLDQNRGLYTFSARVGLVGGGESAR  207
            GASLDQNRGLYTFSARVGLVGGGESAR
Sbjct  181  GASLDQNRGLYTFSARVGLVGGGESAR  207
```

TNMX_01924_ competence protein PilW

```
Score =   496 bits (1277), Expect = 4e-174, Method: Compositional matrix
adjust.
 Identities = 306/320 (96%), Positives = 307/320 (96%), Gaps = 6/320 (2%)

Query  1    VKALLARLATAWRNLPQSTKLLLAGLLLVSAVALWYVGFYLPAQAPMEVQPTPGQPQVPS  60
            +KALLARLATAWRNLPQSTKLLLAGLLLVSAVALWYVGFYLPAQAPMEVQPTPGQPQVPS
Sbjct  1    MKALLARLATAWRNLPQSTKLLLAGLLLVSAVALWYVGFYLPAQAPMEVQPTPGQPQVPS  60

Query  61   QGAEAPKAIEAPPIPPLAETAPQGQAAKPSPEASPSQKAPKAGEPVLTTALPIPKTQQEA  120
            QGAEAPKAIEAPPIPPLAETAPQ QAAKPSPEASPSQK      PVLTTALPIPKTQQEA
Sbjct  61   QGAEAPKAIEAPPIPPLAETAPQDQAAKPSPEASPSQK------PVLTTALPIPKTQQEA  114

Query  121  PLPNPFVPLVVEAPPSPPVASPAPAPRPTPVPTGAPVRVTQGTPLPTPSVQAPPRPLPGS  180
            PLPNPFVPLVVEAPPSPPV SPAPAPRPTPVPTGAPVRVTQGTPLPTPSVQA PRPLPGS
Sbjct  115  PLPNPFVPLVVEAPPSPPVPSPAPAPRPTPVPTGAPVRVTQGTPLPTPSVQASPRPLPGS  174

Query  181  QGALPAPKVLTPAFQVETPKAQVDIPPLLTPPAGLVEAPLPRERQTPEGGKTEPSPATSP  240
            QGALPAPKVLTPAFQVETPKAQVD P LLTPPAGLVEAPLPR  QTPEGGKTEPSPATSP
Sbjct  175  QGALPAPKVLTPAFQVETPKAQVDTPLLLTPPAGLVEAPLPRAPQTPEGGKTEPSPATSP  234

Query  241  KTPLQALVEEKGIKLAGTLLGPVSVAILETKEGYLVLPVGSLLPGSEAVLRRIESDRVVL  300
            KTPLQALVEEKGIKLAGTLLGPVSVAILETKEGYLVLPVGSLLPGSEAVLRRIESDRVVL
Sbjct  235  KTPLQALVEEKGIKLAGTLLGPVSVAILETKEGYLVLPVGSLLPGSEAVLRRIESDRVVL  294

Query  301  ALKDESLEITLEKMQAGGGQ  320
            ALKDESLEITLEKMQAGGGQ
Sbjct  295  ALKDESLEITLEKMQAGGGQ  314
```

TNMX_00030_ competence factor ComEA

```
Score =   166 bits (419), Expect = 8e-51, Method: Compositional matrix
adjust.
 Identities = 85/85 (100%), Positives = 85/85 (100%), Gaps = 0/85 (0%)

Query  18   WPKLVPKPQPVWVETLAEASFEPPAPEPISLNQASLEELMSLPGVGPVLAQRIVAGRPYA  77
            WPKLVPKPQPVWVETLAEASFEPPAPEPISLNQASLEELMSLPGVGPVLAQRIVAGRPYA
Sbjct  18   WPKLVPKPQPVWVETLAEASFEPPAPEPISLNQASLEELMSLPGVGPVLAQRIVAGRPYA  77

Query  78   RVEDLLRVKGIGPATLERLRPYVRP  102
            RVEDLLRVKGIGPATLERLRPYVRP
Sbjct  78   RVEDLLRVKGIGPATLERLRPYVRP  102
```

TNMX_00832_ hypothetical protein TtJL18_0848

```
 Score =  1234 bits (3194),  Expect = 0.0, Method: Compositional matrix
adjust.
 Identities = 602/603 (99%), Positives = 603/603 (100%), Gaps = 0/603 (0%)

Query  1    MRKGFALVSTLIMLVVLMALLTAYFVLTRVELGTTQASVRQTTGFYAAEAGLNLRAEEIR  60
            MRKGFALVSTLIMLVVLMALLTAYFVLTRVELGTTQASVRQTTGFYAAEAGLNLRAEEIR
Sbjct  1    MRKGFALVSTLIMLVVLMALLTAYFVLTRVELGTTQASVRQTTGFYAAEAGLNLRAEEIR  60

Query  61   AKFLGYNRPQGTSPSPTNPCQGSNWGSGDFACKTYTFSGRTVRTYVVEDPNNPRNVTIPP  120
            AKFLGYNRPQGTSPSPTNPCQGSNWGSGDFACKTYTFSGRTVRTYVVEDPNNPRNVTIPP
Sbjct  61   AKFLGYNRPQGTSPSPTNPCQGSNWGSGDFACKTYTFSGRTVRTYVVEDPNNPRNVTIPP  120

Query  121  GELYENLDAQEYRYTLFSEALGPDGRTEVVLELVFKSRLVPMFQFAAFYNKDLEILPGPD  180
            GELYENLDAQEYRYTLFSEALGPDGRTEVVLELVFKSRLVPMFQFAAFYNKDLEILPGPD
Sbjct  121  GELYENLDAQEYRYTLFSEALGPDGRTEVVLELVFKSRLVPMFQFAAFYNKDLEILPGPD  180

Query  181  MTLNGRAHTNGDLYLNAGNTLTIQGQVSTAGTLYRGRKDDNSCGGTVNIYKENNQPAPLP  240
            MTLNGRAHTNGDLYLNAGNTLTIQGQVSTAGTLYRGRKDDNSCGGTVNIYKENNQPAPLP
Sbjct  181  MTLNGRAHTNGDLYLNAGNTLTIQGQVSTAGTLYRGRKDDNSCGGTVNIYKENNQPAPLP  240

Query  241  CNRQNSRRSFSQGDVGAWGSRIQIGVNPVTVPPPEKLDPLPGKTYWEKAELRLGLDLFST  300
            CNRQNSRRSFSQGDVGAWGSRIQIGVNPVTVPPPE+LDPLPGKTYWEKAELRLGLDLFST
Sbjct  241  CNRQNSRRSFSQGDVGAWGSRIQIGVNPVTVPPPEELDPLPGKTYWEKAELRLGLDLFST  300

Query  301  PPTVKVYRVQGGANVNDVIATTALNACVGAVGTSDTFYNWREGKRIRMLDVDLRKLLDCV  360
            PPTVKVYRVQGGANVNDVIATTALNACVGAVGTSDTFYNWREGKRIRMLDVDLRKLLDCV
Sbjct  301  PPTVKVYRVQGGANVNDVIATTALNACVGAVGTSDTFYNWREGKRIRMLDVDLRKLLDCV  360

Query  361  HLNPLAFGFRLDDDTEGGLVFHFTVFGPDSNRVNNYGVRIKNGAEIASTLPGAPRPRGLT  420
            HLNPLAFGFRLDDDTEGGLVFHFTVFGPDSNRVNNYGVRIKNGAEIASTLPGAPRPRGLT
Sbjct  361  HLNPLAFGFRLDDDTEGGLVFHFTVFGPDSNRVNNYGVRIKNGAEIASTLPGAPRPRGLT  420

Query  421  VVTDQAVYIQGDFNKTNWIPAAFLADSLNILSNAWTNDAKSYQTLSQRVASDTEVNAAFL  480
            VVTDQAVYIQGDFNKTNWIPAAFLADSLNILSNAWTNDAKSYQTLSQRVASDTEVNAAFL
Sbjct  421  VVTDQAVYIQGDFNKTNWIPAAFLADSLNILSNAWTNDAKSYQTLSQRVASDTEVNAAFL  480

Query  481  SGTDITGGEEGPSGQNKGNYNGGLENYPRFHETWSGKTLTYRGSFVSLGTPRHVNGPWCG  540
            SGTDITGGEEGPSGQNKGNYNGGLENYPRFHETWSGKTLTYRGSFVSLGTPRHVNGPWCG
Sbjct  481  SGTDITGGEEGPSGQNKGNYNGGLENYPRFHETWSGKTLTYRGSFVSLGTPRHVNGPWCG  540

Query  541  TGGRGTYDPLTGRISGRRGCNIYDPPTRNWSYEVRFSQGQLPPLSPRFVYLRQERFLREF  600
            TGGRGTYDPLTGRISGRRGCNIYDPPTRNWSYEVRFSQGQLPPLSPRFVYLRQERFLREF
Sbjct  541  TGGRGTYDPLTGRISGRRGCNIYDPPTRNWSYEVRFSQGQLPPLSPRFVYLRQERFLREF  600

Query  601  ERP  603
            ERP
Sbjct  601  ERP  603
```

7. References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol,* **215,** 403-410.

Balkwill, D. L., Keift, T. L., Tsukuda, T., Kostandarithes, H. M., Onstott, T. C., Macnaughton, S., Bownas, J. and Fredrickson, J. K. (2004) Identification of iron-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles*, **8,** 347-44.

Bricio, C., Alvarez, L., GÓmez, M. J. and Berenguer, J. (2011) Partial and complete denitrification in *Thermus thermophilus*: lessons from genome drafts. *Biochem. Soc. Trans,* **39,** 249-253.

Cava, F., Zafra, O., Magalon, A., Blasco, F. and Berenguer, J. (2004) A new type of NADH dehydrogenase specific for nitrate respiration in the extreme thermophile *Thermus thermophilus*. The *J. of Bio.Chem.*, vol.279, pp.45369-45378.

Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. (2004) Methods, Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. *Genome Research*, 1403.

de Almeida, J.M.G.C.F. (2010) BiDiBlast: Comparitive Genomics Pipeline for the PC. *Genomics Proteonomics Bioinformatics*, **8**(2)**,** 135-138.

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Nat. Acad. Sci,* U S A, **100,** 8817–8822.

Friedrich, A., Rumszauer, J., Henne, A. and Averhoff, B. (2003) Pilin-like Protein in the Extremely Thermophilic Bacterium *Thermus thermophilus* HB27: Implication in Competence for Natural Transformation and Links to Type IV Pilus Biogenesis. *App*l. *and Enviro. Microbiol*, volume: 69; **7,** 3695-3700.

Ghahramani, Z. (2001) An introduction to Hidden Markov Models and Bayesian Networks. *Inter*. *J*. *of Pattern Recognition and Artificial Intelligence*, **15**(1)**,** 9-42.

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

Integrated Microbial Genome. https://img.jgi.doe.gov/ (Cited 14-11-2012).

International Human Genome Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931-945.

Kosuge, T., Abe, T., Okido, T., Tanaka, N., Hirahata, M., Maruyama, Y., Mashima, J. and Tomiki, A. (2006) "Exploration and Grading of Possible Genes from 183 Bacterial Strains by a Common Protocol to Identification of New Genes: Gene Trek in Prokaryote Space (GTPS)". *DNA Research* **13** (6)**,** 245–254.

Labuschagne, M. and Albertyn, J. (2007) Cloning of an epoxide hydrolase encoding gene from *Rhodotorula mucilaginosa* and functional expresion in *Yarrowi alipolytica*. *Yeast*. **24**(2)**,** 69-78.

Lane, D. J. (1991) 16S/23S Sequencing. *In* E. Stackebrandt & M. Goodfellow (ed.), Nucleic Acid Techniques in *Bact. Sys*, p.115-175. John Wiley and Sons, New York. USA.

Lawrence, J. G. and Ochman H. (1997) Amelioration of bacterial genomes: rates of changed exchange. *J. Mol .Evol*, **44,** 383-397.

Metzker, M. L. (2005) Emerging technologies in DNA sequencing. *Genome Res*. **15,** 1767-1776.

Metzker, M. L. (2009) Sequencing in real time. *Nature Biotech* **27,** 150-151.

Metzker, M. L. (2010) Applications of Next-Generation Sequencing. Sequence technologies – the next-generation. *Nature* Volume 11.

Nowitzki, U., Wyrich, R., Westhoff, P., Henze, K., Schnarrenberger, C. and Martin, W. (1995) Cloning of the amphibolic Calvin cycle/OPPP enzyme D-ribulose-5-phosphate 3-epimerase (EC 5.1.3.1) from spinach chloroplast: functional and evolutionary aspects. *Plant Mol. Bio,* **29,** 1279-1291.

Promega http://www.promega.com/resources/articles/pubhub/enotes/pgemt-easy-vector-system (Cited 17-06-2012).

Quinn, N. L., Levenkova, N., Chow, W., Bouffard, P., Boroevich, K. A., Knight, J. R., Jarvie, T. P., Lubieniecki, K. P., Desany, B. A., Koop, B. F., Harkins, T. T. and Davidson, W. S. (2008) Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome*. BMC Genomics*, **9,** 404.

Wiley, G., Macmil, S., Qu, C., Wang, P., Xing, Y., White, D., Li, J., White, J. D., Domingo, A. and Roe, B. A. (2009) Methods for Generating Shotgun and Mixed Shotgun/Paired-End Libraries for the 454 DNA Sequencer. *Curr. Proto. in Human Genetics,* 18.1.1-18.1.21.

Zafra, O., RamÍrez, S., Castán, P., Moreno, R., Cava, F., Vallés, C., Caro, E. and Berenguer, J. (2002) A cytochrome *c* encoded by the *nar* operon is required for the synthesis of active respiratory nitrate reductase in *Thermus thermophilus*. FEBB letters, **523,** 99-102.

Zhang, X., Ying, Y., Ye, Y, Xu, X., Zhu, X. and Wu, M. (2010) *Thermus arciformis* sp. nov., a thermophilic species from a geothermal area. *Inter. J. of Sys. and Evol. Microbiol,* **60** (pt 4)**,** 834-839.

Zhou, X., Ren, L., Meng, Q., Li, Y., Yu, Y. and Yu, J. (2010) The next-generation sequencing and application. *Protein Cell*, **1**(6)**,** 520–536.

CHAPTER 4

1. Introduction

The genus *Thermus* constitute hundreds of strains commonly found in natural and artificial thermophilic environments throughout the world. Most of these strains can grow optimally under laboratory conditions and some can be genetically manipulated due to the presence of a highly efficient natural competence system (Bricio *et al*., 2011).

Many prokaryotes and archaea are reported to obtain metabolic energy by redox processes involving nitrogen compounds, such as denitrification (Cabello *et al*., 2004). Anaerobic respiration is an alternative form of respiration which occurs in several anaerobic and facultative anaerobic bacteria by the reduction of nitrogen oxides (NO_3^-) to nitrite (NO_2^-), NO (nitric oxide), N_2O (dinitrogen oxide) and nitrogen gas (N_2) in four consecutive steps, by catalysis of respective reductases (Mesa *et al*., 2001) in which nitrate serves as an alternative to oxygen in anaerobic respiration (Ramírez *et al*., 2000).

A recap on what was mentioned in the general introduction chapter is that when a comparison was done amongst *T. scotoductus* SA-01, *T. thermophilus* strains HB8 and HB27 it was evident that they varied from one another on many levels which included the Nar operon. However, when a comparison was done between *T. scotoductus* SA-01 and the *Thermus* sp. NMX2 A.1 strain as pointed out in chapter 3 good synteny is observed and many similarities too. These findings are further supported by a molecular procedure in which the application of genomic fingerprinting of the BOX A repeat sequence showed close similarities between the metal-reducing strains (Figure 4.1, Balkwill *et al*., 2004).

Comparison of nitrate reduction and functionalities

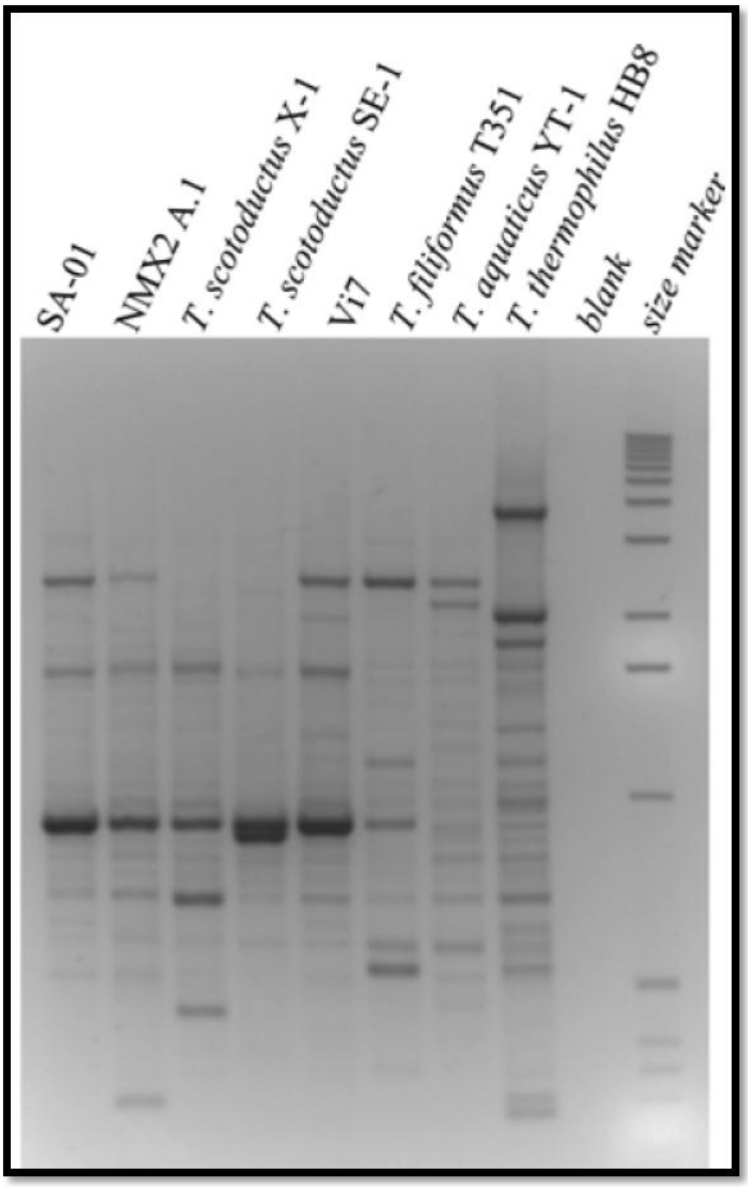


Figure 4.1: BOX- PCR fingerprints of *Thermus* sp. SA-01, *Thermus* sp. NMX2 A.1 strain, *T. scotoductus* X-1, *T. scotoductus* SE-1, *Thermus* sp. VI-7, *T. filiformis* T351, *T. aquaticus* YT-1, and *T.thermophilus* HB8 (Taken from Balkwill *et al*., 2004). In this study the focus will mostly be on *Thermus* sp. SA-01 and *Thermus* sp. NMX2 A.1 strain.

Comparison of nitrate reduction and functionalities

1.1 The aim of the chapter

Genomic sequence data of *Thermus* sp. NMX2 A.1 strain suggest that it might be *T. scotoductus* SA-01. It is reported that *T. scotoductus* can grow under anaerobic conditions when its medium is supplemented with nitrate as the final electron acceptor. All of this is made possible by the presence of the nitrate respiration cassette within *T. scotoductus* SA-01 (Gounder *et al*., 2011).

The aim of this chapter was to determine if *Thermus* sp. NMX2 A.1 strain can reduce nitrate anaerobically, to extend data on the newly sequenced strain abilities and assess if the genome annotation is displayed in functionality under anaerobic respiration.

2. Materials and Methods

2.1 Strains and growth conditions

Strains used in this experiment were from the following individuals:

Table 4.1: Strains used in this chapter

Strains	Candidate
Thermus sp. NMX2 A.1 strain	Prof. Keift (New Mexico, USA)
T. thermophilus HB27	Prof. Berenguer (Madrid, Spain)
T. scotoductus SA-01	Prof. Keift (New Mexico, USA)

The respective strains were grown aerobically for isolation of genomic DNA, *Thermus* sp. NMX2 A.1 strain was cultured in ATTC medium 697 [4 g/L yeast extract, 8 g/L peptone and 2 g/L NaCl; pH 7.5] at 68°C, *T. thermophilus* HB27 was cultured in *Thermus* Broth [8 g/L tryptone, 4 g/L yeast extract and 3 g/L NaCl; pH 7.5] 70°C and *T. scotoductus* SA-01 was cultured in tryptone, yeast extract and glucose (TYG)[5 g/L tryptone, 3 g/L yeast extract and 1 g/L glucose; pH 7.5] at 65°C.

The optical density 600 nm (OD_{600nm}) of the pre-inocula was monitored to a value of ~ 0.6. Anaerobic inoculums were achieved by growing the respective strains in their corresponding media supplemented with (40 mM) KNO_3 and overlaid with liquid paraffin in vial bottles closed with rubber stoppers and crimped, and incubated at their respective temperatures. The (OD_{600nm}) was measured using a spectrophotometer at intervals of 4 h for a period of 72 h.

2.2 Genomic DNA extraction

Genomic DNA extraction was performed using the NucleoSpin® Soil kit; following the manufacturer's instructions. The quantity of the genomic DNA extract was analyzed by measuring DNA concentration and purity using a NanoDrop spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE). The quality and integrity of the genomic DNA was also evaluated by agarose gel electrophoresis and visualized on a 0.8% (w/v) agarose gel containing 2.5 mg/µL ethidium bromide. The agarose gel was prepared in TAE buffer containing 100 mM Tris-HCl (pH 8), 0.05 M ethylenediaminetetraacetic acid (EDTA) and 100 mM glacial acetic acid. Electrophoresis was performed at 6 V/cm for 30 min in TAE buffer and the genomic DNA was visualized using a ChemDoc XRS (Biorad Laboratories) gel documentation system.

2.3 PCR amplification of 16S rRNA gene

The identity of the *Thermus* strains were confirmed by amplification and sequencing of the 16S rRNA gene from the genomic DNA using the universal bacterial primers 27F and 1492R as described in (section 2.3) of chapter 3.

2.4 Cloning of the 16S rRNA gene into pGEM-T Easy vector

The purified product was ligated into pGEM-T Easy vector as described in (section 2.4) of chapter 3.

2.5 Comparative analysis of molecular sequence data

A close comparison amongst the three strains was carried out with the application of multiple alignments using CLUSTAL W (Thompson *et al*., 1994). In addition, *Thermus* sp. NMX2 A.1 strain and *T. scotoductus* SA-01 were closely aligned to evaluate the homology between them.

After aligning the 16S rRNA of the *Thermus* strains, a phylogenetic tree was constructed with addition of related strains such as *Deinococcus* and *Meiothermus* downloaded from the Ribosomal Database Project (RDP, Maidak *et al*., 2001). The main idea was to observe the divergence within the genus *Thermus* and its relatives. Using the Molecular Evolutionary Genetics Analysis (MEGA) 5 software (Tamura *et al*., 2011) maximum likelihood and neighbor-joining trees were constructed.

2.6 Nitrate reduction monitoring

Anaerobic growth was achieved by growing the respective strains in 50 mL of their corresponding media within a 250 mL sized vials, supplemented with (10 mM) KNO_3 and overlaid with liquid paraffin in vials closed with rubber stoppers and crimped shut, followed by incubation at their respective corresponding temperatures. To detect the nitrate reductase activity, the Griess Reagent Kit for nitrite determination (G-792; Griess, 1879) was used and a standard curve was constructed.

2.6.1 Nitrite standard curve

Nitrite dilutions were made by diluting a stock solution with distilled water. A nitrite standard curve (Figure 4.2) was constructed by plotting the absorbance (OD_{548nm}) readings against the known nitrite concentrations in µM.

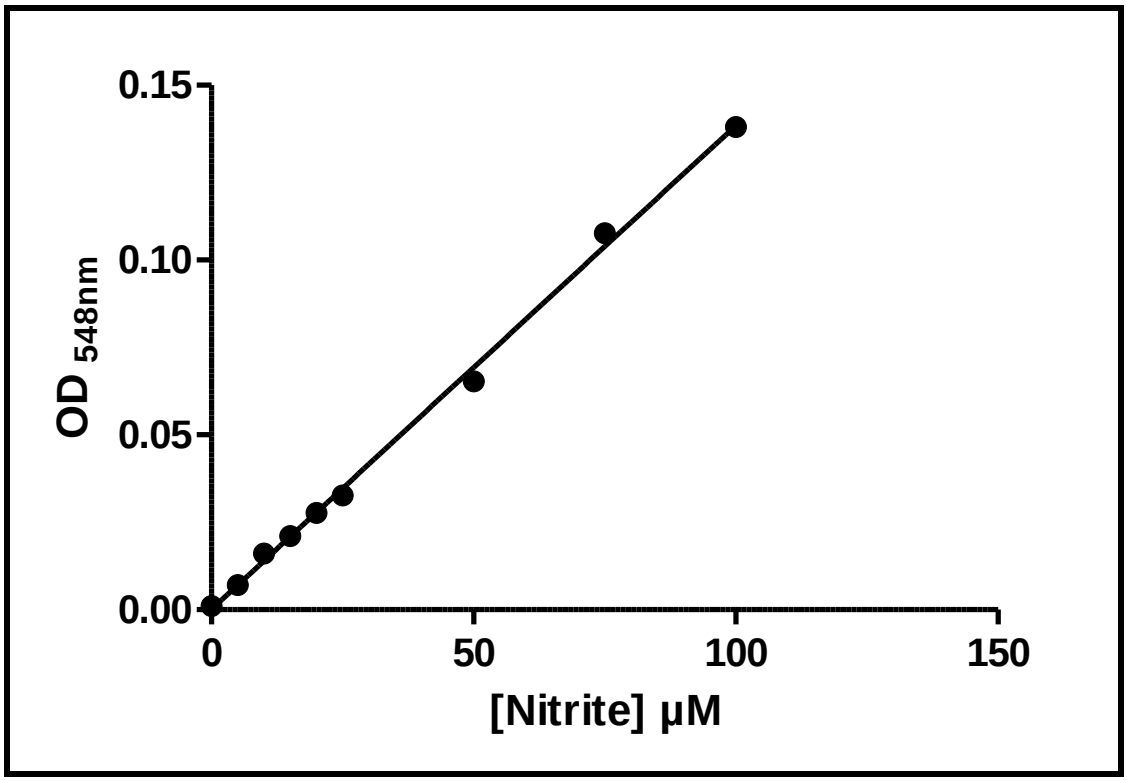


Figure 4.2: Standard curve indicating the relationship between nitrite concentration and OD_{548nm} (R^2= 0.9959). Standard deviations are smaller than the symbols.

3. Results and discussion

3.1 Conformation of *Thermus* strains

Genomic DNA was successfully isolated from the *Thermus* strains as described (section 2.2). The bands showed high integrity DNA and almost no shearing, in Figure 4.3 with concentrations of 174,5 - 291,2 ng/µL. The A_{260}/A_{280} absorbance ratio of 1.8 indicated low protein or RNA contamination.

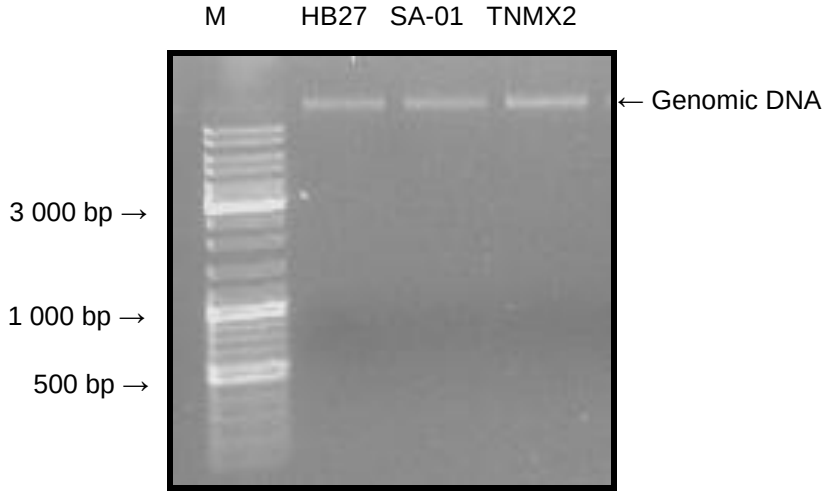


Figure 4.3: Extracted genomic DNA from *Thermus* strains visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; GeneRuler™ DNA ladder (Fermentas), lane 1-3; isolated genomic DNA >10kbp of (*T. thermophilus* HB27, *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain).

The 16S rRNA gene was amplified as described (section 2.3) and a DNA fragment of approximately 1 500 bp was obtained as expected for bacterial amplicons (Figure 4.4). As described in section 2.4 the 16S rRNA gene of the *Thermus* strains were subjected to BLAST analysis against NCBI database to confirm the identity of the organisms, a 99% coverage query showed a 99% identity with an E-vale of zero, in (Supplement B).

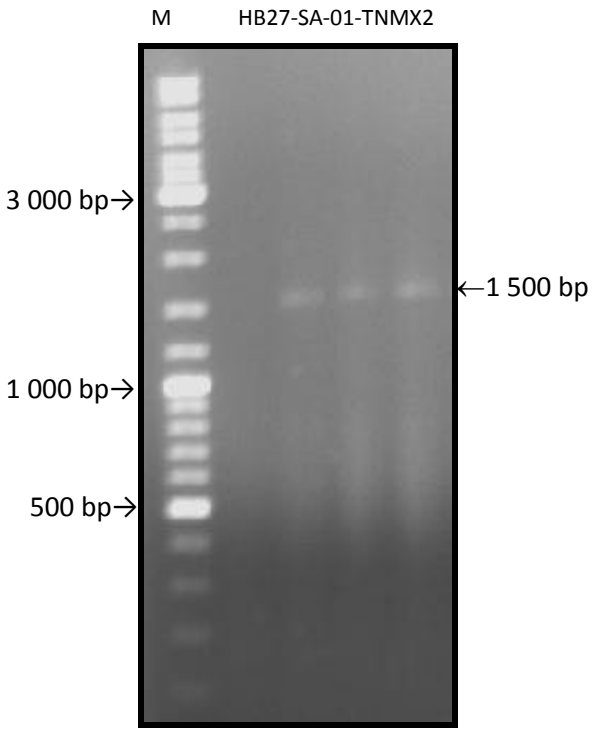


Figure 4.4: Amplification of 16S rRNA fragment from genomic DNA on an ethidium bromide-stained agarose gel 0.8% (w/v). Lane M; GeneRuler™ DNA ladder (Fermentas), lane 1 - 3; positive amplified band of 16S rRNA of (*T. thermophilus* HB27, *T*. *scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain).

3.2 Molecular determination of the phylogenetic positions

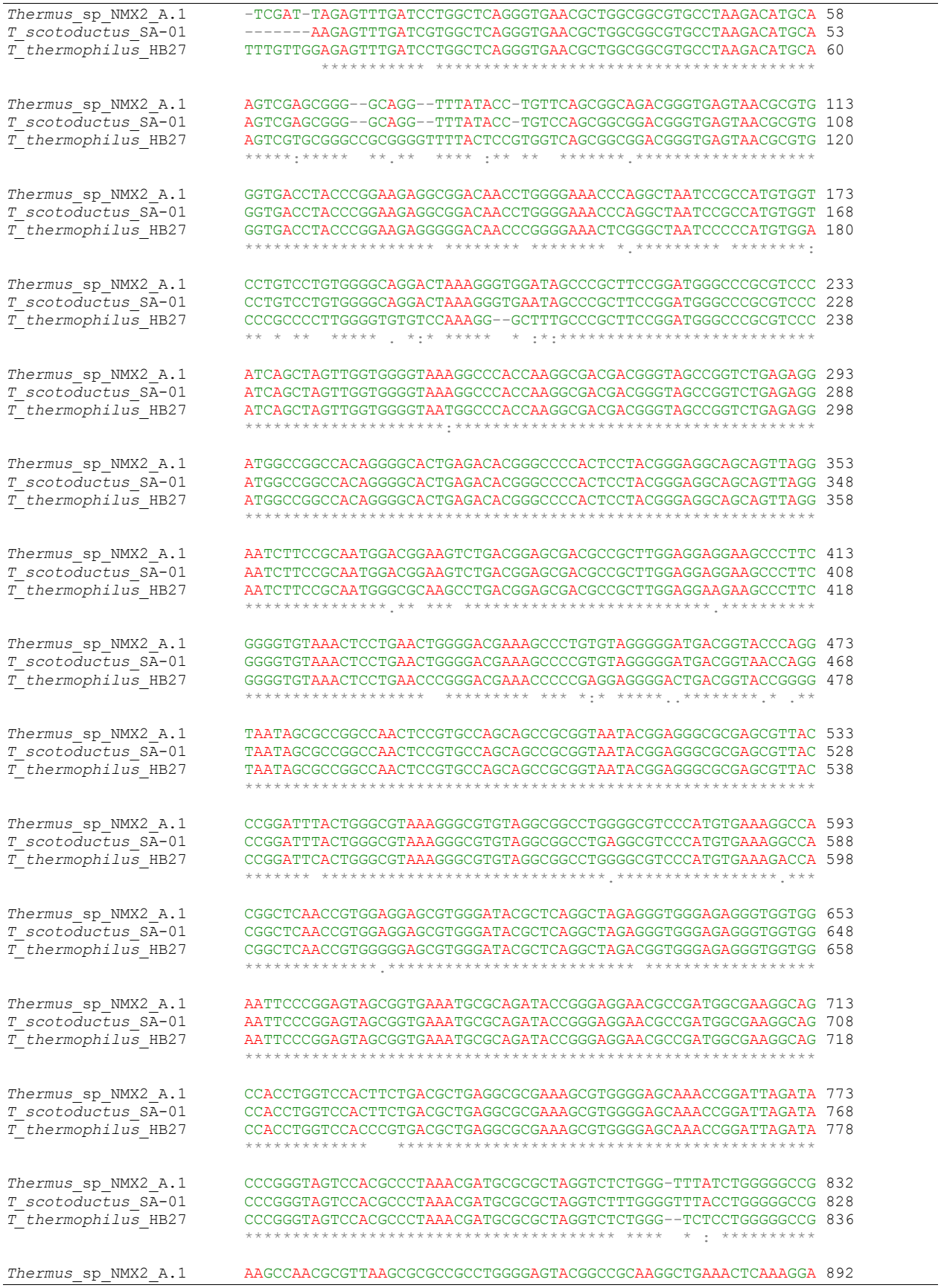
As described in (section 2.5), comparative analysis was done using the 16S rRNA sequences of the three *Thermus* strains at first, then again using the two closely related metal-reducing *Thermus* strains. The outcomes are shown in (Figure 4.5).

CLUSTAL W outcomes therefore summarized the below observed multiple alignment as follow for the first alignment **1** *Thermus* sp. NMX2 A.1 1494 aa; sequence **2**: *T*. *scotoductus* SA-01 1380 aa and sequence **3**: *T*. *thermophilus* HB27 1504 aa gave sequences (**1:2**) an aligned score of: 97%, sequences (**1:3**) an aligned score of: 92%, sequences (**2:3**) an aligned score of: 93%. Meanwhile, the second alignment sequence **1**: *Thermus* sp. NMX2 A.1 strain 1494 aa; sequence **2**: *T*. *scotoductus* SA-01 1380 aa giving a sequences (**1:2**) an aligned score of: 97%.

Comparison of nitrate reduction and functionalities

Phylogenic trees where drawn using MEGA 5 as described in (section 4.5), the application of maximum likelihood and neighbor-joining indicated the following where observed as indicated in (Figure 4.6).

The comparative data shows close relation between *T*. *scotoductus* SA-01 and the *Thermus* sp. NMX2 A.1 strain, which correlates to the experiment performed by (Balkwill et al., 2004) using the BOX-PCR fingerprints of the *Thermus* strains and not much relation is observed between *T. thermophilus* HB27 from both the other strains.


Thermus_sp_NMX2_A.1 -TCGAT-TAGAGTTTGATCCTGGCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCA 58
T_scotoductus_SA-01 -------AAGAGTTTGATCGTGGCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCA 53
T_thermophilus_HB27 TTTGTTGGAGAGTTTGATCCTGGCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCA 60
Thermus_sp_NMX2_A.1 AGTCGAGCGGG--GCAGG--TTTATACC-TGTTCAGCGGCAGACGGGTGAGTAACGCGTG 113
T_scotoductus_SA-01 AGTCGAGCGGG--GCAGG--TTTATACC-TGTCCAGCGGCGGACGGGTGAGTAACGCGTG 108
T_thermophilus_HB27 AGTCGTGCGGGCCGCGGGGTTTTACTCCGTGGTCAGCGGCGGACGGGTGAGTAACGCGTG 120
Thermus_sp_NMX2_A.1 GGTGACCTACCCGGAAGAGGCGGACAACCTGGGGAAACCCAGGCTAATCCGCCATGTGGT 173
T_scotoductus_SA-01 GGTGACCTACCCGGAAGAGGCGGACAACCTGGGGAAACCCAGGCTAATCCGCCATGTGGT 168
T_thermophilus_HB27 GGTGACCTACCCGGAAGAGGGGGACAACCCGGGGAAACTCGGGCTAATCCCCCATGTGGA 180
Thermus_sp_NMX2_A.1 CCTGTCCTGTGGGGCAGGACTAAAGGGTGGATAGCCCGCTTCCGGATGGGCCCGCGTCCC 233
T_scotoductus_SA-01 CCTGTCCTGTGGGGCAGGACTAAAGGGTGAATAGCCCGCTTCCGGATGGGCCCGCGTCCC 228
T_thermophilus_HB27 CCCGCCCCTTGGGGTGTGTCCAAAGG--GCTTTGCCCGCTTCCGGATGGGCCCGCGTCCC 238
Thermus_sp_NMX2_A.1 ATCAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGACGGGTAGCCGGTCTGAGAGG 293
T_scotoductus_SA-01 ATCAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGACGGGTAGCCGGTCTGAGAGG 288
T_thermophilus_HB27 ATCAGCTAGTTGGTGGGGTAATGGCCCACCAAGGCGACGACGGGTAGCCGGTCTGAGAGG 298
Thermus_sp_NMX2_A.1 ATGGCCGGCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGG 353
T_scotoductus_SA-01 ATGGCCGGCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGG 348
T_thermophilus_HB27 ATGGCCGGCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGG 358
Thermus_sp_NMX2_A.1 AATCTTCCGCAATGGACGGAAGTCTGACGGAGCGACGCCGCTTGGAGGAGGAAGCCCTTC 413
T_scotoductus_SA-01 AATCTTCCGCAATGGACGGAAGTCTGACGGAGCGACGCCGCTTGGAGGAGGAAGCCCTTC 408
T_thermophilus_HB27 AATCTTCCGCAATGGGCGCAAGCCTGACGGAGCGACGCCGCTTGGAGGAAGAAGCCCTTC 418
Thermus_sp_NMX2_A.1 GGGGTGTAAACTCCTGAACTGGGGACGAAAGCCCTGTGTAGGGGGATGACGGTACCCAGG 473
T_scotoductus_SA-01 GGGGTGTAAACTCCTGAACTGGGGACGAAAGCCCCGTGTAGGGGGATGACGGTAACCAGG 468
T_thermophilus_HB27 GGGGTGTAAACTCCTGAACCCGGGACGAAACCCCCGAGGAGGGGACTGACGGTACCGGGG 478
Thermus_sp_NMX2_A.1 TAATAGCGCCGGCCAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGCGCGAGCGTTAC 533
T_scotoductus_SA-01 TAATAGCGCCGGCCAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGCGCGAGCGTTAC 528
T_thermophilus_HB27 TAATAGCGCCGGCCAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGCGCGAGCGTTAC 538
Thermus_sp_NMX2_A.1 CCGGATTTACTGGGCGTAAAGGGCGTGTAGGCGGCCTGGGGCGTCCCATGTGAAAGGCCA 593
T_scotoductus_SA-01 CCGGATTTACTGGGCGTAAAGGGCGTGTAGGCGGCCTGAGGCGTCCCATGTGAAAGGCCA 588
T_thermophilus_HB27 CCGGATTCACTGGGCGTAAAGGGCGTGTAGGCGGCCTGGGGCGTCCCATGTGAAAGACCA 598
Thermus_sp_NMX2_A.1 CGGCTCAACCGTGGAGGAGCGTGGGATACGCTCAGGCTAGAGGGTGGGAGAGGGTGGTGG 653
T_scotoductus_SA-01 CGGCTCAACCGTGGAGGAGCGTGGGATACGCTCAGGCTAGAGGGTGGGAGAGGGTGGTGG 648
T_thermophilus_HB27 CGGCTCAACCGTGGGGGAGCGTGGGATACGCTCAGGCTAGACGGTGGGAGAGGGTGGTGG 658
Thermus_sp_NMX2_A.1 AATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAAGGCAG 713
T_scotoductus_SA-01 AATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAAGGCAG 708
T_thermophilus_HB27 AATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAAGGCAG 718
Thermus_sp_NMX2_A.1 CCACCTGGTCCACTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGATA 773
T_scotoductus_SA-01 CCACCTGGTCCACTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGATA 768
T_thermophilus_HB27 CCACCTGGTCCACCCGTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGATA 778
Thermus_sp_NMX2_A.1 CCCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTCTGGG-TTTATCTGGGGGCCG 832
T_scotoductus_SA-01 CCCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTTTGGGGTTTACCTGGGGGCCG 828
T_thermophilus_HB27 CCCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTCTGGG--TCTCCTGGGGGCCG 836
Thermus_sp_NMX2_A.1 AAGCCAACGCGTTAAGCGCGCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGA 892

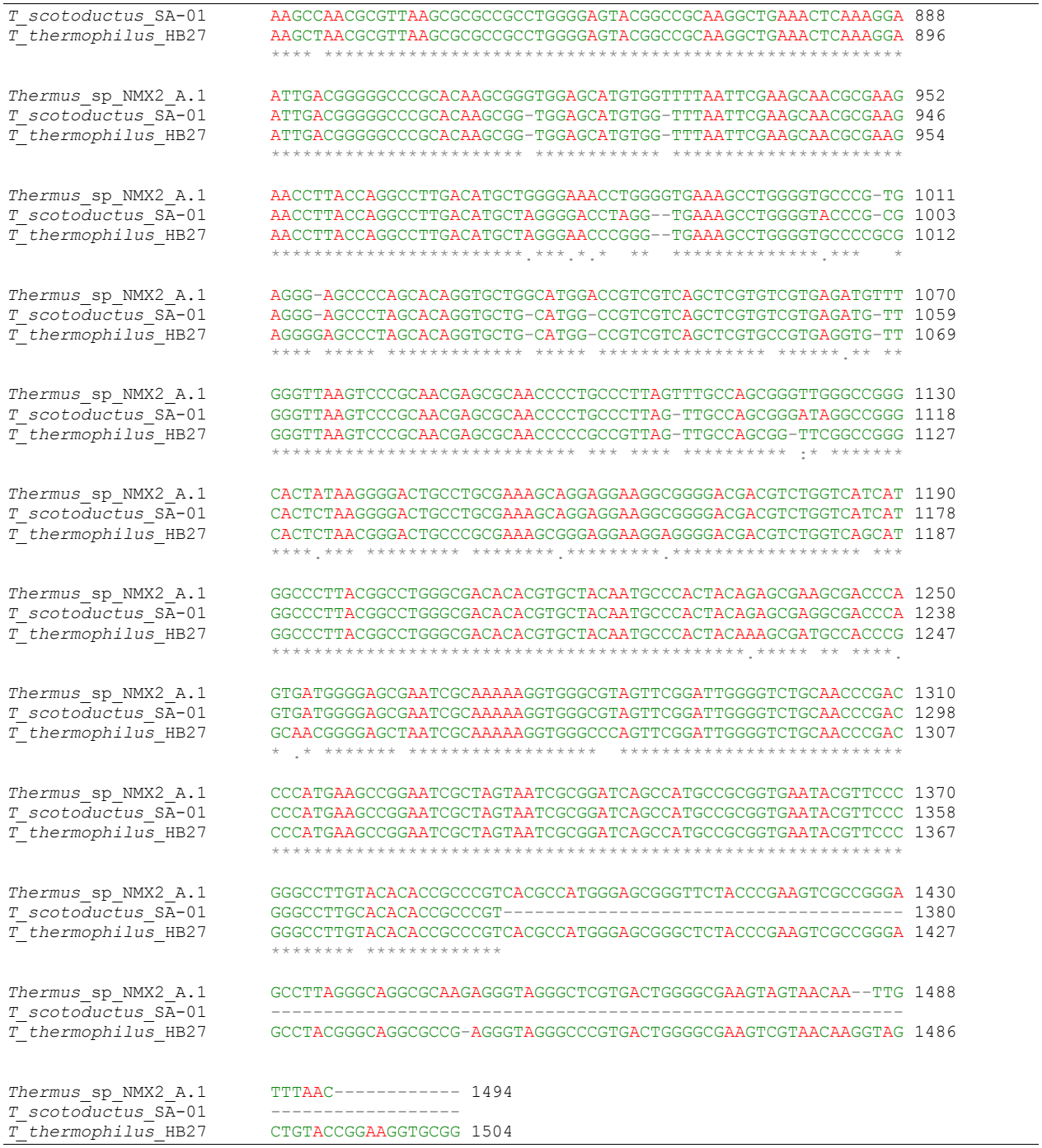


Figure 4.5: Multiple sequence (16S rRNA) alignment of three *Thermus* strains.

```
T_scotoductus_SA-01      -----AAGAGTTTGATCGTGGCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCAAG 55
T_NMX2_A.1               TCGATTAGAGTTTGATCCTGGCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCAAG 60
                              :********** ********************************************

T_scotoductus_SA-01      TCGAGCGGGGCAGGTTTATACCTGTCCAGCGGCGGACGGGTGAGTAACGCGTGGGTGACC 115
T_NMX2_A.1               TCGAGCGGGGCAGGTTTATACCTGTTCAGCGGCAGACGGGTGAGTAACGCGTGGGTGACC 120
                         ************************ *******.**************************

T_scotoductus_SA-01      TACCCGGAAGAGGCGGACAACCTGGGGAAACCCAGGCTAATCCGCCATGTGGTCCTGTCC 175
T_NMX2_A.1               TACCCGGAAGAGGCGGACAACCTGGGGAAACCCAGGCTAATCCGCCATGTGGTCCTGTCC 180
                         ************************************************************

T_scotoductus_SA-01      TGTGGGGCAGGACTAAAGGGTGAATAGCCCGCTTCCGGATGGGCCCGCGTCCCATCAGCT 235
T_NMX2_A.1               TGTGGGGCAGGACTAAAGGGTGGATAGCCCGCTTCCGGATGGGCCCGCGTCCCATCAGCT 240
                         **********************.*************************************

T_scotoductus_SA-01      AGTTGGTGGGGTAAAGGCCCACCAAGGCGACGACGGGTAGCCGGTCTGAGAGGATGGCCG 295
T_NMX2_A.1               AGTTGGTGGGGTAAAGGCCCACCAAGGCGACGACGGGTAGCCGGTCTGAGAGGATGGCCG 300
                         ************************************************************

T_scotoductus_SA-01      GCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGGAATCTTC 355
T_NMX2_A.1               GCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGGAATCTTC 360
                         ************************************************************

T_scotoductus_SA-01      CGCAATGGACGGAAGTCTGACGGAGCGACGCCGCTTGGAGGAGGAAGCCCTTCGGGGTGT 415
T_NMX2_A.1               CGCAATGGACGGAAGTCTGACGGAGCGACGCCGCTTGGAGGAGGAAGCCCTTCGGGGTGT 420
                         ************************************************************

T_scotoductus_SA-01      AAACTCCT-GAACTGGGGACGAAAGCCCCGTGTAGGGGGATGACGGTAACCAGGTA-ATA 473
T_NMX2_A.1               AAACTCCT-GAACTGGGGACGAAAGCCCTGTGTAGGGGGATGACGGTACCCAGGTA-ATA 478
                         ******** ****************** ********************.******* *** 

T_scotoductus_SA-01      GCGCCGGCCAACTCCGTGCCAGCAGCCGCGGTAATACGGAG-GGCGCGAGCGTTACCCGG 532
T_NMX2_A.1               GCGCCGGCCAACTCCGTGCCAGCAGCCGCGGTAATACGGAG-GGCGCGAGCGTTACCCGG 537
                         ****************************************** ******************

T_scotoductus_SA-01      ATTTACT-GGGCGTAAAGGGCGTGTAGGCGGCCTGAGGCGTCCCATGTGAAAGGCCACGG 591
T_NMX2_A.1               ATTTACT-GGGCGTAAAGGGCGTGTAGGCGGCCTGGGGCGTCCCATGTGAAAGGCCACGG 596
                         ******* **************************.************************

T_scotoductus_SA-01      CTCAACCGTGGAGGAGCGTGGGATACGCTCAGGCTAGAGGGTGGGA-GAGGGTGGTGG-A 649
T_NMX2_A.1               CTCAACCGTGGAGGAGCGTGGGATACGCTCAGGCTAGAGGGTGGGA-GAGGGTGGTGG-A 654
                         ********************************************** *********** * 

T_scotoductus_SA-01      ATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAAGGCAGC 709
T_NMX2_A.1               ATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAAGGCAGC 714
                         ************************************************************

T_scotoductus_SA-01      CACCTGGTCCACTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGATAC 769
T_NMX2_A.1               CACCTGGTCCACTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGATAC 774
                         ************************************************************

T_scotoductus_SA-01      CCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTTTGGGGTTTA-----CCTGGGG 824
T_NMX2_A.1               CCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTCTGGGTTTAT------CTGGGG 828
                         ************************************* **** **::      ****** 

T_scotoductus_SA-01      GCCGAAGCCAACGCGTTAAGCGCGCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAA 884
T_NMX2_A.1               GCCGAAGCCAACGCGTTAAGCGCGCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAA 888
                         ************************************************************

T_scotoductus_SA-01      AGGAATTGACGGGGGCCCGCACAAGCGG-TGGAGCATGTGG-TTTAATTCGAAGCAACGC 942
T_NMX2_A.1               AGGAATTGACGGGGGCCCGCACAAGCGGGTGGAGCATGTGGTTTTAATTCGAAGCAACGC 948
                         **************************** ************ ******************

T_scotoductus_SA-01      GAAGAACCTTACCAGGCCTTGACATGCTAGGG-GACCTAGG-TGAAAGCCTGGGGTACCC 1000
T_NMX2_A.1               GAAGAACCTTACCAGGCCTTGACATGCTGGGGAAACCTGGGGTGAAAGCCTGGGGTGCCC 1008
                         ****************************.*** .****.** **************.*** 

T_scotoductus_SA-01      GCGAGGGAGCCCTAGCACAGGTGCTG-CATGG-CCGTCGTCAGCTCGTGTCGTGAGATG- 1057
T_NMX2_A.1               GTGAGGGAGCCCCAGCACAGGTGCTGGCATGGACCGTCGTCAGCTCGTGTCGTGAGATGT 1068
                         * ********** ************* ***** ************************* 
```

```
T_scotoductus_SA-01      TTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCCTTAG-TTGCCAGCGGGATAGGCCG 1116
T_NMX2_A.1               TTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCCTTAGTTTGCCAGCGGGTTGGGCCG 1128
                         ************************************ ***********:*.***** 

T_scotoductus_SA-01      GGCACTCTAAGGGGACTGCCTGC-GAAAGCAGGAGGAAGGCGGGGACGACGTCTGGTCAT 1175
T_NMX2_A.1               GGCACTATAAGGGGACTGCCTGC-GAAAGCAGGAGGAAGGCGGGGACGACGTCTGGTCAT 1187
                         ******.*************** ********************************** 

T_scotoductus_SA-01      CATGGCCCTTACGGCCTGGGCGACACACGTGCTACAATGCCCACTACAGAGCGAGGCGAC 1235
T_NMX2_A.1               CATGGCCCTTACGGCCTGGGCGACACACGTGCTACAATGCCCACTACAGAGCGAAGCGAC 1247
                         ************************************************.***** 

T_scotoductus_SA-01      CCAGTGATGGGGAGCGAATCGCAAAAAGGTGGGCGTAGTTCGGATTGGGGTCTGCAACCC 1295
T_NMX2_A.1               CCAGTGATGGGGAGCGAATCGCAAAAAGGTGGGCGTAGTTCGGATTGGGGTCTGCAACCC 1307
                         ************************************************************ 

T_scotoductus_SA-01      GACCCCATGAAGCCGGAATCGCTAGTAATCGCGGATCAGCCATGCCGCGGTGAATACGTT 1355
T_NMX2_A.1               GACCCCATGAAGCCGGAATCGCTAGTAATCGCGGATCAGCCATGCCGCGGTGAATACGTT 1367
                         ************************************************************ 

T_scotoductus_SA-01      CCCGGGCCTTGCACACACCGCCCGT---------------------------------- 1380
T_NMX2_A.1               CCCGGGCCTTGTACACACCGCCCGTCACGCCATGGGAGCGGGTTCTACCCGAAGTCGCCG 1427
                         *********** ************* 

T_scotoductus_SA-01      ------------------------------------------------------------
T_NMX2_A.1               GGAGCCTTAGGGCAGGCGCAAGAGGGTAGGGCTCGTGACTGGGGCGAAGTAGTAACAATT 1487


T_scotoductus_SA-01      -------
T_NMX2_A.1               GTTTAAC 1494
```

Figure 4.6: Sequence (16S rRNA) alignment of the two closely related metal-reducing *Thermus* strains

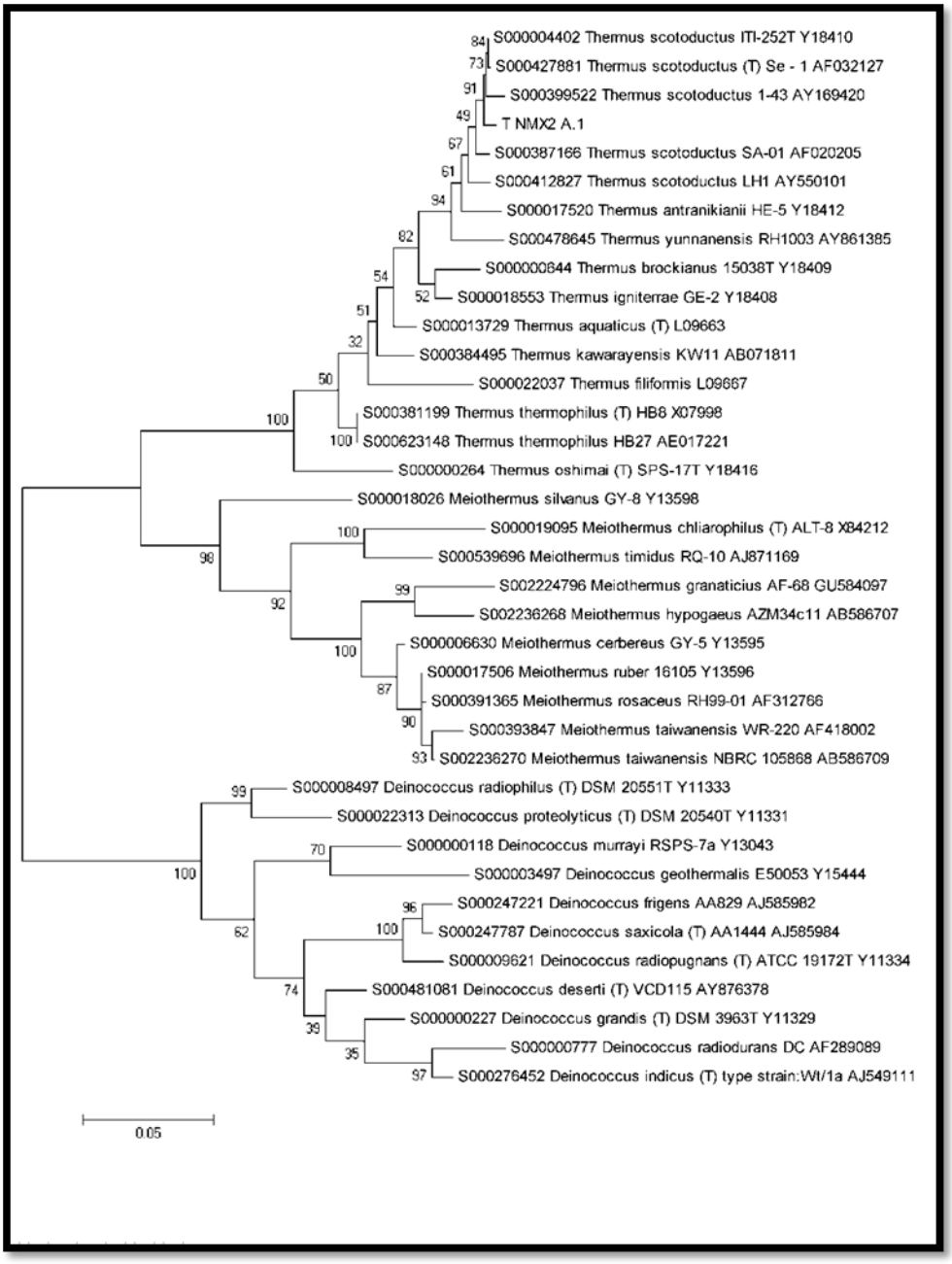


Figure 4.7 (a): Phylogenetic tree based on 16S rRNA gene sequences comparisons of validly described type of species of the Phylum *Deinococcus-Thermus* and close relatives of the genus *Meiothermus*. Construction of dendogram was done from evolutionary distances using the maximum likelihood method. The scale bar represents 0.05 inferred nucleotide changes per 500 nucleotides.

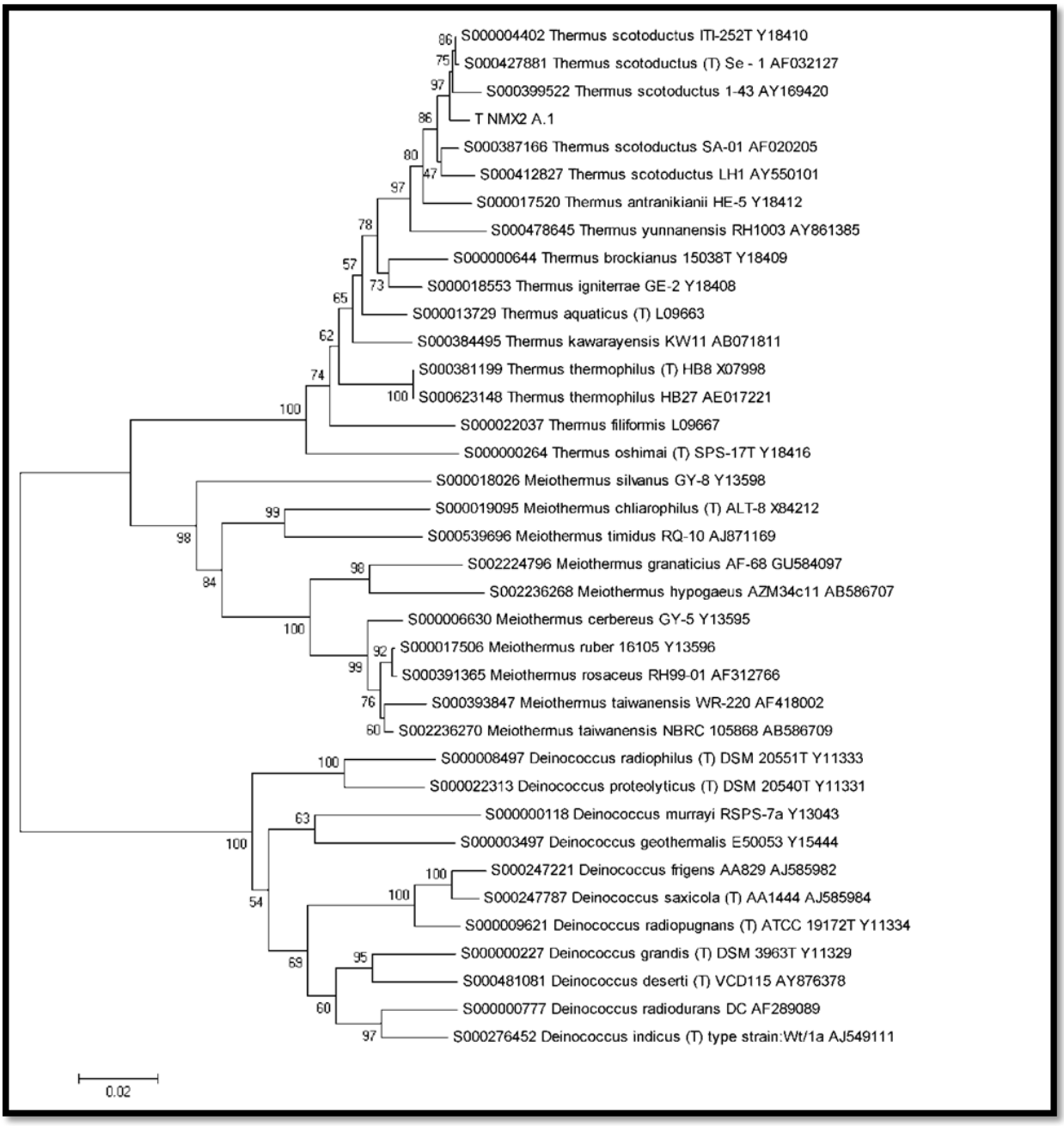


Figure 4.7 (b): Phylogenetic tree based on 16S rRNA gene sequences comparisons of validly described type of species of the Phylum *Deinococcus-Thermus* and close relatives of the genus *Meiothermus*. Construction of dendogram was done from evolutionary distances using the neighbor-joining method. The scale bar represents 0.02 inferred nucleotide changes per 500 nucleotides.

3.3. Anaerobic growth analysis

The growth was monitored as described in (section 2.6). The media was done with varying concentrations of KNO_3 (10 mM, 20 mM and 40 mM) from which the (40 mM) was used to supplement the media. Below in Figure 4.8 the monitoring of growth over a period of 72 h is observed.

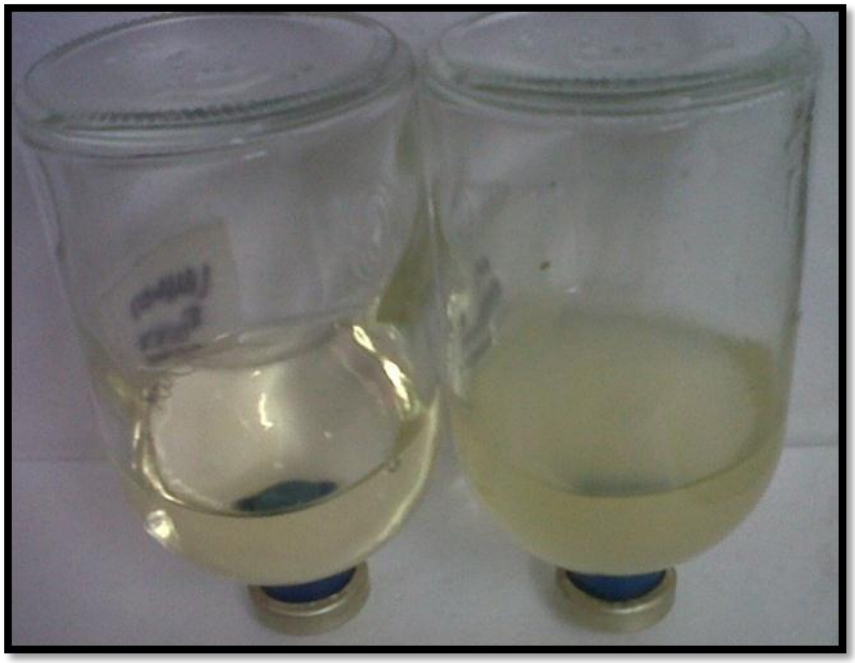


Figure 4.8: Growth monitored over a period of 72 h. Low biomas was obtained under anaerobic conditions,turbidity was measured over time, clearer volume was used as a control and turbid volume was the experimental sample.

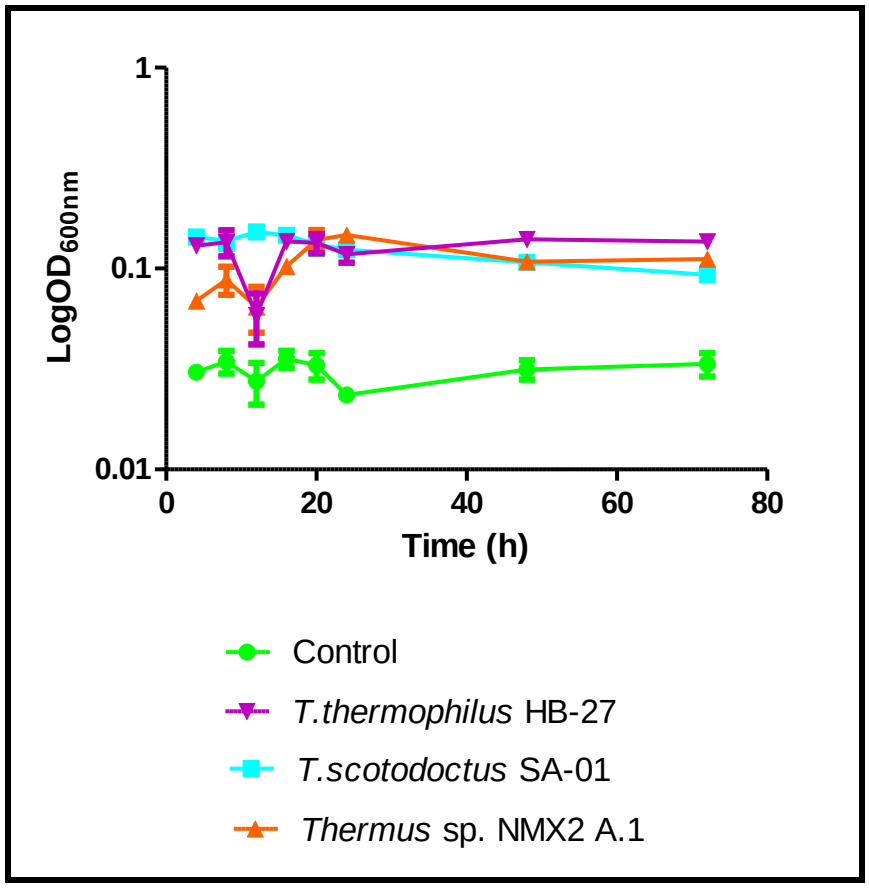


Figure 4.9: Representation of *Thermus* strains ability to grow under anaerobic conditions using KNO_3 as final electron acceptor.

3.4 Nitrate reduction

Comparative functionality studies of the ability to reduce NO_3^- by *T. thermophilus* HB27, *T. scotoductus* SA-01 and *Thermus* sp.NMX2 A.1 strain, as described in section 2.6 is shown in Figure 4.10.

Balkwill and co-workers stated in literature that *T. scotoductus* SA-01 previously isolated in a South African deep (3.2 km) gold mine is closely related to *Thermus* sp. NMX2 A.1 strain previously isolated from Thermal Springs in New Mexico, USA; and that the *Thermus* sp. NMX2 A.1 strain was identified to be a *T. scotoductus*. Both these strains have been shown to use nitrate as terminal electron acceptors (Balkwill *et al*., 2004). Moreover, Gounder and co-workers showed in 2011 that *T. scotoductus* SA-01 contains

genes responsible for carrying out denitrification (Gounder *et al*., 2011), which indicate that the same results may occur in *Thermus* sp. NMX2 A.1 strain.

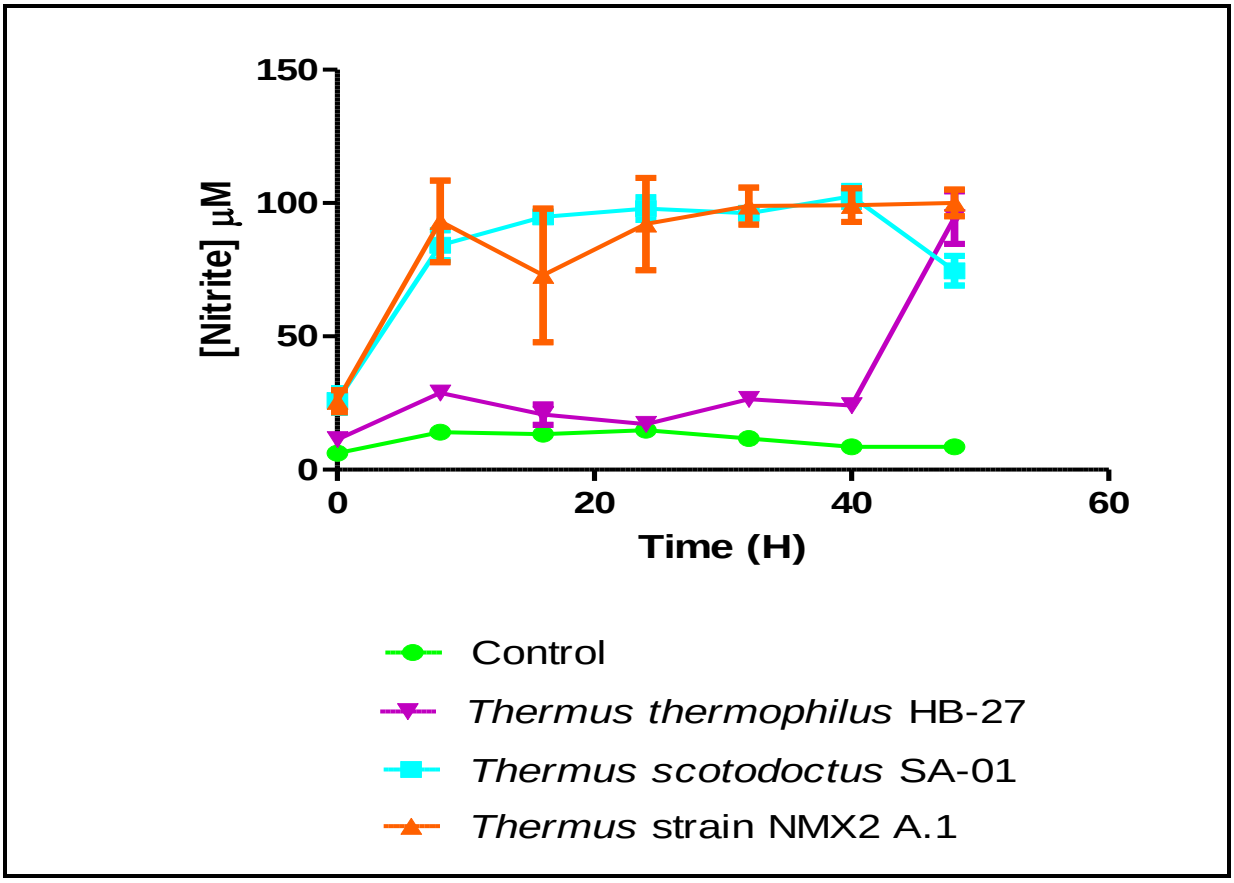


Figure 4.10: A graph of anaerobic growth carried with three *Thermus* strains. The respective media (TYG, *Thermus* Broth and ATTC 697) were supplemented with 10 mM of Potassium nitrate, the *Thermus* strains were then grown to optical densities (600 nm) of 0.6 and then incubation was carried out for 48 hours and nitrite was done over time.

The turnover of NO_3^- to NO_2^- by *T. thermophilus* HB27 is low in comparison to the turnover of *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain. These outcomes are supported by literature of Bricio and co-workers where he mentioned that the sequenced *T. thermophilus* HB27 cannot grow in the absence of oxygen (Bricio *et al*., 2011). Subsequently, both *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain showed a high turnover of NO_3^- to NO_2^- from the beginning of the analysis which indicate that the genes responsible are present hence they got activated to perform their function right from the start.

4. Conclusions

As shown in the previous chapter when comparison was done on genome level great similarities suggested that *Thermus* sp. NMX2 A.1 strain is most likely to be a *T. scotoductus*. Additional supporting data was performed in which the nitrate reduction was carried out, what was expected was to see a similar profile between *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain.

Among the three strains used namely, *T. thermophilus* HB27, *T. scotoductus* SA-01and *Thermus* sp. NMX2 A.1 strain; the phylogenetic results from the BOX-PCR fingerprints, CLUSTAL W alignment as well as the phylogenetic trees indicated similarities between *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain in all three instances.

Moreover, as the nitrite detection was carried out over a period of 72 hours, amongst the three strains, *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain had similar profiles and reduced nitrate from the first 8 hours of incubation; meanwhile *T. thermophilus* HB27 didn't show significant reduction of nitrate, these outcomes are supported by literature of Bricio and co-workers 2011.

It is therefore evident that *T. scotoductus* SA-01 and *Thermus* sp. NMX2 A.1 strain has similar functionality which is a confirmation of the operon comparison in chapter 3.

5. Supplement B

Table 4.2 (a): The 16S rRNA sequence of *Thermus* sp. NMX2 A.1 strain used for identification

tcgattagagtttgatcctggctcagggtgaacgctggcggcgtgcctaagacatgcaagtcgagcggggcaggttta
tacctgttcagcggcagacgggtgagtaacgcgtgggtgacctacccggaagaggcggacaacctggggaaacccag
gctaatccgccatgtggtcctgtcctgtggggcaggactaaagggtggatagcccgcttccggatgggcccgcgtcc
catcagctagttggtggggtaaaggcccaccaaggcgacgacgggtagccggtctgagaggatggccggccacaggg
gcactgagacacgggccccactcctacgggaggcagcagttaggaatcttccgcaatggacggaagtctgacggagc
gacgccgcttggaggaggaagcccttcggggtgtaaactcctgaactggggacgaaagccctgtgtaggggatgac
ggtacccaggtaatagcgccggccaactccgtgccagcagccgcggtaatacggagggcgcgagcgttacccggatt
tactgggcgtaaagggcgtgtaggcggcctggggcgtcccatgtgaaaggccacggctcaaccgtggaggagcgtgg
gatacgctcaggctagagggtgggagagggtggtggaattcccggagtagcggtgaaatgcgcagataccgggagga
acgccgatggcgaaggcagccacctggtccacttctgacgctgaggcgcgaaagcgtggggagcaaaccggattaga
tacccgggtagtccacgccctaaacgatgcgcgctaggtctctgggtttatctgggggccgaagccaacgcgttaag
cgcgccgcctggggagtacggccgcaaggctgaaactcaaaggaattgacgggggcccgcacaagcgggtggagcat
gtggttttaattcgaagcaacgcgaagaaccttaccaggccttgacatgctggggaaacctggggtgaaagcctggg
gtgcccgtgagggagccccagcacaggtgctggcatggaccgtcgtcagctcgtgtcgtgagatgtttgggttaagt
cccgcaacgagcgcaacccctgcccttagtttgccagcgggttgggccgggcactataaggggactgcctgcgaaag
caggaggaaggcggggacgacgtctggtcatcatggcccttacggcctgggcgacacacgtgctacaatgcccacta
cagagcgaagcgacccagtgatggggagcgaatcgcaaaaaggtgggcgtagttcggattggggtctgcaacccgac
cccatgaagccggaatcgctagtaatcgcggatcagccatgccgcggtgaatacgttcccgggccttgtacacaccg
cccgtcacgccatgggagcgggttctacccgaagtcgccgggagccttagggcaggcgcaagagggtagggctcgtg
actggggcgaagtagtaacaattgtttaac

Table 4.2 (b): Obtained results after BLAST analysis of the 16S rRNA sequence of the *Thermus* sp. NMX2 A.1 strain

Accession	Description	Max score	Total score	Query coverage	E-value	Max indent
L09661.1	*Thermus* sp. NMX2 A.1 16S ribosomal RNA (rna) gene sequence	822	822	99%	0.0	99%
AF255590.1	*Thermus* sp. SRI-96 16S ribosomal RNA gene, partial sequence	817	817	99%	0.0	99%
DQ861411.1	Uncultured *Thermus* sp. clone c108 16S ribosomal RNA gene, partial sequence	811	811	99%	0.0	98%
DQ414817.1	*Thermus* sp. GD31 16S ribosomal RNA gene, partial sequence	811	811	99%	0.0	98%
NR_024867.1	*Thermus scotoductus* strain Se-1 16S ribosomal RNA, partial sequence >gb|AF032127.1| *Thermus scotoductus* 16S ribosomal RNA gene, complete sequence	808	808	99%	0.0	98%

Table 4.3 (a): The 16S rRNA sequence of *Thermus scotoductus* SA-01 strain used for identification

```
tcgattggttaccttgttactacttcgccccagtcacgagccctaccctcggcgcctgccctaaggctcccggcgac
ttcgggtagaacccgctcccatggcgtgacgggcggtgtgtacaaggcccgggaacgtattcaccgcggcatggctg
atccgcgattactagcgattccggcttcatggggtcgggttgcagaccccaatccgaactacgcccacctttttgcg
attcgctccccatcactgggtcgcctcgctctgtagtgggcattgtagcacgtgtgtcgcccaggccgtaagggcca
tgatgaccagacgtcgtccccgccttcctcctgctttcgcaggcagtccccttagagtgcccggcctatcccgctgg
caactaagggcaggggttgcgctcgttgcgggacttaacccaacatctcacgacacgagctgacgacggccatgcag
cacctgtgctagggctccctcgcgggtaccccaggctttcacctaggtcccctagcatgtcaaggcctggtaaggtt
cttcgcgttgcttcgaattaaaccacatgctccaccgcttgtgcgggcccccgtcaattcctttgagtttcagcctt
gcggccgtactccccaggcggcgcgcttaacgcgttggcttcggcccccaggtaaaccccaaagacctagcgcgcat
cgtttagggcgtggactacccgggtatctaatccggtttgctccccacgctttcgcgcctcagcgtcagaagtggac
caggtggctgccttcgccatcggcgttcctcccggtatctgcgcatttcaccgctacttcgggaattccaccaccct
ctcccaccctctagcctgagcgtatcccacgctcctccacggttgagccgcggccctttcacatgggacgccccagg
ccgccctacacgccctttacgcccagtaaatccggg
```

Table 4.3 (b): Obtained results after BLAST analysis of the 16S rRNA sequence of the *Thermus scotoductus* SA-01 strain

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
AF020205.1	*Thermus scotoductus* SA-01 16S ribosomal RNA gene, partial sequence	2542	2542	100%	0.0	99%
CP001962.1	*Thermus scotoductus* SA-01, complete genome	2518	5028	99%	0.0	99%
EU330195.1	*Thermus scotoductus* strain SA-01 16S ribosomal RNA gene, partial sequence	2518	2518	99%	0.0	99%
AB661716.1	*Thermus* sp. IC-174 gene for 16S rRNA, partial sequence	2464	2464	97%	0.0	99%
AB183863.1	Uncultured bacterium gene for 16S ribosomal RNA, partial sequence, clone: GAB-B07	2462	2462	98%	0.0	99%
AF255590.1	*Thermus* sp. SRI-96 16S ribosomal RNA gene, partial sequence	2459	2459	98%	0.0	99%

Table 4.4 (a): The 16S rRNA sequence of *Thermus thermophilus* HB27 strain used for identification

```
aaagttaccttgttactacttcgccccagtcacgggccctaccctcggcgcctgcccgtaggctcccggcgacttcg
ggtagagcccgctcccatggcgtgacgggcggtgtgtacaaggcccgggaacgtattcaccgcggcatggctgatcc
gcgattactagcgattccggcttcatggggtcgggttgcagaccccaatccgaactgggcccacctttttgcgatta
gctccccgttgccgggtggcatcgctttgtagtgggcattgtagcacgtgtgtcgcccaggccgtaagggccatgct
gaccagacgtcgtcccctccttcctcccgctttcgcgggcagtcccgttagagtgcccggccgaaccgctggcaact
aacggcgggggttgcgctcgttgcgggacttaacccaacacctcacggcacgagctgacgacggccatgcagcacct
gtgctagggctcccctcgcggggcaccccaggctttcacccgggttccctagcatgtcaaggcctggtaaggttctt
cgcgttgcttcgaattaaaccacatgctccaccgcttgtgcgggcccccgtcaattcctttgagtttcagccttgcg
gccgtactccccaggcggcgcgcttaacgcgttagcttcggcccccaggagacccagagacctagcgcgcatcgttt
agggcgtggactacccgggtatctaatccggtttgctccccacgctttcgcgcctcagcgtcacgggtggaccaggt
ggctgccttcgccatcggcgttcctcccggtatctgcgcatttcaccgctactccgggaattccaccaccctctccc
accgtctagcctgagcgtatcccacgctccccacggttgagccgtggtctttcacatgggacgccccaggccgcct
acacgccctttacgcccagtgaatccgggtaacgctcgcgccctccgtattaccgcggctgctggcacggagttggc
cggcgctattaccccggtaccgtcagtcccctcctcgggggtttcgtcccgggttcaggagtttacaccccgaaggg
cttcttcctccaagcggcgtcgctccgtcaggcttgcgcccattgcggaagattcccaactgctgcctcccgtagga
gtggggcccgtgtctcagtgcccctgtggccggccatcctctcagaccggctacccgtcgtcgccttggtgggccat
taccccaccaactagctgatgggacgcgggcccatccggaagcgggcaaagccctttggacacaccccaaggggcgg
gtccacatgggggattagcccgagtttccccgggttgtccccctcttccgggtaggtcacccacgcgttactcaccc
gtccgccgctgaccacggagtaaaacccggcggcccgcacgacttgc
```

Table 4.4 (b): Obtained results after BLAST analysis of the 16S rRNA sequence of the *Thermus thermophilus* HB27 strain

Accession	Description	Max score	Total score	Query coverage	E-value	Max indent
AE017221.1	*Thermus thermophilus* HB27, complete genome	2621	5243	99%	0.0	99%
L09659.1	*Thermus thermophilus* HB27 16S ribosomal RNA gene sequence	2621	2621	99%	0.0	99%
L09660.1	*Thermus flavus* AT-62 16S ribosomal RNA (rnn) gene sequence >gb|EU214608.1| *Thermus thermophilus* strain mc-13 16S ribosomal RNA gene, partial sequence	2621	2621	99%	0.0	99%
EU652052.1	*Thermus thermophilus* strain it-1 16S ribosomal RNA gene, partial sequence	2619	2619	99%	0.0	99%
DQ087525.2	*Thermus* sp. E26 16S ribosomal RNA gene, partial sequence	2619	2619	99%	0.0	99%

6. References

Balkwill, D. L., Keift, T. L., Tsukuda, T., Kostandarithes, H. M., Onstott, T. C., Macnaughton, S., Bownas, J. and Fredrickson, J. K. (2004) Identification of iron-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles*, **8,** 347-44.

Bricio, C., Alvarez, L., Gómez, M. J. and Berenguer, J. (2011) Partial and complete denitrification in *Thermus thermophilus*: lessons from genome drafts. *Biochem. Soc. Trans,* **39,** 249-253.

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

Griess, P. (1879) *Chem. Ber*, **v. 12.** 426.

Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T (Jr)., Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M. and Tiedje, J. M. (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res,* **298,** 173-174.

Mesa, S., Göttfert M. and Bedmar E. J. (2001) Thenir, nor, and nos denitrification genes are dispersed over the *Bradyrhizobium japonicum* chromosome. *Arch Microbiol,* **176,** 136-142.

Ramírez, S., Moreno, R., Zafra, O., Castán, P., Vallès, C and Berenguer, J. (2000) Two nitrate/nitrite transporters are encoded within the mobilizable plasmid for nitrate respiration of *Thermus thermophilus* HB8. *J. of. Bacter,* p. 2179-2183.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) **MEGA5:** Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods**.** *Mol. Biol. and Evol,* **28,** 2731-2739.

Thompson, J. D., Higgens, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* **22,** 4673-4680.

CHAPTER 5

Microbial isolation

1. Background to this chapter

The initial chapters allowed comparison between known *Thermus* sp. and their genomes, as well as functionality. Several phylogenetic studies of South African deep mine environments as well as the isolation of *T. scotoductus* SA-01 suggest that novel *Thermus* sp. could be part of the subsurface biome. This chapter deals with attempts to isolate new *Thermus* sp. to extend our knowledge on metabolic capabilities of the genus *Thermus*.

Previous studies done by Abhita Judgave (PhD) on a fissure water sample (NO212FW050508) isolated from the Northam platinum mine of South Africa at (1.9 km) below surface showed a consortium that revealed a possibility in a isolation that contained at least one *Thermus* sp. that aligned closely to a previous sequence of *Thermus* sp. Tibetan G-7.

The consortium in a form of glycerol stock was retrieved for the isolation of this *Thermus* sp. Two sets of media where used for the attempted enrichments at a pH of 7.6 shown in (Table 5.1), and an optimal temperature of 65°C was decided on for incubation.

Table 5.1: Media used to enrich the consortium from the Northam platinum mine

Heterotrophic (HB)	**Tryptone, Yeast extract and Glucose (TYG)**
0.1 g/L glucose	5 g/L tryptone
0.1 g/L yeast extract	3 g/L yeast extract
0.05 g/L peptone	1 g/L glucose
0.05 g/L tryptone	16 g/L agar
0.6 g/L $MgSO_4.7H_2O$	
0.07 g/L $CaCl_2.2H_2O$	
0.1 g/L MOPS	
16 g/L agar	

Microbial isolation

Additional deep gold mine samples namely (KI445FW190711), (BE326FW071211), (Dr51PC150711) and (TT107FW081111) were selected to extend attempts of *Thermus* sp. isolation since the phylum *Deinococcus- Thermus* was previously found in these environments (Kieft *et al*., 1999) the (Table 5.2 a) below gives description of the samples.

Table 5.2 (a): Sample names; parameters and depth of the site as well their respective owners

Sample name	Depth below surface	Parameters	Owned by
TT107FW081111	3.1 km	Temp: 52.1ºC, pH: 8.57,	Ashanti Gold
BE326FW071211	1.34 km	Temp: 31.4ºC, pH: 8.68,	Sibanye Gold
KI445FW190711	3.28 km	Temp: 54.5ºC, pH: 8.0,	Sibanye Gold
Dr51PC150711	1.2 km	Temp: 26.7ºC, pH: 7.32	Sibanye Gold
NO212FW050508	1.9 km	Temp: 53.7ºC, pH: 9.30	Northam Platinum Limited

TT107FW081111> **TT:** Tau-tona; **107:** Level; **FW:** Fissure Water; **081111:** Date.

BE326FW071211> **BE:** Beatrix; **326**: Shaft 3 Level 26; **FW:** Fissure Water; **071211:** Date.

KI445FW190711> **KI:** Kloof; **445:** Shaft 4 Level 45; **FW:** Fissure Water; **190711:** Date.

Dr51PC150711> **Dr:** Driefontein; **51PC:** Shaft 51 Chamber; **150711:** Date.

NO212FW050508> **NO:** Northam; **212**: Shaft 2 Level 12; **FW:** Fissure Water; **050508**: Date.

Microbial isolation

1.1 The aim of the chapter

The aim of this chapter was to attempt to isolate new *Thermus* spp. from various fissure water samples. Since all sites and initial enrichments indicated phylogenetically that the phylum *Deinococcus- Thermus* was present in these environments and therefore the likelihood of new *Thermus* strains was explored.

2. Materials and Methods

2.1 Revival of cells

Glycerol stocks were obtained from Abhita Judgave (PhD), from the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State; which had shown from the BLAST results to contain at least a *Thermus* sp. The samples were plated out on various agar plates (Table 5.2 b) at pH 7.6 followed by incubation at 65°C for three days.

Table 5.2 (b): Media used to enrich the samples and to evaluate where the most growth occurs

Nutrient Broth (NB)	**Luria- Bertani (LB)**	**Tryptone, Yeast extract and Glucose (TYG)**	****Violet Red Bile Agar (VRBA)**	**Heterotrophic (HB)**
1 g/L meat extract	10 g/L tryptone	5 g/L tryptone	7 g/L peptone	0.1 g/L glucose
2 g/L yeast extract	5 g/L yeast extract	3 g/L yeast extract	3 g/L yeast extract	0.1 g/L yeast extract
5 g/L peptone	10 g/L NaCl	1 g/L glucose	10 g/L lactose	0.05 g/L peptone
8 g/L sodium chloride	16 g/L agar	16 g/L agar	1.5 g/L bile salts NO.3	0.05 g/L tryptone
16 g/L agar			5 g/L sodium chloride	0.6 g/L $MgSO_4.7H_2O$
			0.03 g/L neutral red	0.07 g/L $CaCl_2.2H_2O$
			0.002 g/L crystal violet	0.1 g/L MOPS
			12 g/L agar	16 g/L agar

Rich media for proliferation and VRBA selection media for Gram negative isolation as well minimal media.

Colonies were re-plated to retrieve pure cultures as many times as required. Single colonies were then incubated into liquid media (50 mL) and allowed to grow at 65°C with agitation (200 rpm). In addition, pure cultures were used for Gram staining and evaluated microscopically at 1000x magnification (Bergey *et al*., 1994).

2.2 Genomic DNA extraction

Genomic DNA extraction, quantity and purity of the genomic DNA were performed according to (section 2.2) of chapter 3.

2.3 PCR amplification of the 16S rRNA gene

The identity of the consortium and enriched isolates were confirmed by amplification and sequencing of the 16S rRNA gene from the genomic DNA using the universal bacterial primers 27F and 1492R as described in (section 2.3) of chapter 3.

2.4 Cloning of the 16S rRNA gene into pGEM-T Easy vector and sequencing

The purified products were ligated into pGEM-T Easy vector as described in (section 2.4) of chapter 3.

2.5 Fissure water "Bio" samples from deep gold mines of South Africa

Samples from deep gold mines of South Africa shown in (Table 5.2) where sampled to perform the isolation of *Thermus* sp. The samples were stained to confirm the presence of bacterial cells and obtain an estimate of the numbers of cells present. A volume of 1 mL from each of the fissure water samples was mixed with 108 µL of 40% (v/v) formaldehyde and incubated at 4°C for 2 h. Subsequently, the mixture was filtered using 0.2 µm membrane filter and then placed onto a clean petri-dish with the cells on the upper surface. A portion of the membrane filter was cut using a sterile scalpel blade and the filter pieces were placed facing down onto 10 µL of 4', 6-diamidino-2-phenylindole (DAPI) staining solution with a concentration of 10 µg/mL. It was left for 2 min covered with aluminum foil to protect the membrane from light.

A volume of 10 µL citifluor (Citiflour Ltd, London) was added onto a clean microscopic slide followed by placing the stained portion of the membrane filter onto the slide, after covering the slide with a cover slip, visualization was done under a fluorescence

microscope equipped with the appropriate filter set (Porter and Feig, 1980). Another 1 mL aliquot was filtered using a 0.2 µm membrane filter to concentrate the cells and grown on agar plates of TYG media supplemented with the components listed in (Table 5.3 (a).

Table 5.3 (a): Supplements used for the TYG media

0.5 ml/L of trace elements
12.8 g/L Nitrilotriacetic acid
1 g/L $FeCl_2.4H_2O$
0.5 g/L $MnCl_2.4H_2O$
0.3 g/L $CoCl_2.4H_2O$
50 mg/L $CuCl_2.2H_2O$
50 mg/L $Na_2MoO_4.2H_2O$
20 mg/L H_3BO_3 and 20 mg/L $NiCl_2.6H_2O$

Modified *Thermus* 162 medium (DSMZ) was simultaneously used for incubation of the samples to target specifically *Thermus* sp., (Table 5.3 b). The media pH was adjusted to 7.6 and incubation was done at 65°C and 70°C for a period of 2-3 days.

Table 5.3 (b): Modified *Thermus* 162 medium (DSMZ)

Modified *Thermus* 162 medium	**100 ml/L of phosphate buffer pH 7.2**
2.5 g/L yeast extract	5.44 g/L KH_2PO_4
2.5 g/L tryptone	43 g/L $Na_2HPO_4.12H_2O$
28 g/L agar	
100 mg/L Nitrilotriacetic acid	
40 mg/L $CaSO_4.2H_2O$	
200 mg/L $MgCl_2.6H_2O$	
0.5 ml/L of 0.01 M Fe citrate	
0.5 ml/L of trace elements	
100 ml/L of phosphate buffer	

Once growth was obtained, re-plating was done as described in (section 2.1) of this chapter as well as the Gram staining procedures. This was followed by genomic DNA extraction, PCR amplification of 16S rRNA gene and cloning of the purified 16S rRNA gene PCR product into pGEM-T Easy vector as described in (section 2.2-2.4) of chapter 3.

2.6 Isolation using selective media components

Additions of selective components where used to select specifically for Gram negative bacteria as shown in (Table 5.4). In addition to the media that enhanced the growth of selected cells, specific reagents namely crystal violet and bile salts which inhibit Gram positive bacteria where used and grown at 65°C and 70°C for a maximum of 3 days.

Table 5.4: Selective components used to enhance the growth of the Gram negative bacteria

Mac-Conkey agar pH of 7.6	Eosin-methylene blue (EMB) agar pH of 7.6	Combination 1 pH of 7.6	Combination 2 pH of 7.6	Combination 3 pH of 7.6
5 g/L tryptone	5 g/L tryptone	EMB media	EMB media	EMB media
3 g/L yeast extract	3 g/L yeast extract	0.002 g/L crystal violet	0.002 g/L crystal violet	1.5 g/L bile salts
1 g/L glucose	1 g/L glucose			
0.002 g/L crystal violet	0.065 g/L methylene blue	1.5 g/L bile salts		
1.5 g/L bile salts				

Once growth was obtained, re-plating was done as described in (section 2.1) of chapter 5 as well as the Gram staining procedures. This was followed by genomic DNA extraction, PCR amplification of 16S rRNA gene and cloning of the purified 16S rRNA gene PCR product into pGEM-T Easy vector as described in (section 2.2-2.4) of chapter 3.

2.7 Isolation with application of antibiotics

Antibiotics where also used as an alternative to selective components. Literature suggest that the genus *Thermus* are heterotrophic thermophilic Gram negative bacteria that grow aerobically with high growth rates and good yields on complex medium at an optimum growth temperature of 62 °C to 75 °C, and they do not require specific amino

acids or vitamins (Zhang et al., 2010). One can increase the chances of Gram negative isolation by using antibiotics which are effective in eliminating Gram positive bacteria.

A concentration of 25 µg/mL of penicillin was used to supplement TYG media, the pH was adjusted to 7.6 and inocula were grown at 65°C and 70°C for a day. Also, as no growth control; ampicillin at a concentration of 30 µg/mL was used as a control since it is known to be effective against both Gram negative as well as Gram positive bacteria. Single colonies obtained were then diluted in 50 µL of 1% (w/v) physiological salt solution and again streaked out onto TYG plates supplemented with antibiotics followed by the same incubation procedures.

Once growth was obtained, re-plating was done as described in (section 2.1) of chapter 5 as well as the Gram staining procedures. This was followed by genomic DNA extraction, PCR amplification of 16S rRNA gene and cloning of the purified 16S rRNA gene PCR product into pGEM-T Easy vector as described in (section 2.2-2.4) of chapter 3.

2.8 Denaturing Gradient Gel Electrophoresis (DGGE) analysis

In addition, since no effective isolation attempts yield the desired micro-organisms, two initial consortium enrichments (NO212FW050508) and (TT107FW081111) were incubated anaerobically with TYG media supplemented with 10 mM of KNO_3, Fe (NTA), Lactate and Acetate respectively at 65°C. Genomic DNA was then extracted using the NucleoSpin® Soil kit according to the manufacturer's instructions. The evaluation of the extracted gDNA was done as described in (section 2.2) of chapter 3.

2.8.1 Nested DGGE-PCR Amplification

A small portion of the 16S rRNA was amplified using set of primers described in (Table 5.5) with the extracted gDNA products as templates. Each nested PCR reaction contained 1 µL template, 1 µL dNTPs (10 mM), 1 µL of each primer (10 µM), 1 µL BSA (10 µg/µL), 5 µL 10X Buffer and 0.25 µL (5 U/µL) of *Taq* DNA polymerase from New

England BioLabs. The volume was adjusted to a final volume of 50 µL with sterile milliQ water.

Table 5.5: Primer sequences for nested-DGGE profiling

Primers	Length	Sequence of primers	Template region
341F-GC	341–357	5′-CCT ACG GGA GGC AGC AG-3′	V3 high variable region
907R	907–926	5′-CCG TCA ATT CMT TTGAGT TT-3′	626 bp

F, Forward primer; R, Reverse primer; GC was a 40-base GC clamp, 5'-CGC CCGCCGCGC CCC GCGCCC GTC CCGCCGCCC CCGCCC-3' (Taken from Wang *et al*., 2008)

For each nested PCR product (5 µl) was added to 2 µl of the loading dye (Fermentas) before loading on the 1% (w/v) agarose gel for evaluation under UV light. Thereafter re-amplification was performed using primer sets in (Table 2.9.1) without the GC-clamp followed by sequencing and identification using nBLAST alogarithm at NCBI.

2.8.2 Preparation of Denaturing Gradient Gel

The prepared stock solutions included, 0% Urea-Formamide [40% acrylamide/bis (10 ml), 50X TAE (1 ml) and ultra-pure water (39 ml)] and 80% Urea-Formamide [40% acrylamide/bis (10 ml), 50X TAE (1 ml), formamide (16 ml), urea (16.8 g) and filled with ultra-pure water to a volume of 50 ml]. The gradient used was 40% UF [0% stock solution (5 ml) and 80% stock solution (5 ml)] and 60% [0% stock solution (2.5 ml) and 80% stock solution (7.5 ml)]. APS (63 µl) and TEMED (7 µl) were added to each gradient solution.

The 8% acrylamide/bis gel was cast, after gel polymerization, the inserted comb removed and the polymerized gel released from the casting stand and placed into the pre-heated buffer [140 ml filtered 50x TAE]. The remaining nested-PCR amplicons were subjected to denaturing gradient gel at a voltage of 100 V for 16 hours at 60^{o}C. The gel was stained with ethidium bromide for 15 min, washed with distilled water and visualized under UV light. Different bands obtained were excised, each band was gel

eluted with autoclaved double distilled water (50 µL) in a 1.5 mL tube and incubated overnight at 55°C. The gel eluted bands were re-amplified using primers set up in (Table 5.6) without the GC-clamp.

3. Results and discussions

3.1 Cell morphology

Microscopic analyses of the stained isolates showed that the cultures were not homogeneous. The Gram stain procedure (section 2.1) showed both Gram positive and Gram negative bacteria with rod shape morphology presented in (Figure 5.1). It was also noted that for different media different morphologies where observed some had longer cells in comparison to one another.

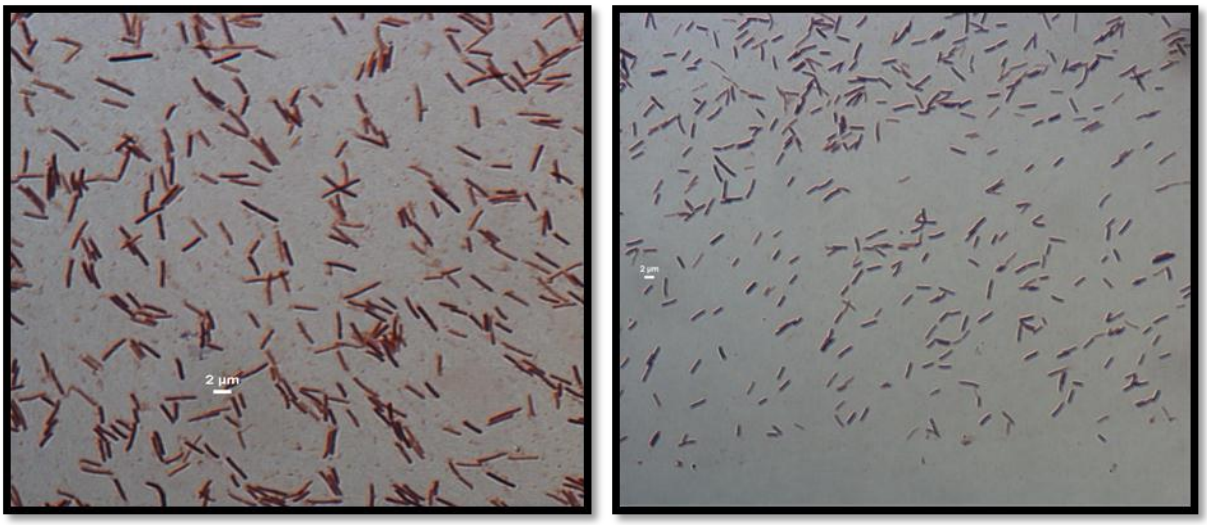


Figure 5.1: The isolates morphology from all the biological isolation attempts indicated both Gram positive and Gram negative rod shaped cells. The scale bar is indicated in each figure as 2 µm.

3.2 Genomic DNA extraction and PCR amplification of the 16S rRNA

Even though no single morphology was observed, genomic DNA was isolated from the consortium samples; (Table 5.6) indicates the successful DNA yields from consortium enrichments. The obtained bands showed high integrity DNA and almost no shearing, indicated in (Figure 5.2). The A_{260}/A_{280} absorbance ratio of 1.8 indicated low protein or RNA contamination.

Table 5.6: Genomic DNA concentrations from consortium samples

Consortium samples	Application of various media (ng/ul)	Selective media components (ng/ul)	Application of antibiotics (ng/ul)
NO212FW050508	555.18	1322.39	1009.11
NO212FW050508	458.23	488.91	685.30
TT107FW081111	300.26	286.32	465.27
BE326FW071211	1057.05	558.41	500.95
KI445FW190711	1866 61	2729.76	1658.32
Dr51PC150711	806.01	666.54	379.81

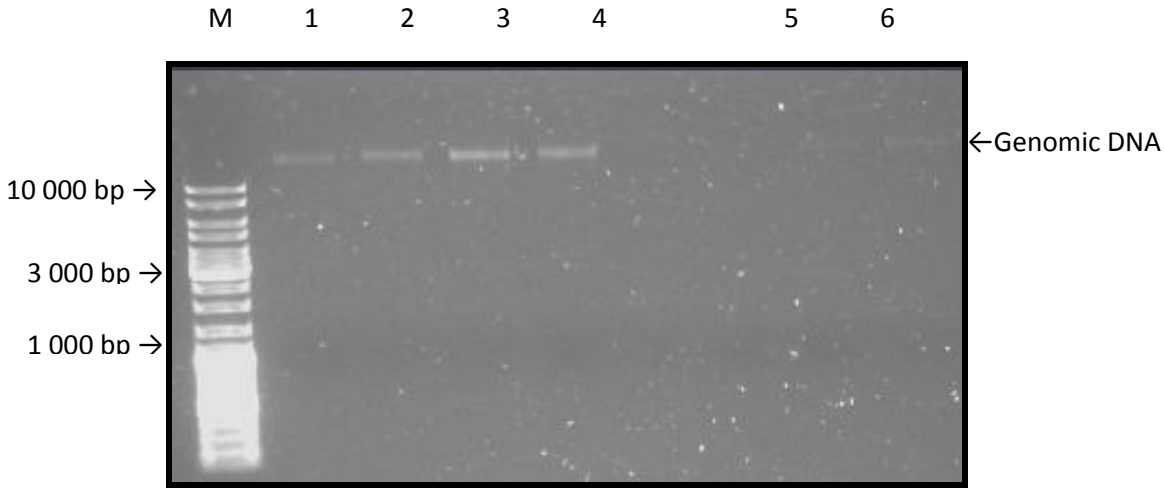


Figure 5.2: Extracted genomic DNA all the enrichments visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; GeneRuler™ DNA ladder (Fermentas), lanes (**1:** NO212FW050508; **2:** NO212FW050508, **3:** TT107FW081111; **4:** BE326FW071211; **5:** KI445FW190711 and **6:** Dr51PC150711).

The DNA was used as a template for the 16S rRNA amplification as described in (sections 2.3). A fragment of approximately 1 500 bp was obtained as expected for bacterial 16S rRNA gene amplicons indicated in (Figure 5.3).

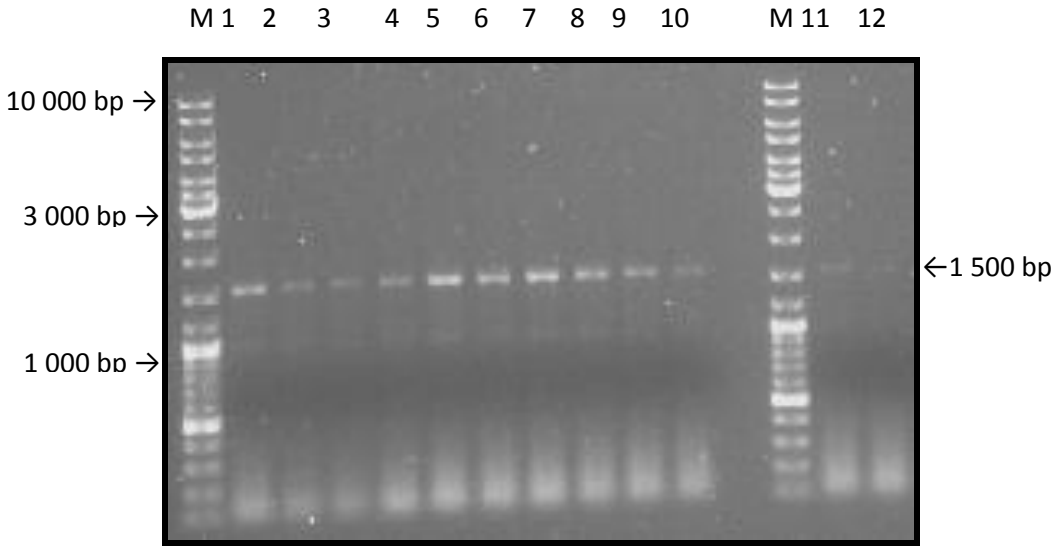


Figure 5.3: Amplification of 16S rRNA fragment from genomic DNA on an ethidium bromide-stained agarose gel 0.8% (w/v). Lane M; GeneRuler™ DNA ladder (Fermentas), lane 1-12; positive amplified band of 16S rRNA from (**1-2:** NO212FW050508; **3-4:** NO212FW050508, **5-6:** TT107FW081111; **7-8:** BE326FW071211; **9-10:** KI445FW190711 and **11-12:** Dr51PC150711).

Most clones that were sequenced and analyzed using nBLAST results indicated that the majority of the organisms present were *Geobacillus* sp. at which 92% - 100% coverage showed 98% - 99% identity with an E-value of zero. This was extremely disappointing since great care and several attempts to eliminate Gram positive dominance still yield Gram positive isolations or at least dominance in consortia of Gram positive microorganisms (as seen in the Gram stain figures). The obtained sequences used of the samples (NO212FW050508, TT107FW081111; BE326FW071211; KI445FW190711 and Dr51PC150711) are shown in Supplement C.

3.3 Fissure water samples identification

Counts of viable cells were carried out using the DAPI staining as described in (section 2.5). Most of the cells from the deep gold mines of South Africa samples showed a few viable cell counts as indicated in (Figure 5.4), however; it is also known that fissure water contain low values of viable cells of ranges between 10^2 to 10^3 cells/mL therefore these outcomes were not surprising as the fissure water was also old and stored away for long periods of time.

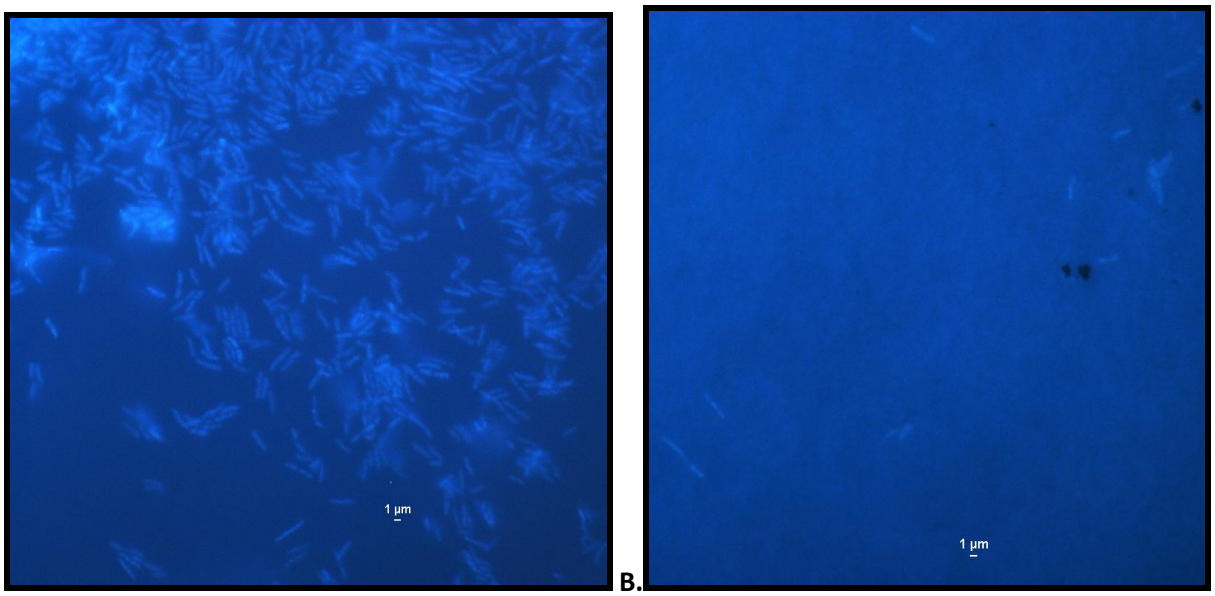


Figure 5.4: DAPI staining for the determination of viable cells within the fissure water from the deep gold mine of South Africa **A)** Positive control (B_4H_3) and **B)** (TT107FW081111) sample. Scale bar 1 µm.

However, even though very few viable cells where observed with most of the deep gold mines of South Africa fissure water samples, the 0.2 µm filter membranes through which 1 mL of fissure water was filtered, as described in (section 2.5), did show growth on the complex medium as indicated in (Figure 5.5).

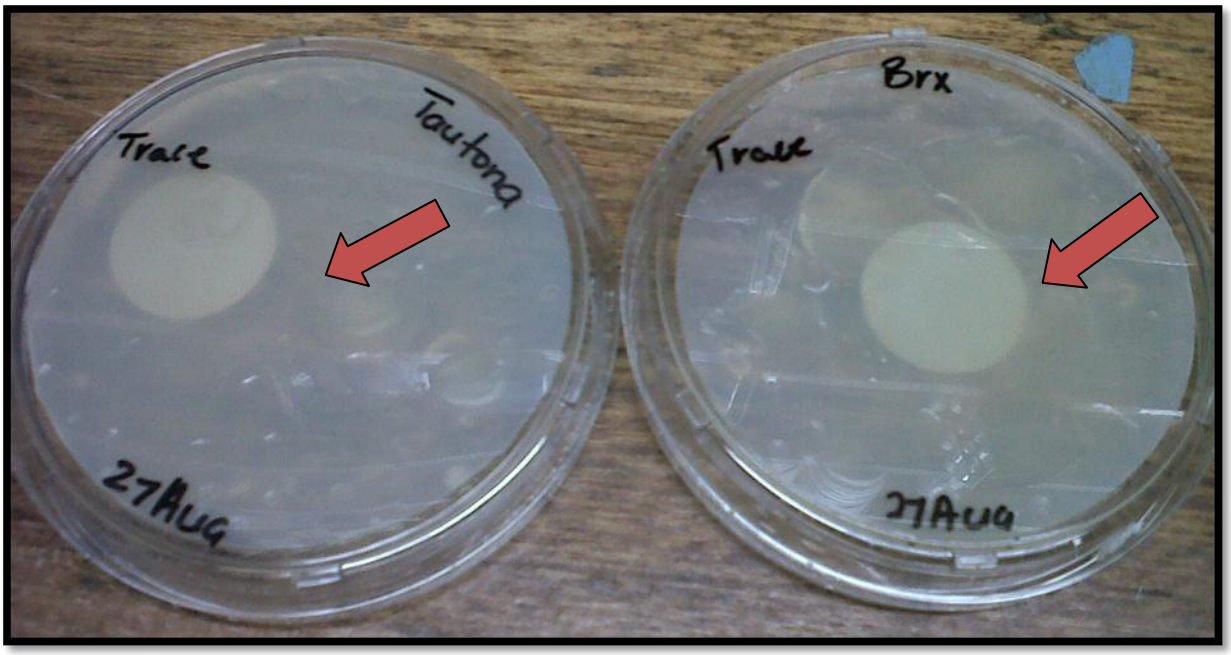


Figure 5.5: Concentrated cells onto a 0.2 µm filter membrane grown onto TYG medium supplemented with trace elements. Two plates represent the fissure water samples of (TT107FW081111) and (BE326FW071211) on which growth was observed after incubation. The arrows are pointing on the area in which a slight cream mass growth was observed.

3.4 Selective media

Using selective media to obtain homogeneity within the cultures grown as described in (section 2.6) showed no additional positive results regarding our search for novel *Thermus* sp. All attempts even from what was observed as single colonies still gave the same outcomes of mixed cultures

Another frustrating aspect was that even observed single colonies from one plate that were subjected to Gram staining as described in (section 2.1) could display mixed morphology of both Gram negative and Gram positive rods of different sizes were observed as shown in (Figure 5.6).

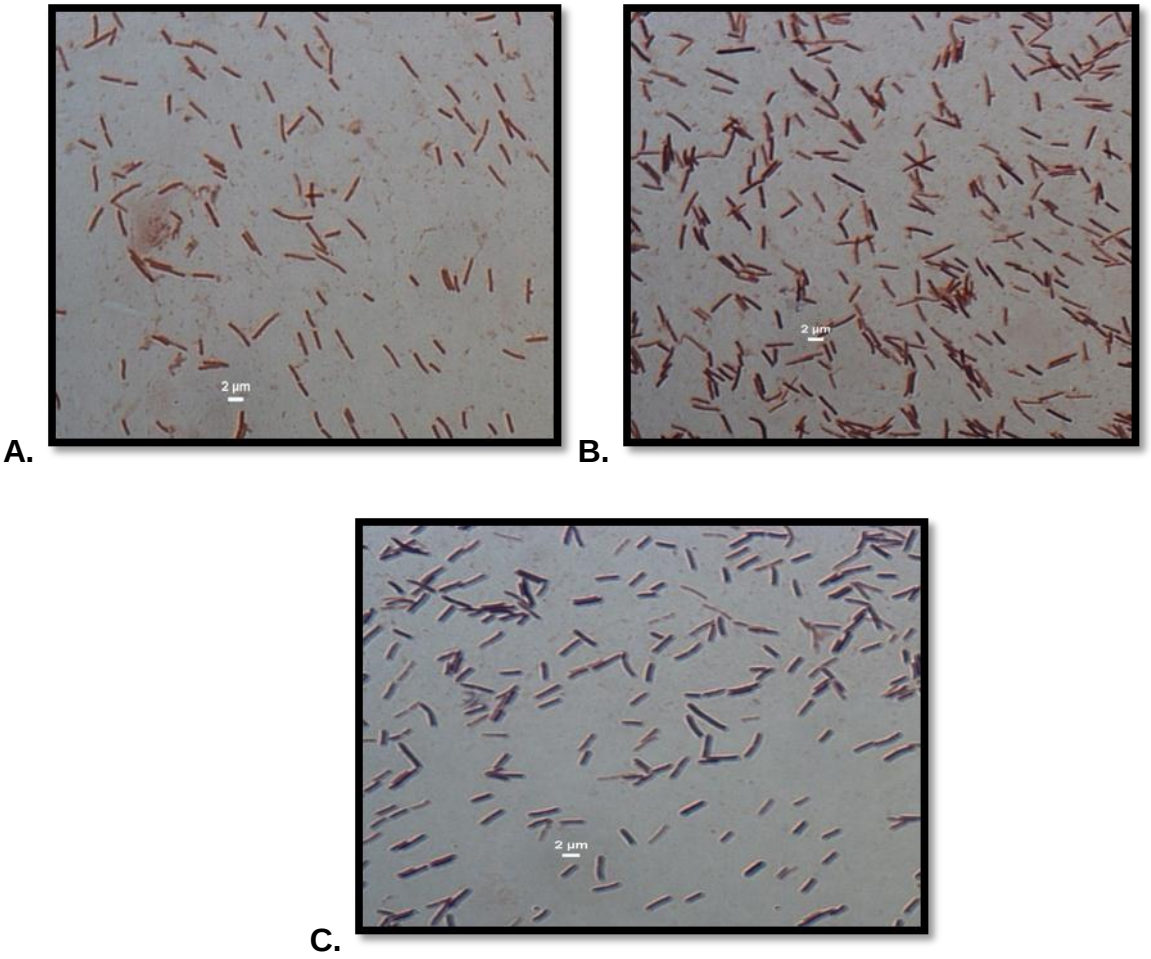


Figure 5.6: Fissure water culture (TT107FW081111) morphology. Gram stain identified both Gram positive and Gram negative rod shaped cells. This was observed within a single colony. **A)** Domination of Gram negative rods **B)** followed by a mixture of cells observed in the center of the microscope slide and on the other end of the slide **C)** a domination of Gram positive rods.

3.5 Application of antibiotics

Our final attempt to remove the Gram positive dominance was to use antibiotic selection in addition to selective media components, antibiotics were used as described in (section 2.7).

The observed morphology of cells supplemented with penicillin [25 µg/mL] and ampicillin [30 µg/mL] showed no homogeneity, however of interest a special characteristic of the genus *Thermus* sp. was observed in which “rotund bodies” were

seen as indicated in Figure 5.7. This isolation gave some phenotypically hope of success. However, again the final colony 16S rRNA analysis showed no positive *Thermus* sp. enrichment.

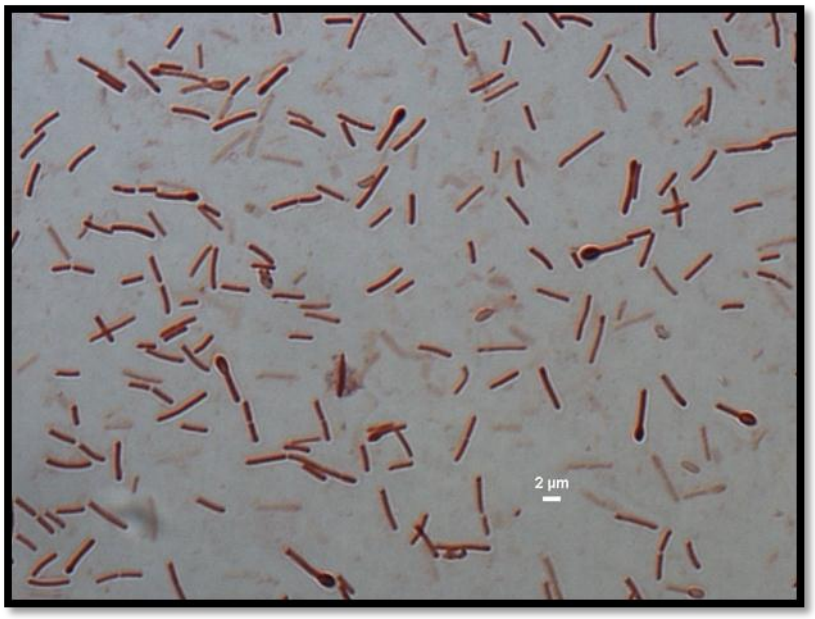


Figure 5.7: Fissure water culture (TT107FW081111) morphology. The presence of "rotund bodies" observed in cell morphology supplemented with antibiotics. Scale bar 2 µm.

3.6 Denaturing Gradient Gel Electrophoresis

In attempt to see if the mixed consortium contained any *Thermus* sp. or at least the phylum *Deinococcus* an extraction of genomic DNA using the NucleoSpin® Soil kit as described in (section 2.8) followed by amplification of the amplicons the obtained results are indicate in Figure 5.9 a). Consortium sample from (TT107FW081111) supplemented with Lactate (electron donor) and KNO_3 (electron acceptor) showed positive amplification while the other (NO212FW050508) consortia sample had no amplicons.

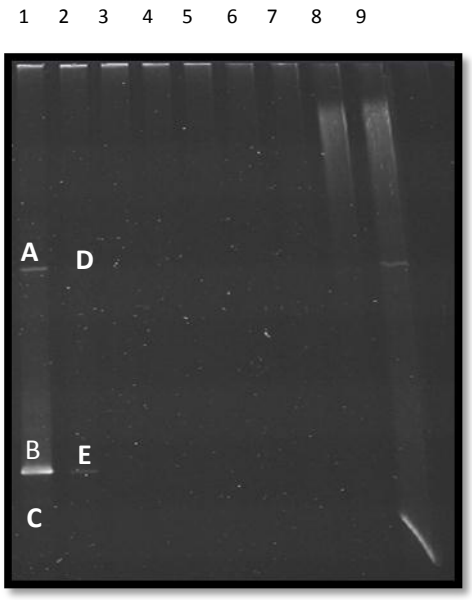


Figure 5.9 (a): Representation of the DGGE fingerprint of bacterial fragments. lane 1, (TT107FW081111) Lactate, lane 2, (TT107FW081111) KNO_3, the remaining lanes represent the (NO212FW050508;) samples which showed no positive amplicons.

Excision of bands was carried out as described in section 2.8.1, followed by re-amplification (Figure 5.9 (b). Positive results were obtained for all of the excised bands of 700 bp.

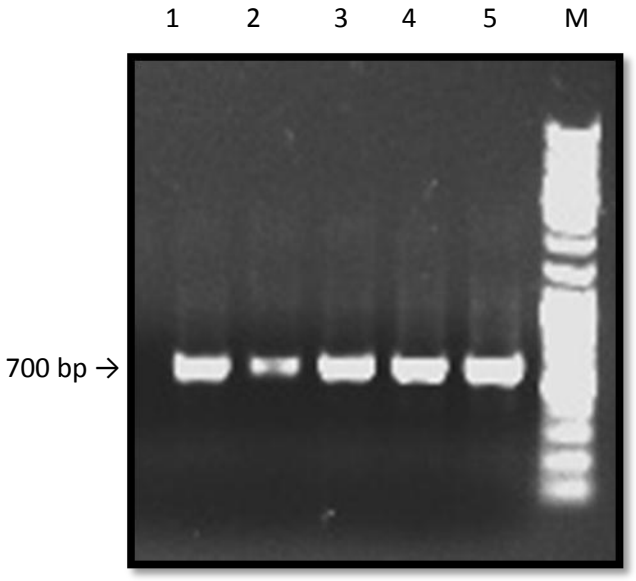


Figure 5.9 (b): Representation of re-amplified amplicons excised from the DGGE gel. Expected band sized of 700 bp were visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; GeneRuler™ DNA ladder (Fermentas), lane 1 - 5; positive amplified amplicons for the obtained symbols in Figure 5.9 (a).

BLAST results indicated that the unknown organisms were likely uncultured *Bacillus* sp. at which a 72% coverage query showed 79% identity with an expected value of relatively close to zero. At this stage it was clear the dominance of Gram positive cells in every sample was observed.

In addition to this stage all attempts to isolate a novel *Thermus* sp. were abandoned and new strategies were discussed with international collaborators and the samples initial data will be transferred to an exchange undergraduate from the USA, alternatively a fellow colleague within the laboratory will further work on the samples.

However, this was not entirely unsuccessful as the majority of the subsurface microorganisms are reported to closely affiliate to the phylum *Firmicutes* on

phylogenetic level (Gihring *et al*., 2006). Also, due to the long storage periods of the fissure waters, could also contribute to the increase of the dominating organisms to lead even during the clone library sequences. Therefore all these new *Firmicutes* isolations were stored for future characterizations at -80˚C.

4. Conclusions

The objective of this chapter was to isolate a thermophilic bacterium from the genus *Thermus* or a relative thereof. Microbiological enumeration procedures are known to give an underestimation of the true microbial community, as all microorganisms are not recovered.

Microorganisms, especially bacteria continue to be discovered living under extreme environmental conditions previously thought to be unable to sustain life in such environments. Included are thermophilic microorganisms which are a group of extremophiles with optima growth temperatures of between 45°C and 75°C, while the hyperthermophiles have the ability of growing optimally beyond 80°C.

Even though the last objective was not successfully carried out to final expected outcomes, a few important aspects remained with the research project in isolation attempts where Gram positive groups might dominate at higher temperatures. The use of selective media including trace elements, antibiotics and metals did not have the desired selection of homogeneous Gram negative morphology. This led us to believe we may strongly be dealing with microorganisms that may be affiliated with the phylum *Firmicutes*.

The adjustment of conditions to accommodate the growth of Gram negative cells entailed an increase of temperature to 65°C and at times 70°C, which are beyond the usual *Geobacillus* sp. optima temperature. This, however, didn't prevent the *Geobacillus* sp. to occur within the BLAST results which was a clear demonstration of the resilience of the Gram positive isolates, and it was indeed a big curve ball to overcome.

The project was therefore completed without fulfilling the objective due to major difficulties encountered. Significant amount of molecular techniques were applied in attempt to find a homologous morphology; possibly a Gram-negative, rod cells indicating the presence of a *Thermus* sp. However, a great amount of *Geobacillus* sp.

and *Bacillus* sp. are obtained from all the experimental procedures carried out as indicated above.

Therefore, it is evident that the isolation of the genus *Thermus* requires more than the application of modified media, selective compounds, antibiotics or even anaerobic respiratory metabolism to break down the nutrients. A lot of specifications do enhance the growth of bacterial species; but when it comes to the genus *Thermus* that does not seem to be an easy task to carry out.

5. Supplement C

Molecular techniques where performed as described in the materials and methods for the identification of the fissure water samples from the deep mines of South Africa, Tables therefore show the sequences obtained and used for nBLAST.

Table 5.7: The 16S rRNA sequence of the Northam platinum mine sample (NO212FW050508) used for identity

AAGGCGGCCTTTGGCTGTCACTTGCGGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAG
GCGACGATGCGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACGCCGCGTGAGCGAAGAAGGCCTTC
GGGTCGTAAAGCTCTGTTGTGAGGGACGAAGGAGCGCCGTTCGAAGAGGGCGGCGCGGTGACGGTACCTCACG
AGGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAATTATTGGG
CGTAAAGCGCGCGCAGGCGGTTCCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAAC
TGGGGGACTTGAGTGCAGGAGAGGAGAGCGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAA
CACCAGTGGCGAAAGGCGGCTCTCTGGGCCTGCAACTGACGCTGAGGCGCGAAGGCGTGGGGAGCAAACAGGA
TTAGATACCCTGGT

Table 5.8: Obtained results after BLAST analysis of the 16S rRNA sequence of the fissure water sample (NO212FW050508) from the Northam platinum mine

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
HQ785619.1	Uncultured organism clone ELU0104-T246-S-NI_000196 small subunit ribosomal RNA gene, partial sequence	1103	1103	100%	0.0	99%
JN692243.1	*Geobacillus* sp. SP38 16S ribosomal RNA gene, partial sequence	1103	1103	100%	0.0	99%
JN692242.1	*Geobacillus* sp. SP37 16S ribosomal RNA gene, partial sequence	1103	1103	100%	0.0	99%
HM059717.1	*Geobacillus* sp. enrichment culture clone W2-10 16S ribosomal RNA gene, partial sequence	1103	1103	100%	0.0	99%
HQ744054.1	Uncultured organism clone ELU0016-T94-S-NI_000359 small subunit ribosomal RNA gene, partial sequence	1099	1099	100%	0.0	99%
JQ743052.1	*Geobacillus stearothermophilus* strain D2-3-3-1 16S ribosomal RNA gene, partial sequence	1098	1098	100%	0.0	99%
JN366780.1	Bacillaceae bacterium 55AA1-1 16S ribosomal RNA gene, partial sequence	1098	1098	100%	0.0	99%
HQ792244.1	Uncultured organism clone ELU0126-T312-S-NI_000169 small subunit ribosomal RNA gene, partial sequence	1098	1098	100%	0.0	99%

Table 5.9: The 16S rRNA sequence of the deep gold mine sample (TT107FW081111) used for identity after anaerobic incubation with metals and nitrogen compounds supplements

GATTGCCGAATAAAAAAGGGCCCTTTTCCGGGTTCCTAAAAAGGCTCTTGTTTGGGGAGGACACAAAGGGAGGC
GCCCGTTCGAAAGAGGGGCGGGCGGGGGGCACTGGTTCCTCCCAGAGAAAACCCCGGGGAATAACTACGGGGC
AACAGCCGCGGTAATACGTAGGGGGCAAGCTTGGCCGAAATTTTTTGGGGGGCGAAAAGCGGGCGCAGGGGGGT
CCTTAAGTCTGAAGGGAAAGCCCCGGTTCTCCGGGGGAGGGGCTTTGGAAATGGGGGGGCTTGTGTGCGGAAAG
TTTAGGAATTCCC

Table 5.10: Obtained results after BLAST analysis of the 16S rRNA sequence of the deep gold mine sample (TT107FW081111) after anaerobic incubation and metals and nitrogen compounds supplements.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
HQ808916.1	Uncultured organism clone ELU0161-T363-S-NIPCRAMgANa_000547 small subunit ribosomal RNA gene, partial sequence	135	135	72%	1e-28	79%
HM598401.1	*Bacillu*s sp. 15-2 16S ribosomal RNA gene, partial sequence	130	130	58%	6e-27	80%
JN699191.1	Bacterium MVI.17 16S ribosomal RNA gene, partial sequence	128	128	45%	2e-26	83%
JN645847.1	*Thalassobacillus devorans* strain SL14 16S ribosomal RNA gene, partial sequence	128	128	45%	2e-26	83%
FJ384500.1	Uncultured Bacillus sp. clone QNSW24 16S ribosomal RNA gene, partial sequence	128	128	51%	2e-26	82%
EU186648.1	Bacillaceae bacterium ML-SRAO 16S ribosomal RNA gene, partial sequence	128	128	58%	2e-26	80%
EF503549.1	*Halobacillus* sp. HPC966 16S ribosomal RNA gene, partial sequence	128	128	45%	2e-26	83%

6. References

Bergey, D H.; John, G. H.; Noel, R. K. and Peter, H. A. S. (1994). Bergey's Manual of Determinative Bacteriology (9th ed.); *Lippincott Williams & Wilkins*.

DSMZ. http://www.dsmz.de/ (Cited 07-08-2011).

Ghiring, T. M., Moser, D. P., Lin, L. H., Davidson, M., Onstott, T. C., Morgan, L., Milleson. M., Kieft, T. L., Trimarco, E., Balkwill, D. L. and Dollhopf, M. E. (2006) The distribution of the microbial taxa in subsurface water of the Kalahari shield, South Africa. *Geomicrobiology J.* **23,** 415–430.

Kieft, T. L., Fredrickson, J. K., Onstott, T. C, Gorby, Y. A., Kostandarithes, H. M. and Bailey, T. J. (1999) Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Appl Environ Microbiol*, **65,** 1214-1221.

Porter, K. G. and Feig, Y. S. (1980) The use of DAPI for indentifying and counting aquatic *microflora. Limnol. Oceanogr*. **25,** 943-948.

Wang, J., Ma, T., Zhao, L., Lv J., Li, G., Liang, F. and Liu, R. (2008) PCR-DGGE method for analyzing the bacterial community in a high temperature petroleum reservoir. *World J Microbiol Biotechnol*. **24,** 1981-1987.

Zhang, X., Ying, Y., Ye, Y, Xu, X., Zhu, X. and Wu, M. (2010) *Thermus arciformis* sp. nov., a thermophilic species from a geothermal area. *Inter. J. of Systematic and Evolutionary Microbiol.* **60** (pt 4)**,** 834-839.

CHAPTER 6

Concluding remarks

General summary

The aim of this project was to sequence the whole genome of *Thermus* sp. NMX2 A.1 strain and compare it to the whole genome sequence of *T. scotoductus* SA-01.Therafter attempt to use experimental data to confirm functionality within the genomes and lastly isolate a new *Thermus* sp. from the fissure water samples routinely collected from the deep gold mines of South Africa and the Northam Platinum mine.

The genus *Thermus* has been extensively studied since the discovery of *T. aquaticus* in 1969 by Brock and Freeze and hundreds of species had been isolated ever since. However, up to date only eight validly described species are comprised in the genus *Thermus*.

Moreover, amongst this handful amount of species from the genus *Thermus*; great discoveries have been made and highlighted parts drew further attention in studying this genus even more. Their metabolism is one of the aspects looked into especially the denitrification respiration. With the application of a nitrate operon it is known that a few species within this genus are able to grow under such conditions, meanwhile; others are modified by genetically manipulating them to do so, for instance *T. thermophilus* HB27.

Subsequently, *T. scotoductus* SA-01 is reported to naturally possess the nitrate operon allowing it to grow in oxygen restricted conditions without any genetic manipulation applied. Recently, another strain was discovered to have the same functionalities as *T. scotoductus* SA-01 when grown under denitrification respiration with the supplement of potassium nitrate.

In addition to that, many phylogenetic similarities and identical remarks were also observed between *T. scotoductus* SA-01 and the newly sequence *Thermus* sp. NMX2 A.1 strain such as the ones carried out using the BOX-PCR fingerprinting. Comparison was, therefore, carried out on genome level to verify the phylogenetic similarities; of which was seen that the two strains shared up to (81.6%) similarities. The remaining percentage represented a set of genes that where uniquely found in the newly sequenced strain such entailed the Calvin cycle.

Concluding remarks

It is however, understandable as to why there might be differences between the two strains since the *T. scotoductus* SA-01 was isolated in the deep gold mines of South Africa and *Thermus* sp. NMX2 A.1 strain from the thermal hot springs of New Mexico in the USA. As a result; *T. scotoductus* SA-01 also indicated islands within its genome in which were not found in the newly sequenced strain.

A detailed experimental procedure was carried out to further support the theoretical similarities of the two strains with the presence of their genes that participate within the nitrate operon. Their functionalities were therefore analyzed using the profile of the nitrite detection as initial step of nitrate reduction, of which they showed an identical profile from the beginning of the incubation period until the end.

Lastly, with the advantages of a variety of fissure water samples collected from the different deep gold mines of South Africa and the Northam platinum mine a search for possible *Thermus* isolate was in addition performed. A range of different applications were implemented to obtain pure Gram negatives cells. This however, didn't successfully yield the expected results; even though there were numerous indications that the Gram negative cells where present and one could even hypothesize that those Gram negative cells were a genus *Thermus* since the "rotund bodies" characteristics were seen amongst these fissure water-samples. Unfortunately this directive was not achieved but the search continues to extend the known *Thermus* sp. isolations and exploration into their metabolic versatility.

Key words: Calvin cycle; Genome, Next Generation Sequence; Nitrate reduction Thermophiles, *Thermus*, *Thermus* sp. NMX2 A.1.

Concluding remarks

Algemene opsomming

Die doel van hierdie projek was om die reeks van die hele genoom van *Thermus* sp. NMX2 A.1 stam te bepaal en dit te vergelyk met die hele genoom van *T. scotoductus* SA-01. Daarna is gepoog om eksperimentele data te gebruik om funksies binne die genome te bevestig en laastens 'n nuwe *Thermus* spesie uit die groef watermonsters te isoleer wat op gereelde basis uit die diep myne van Suid-Afrika en die Northam platinummyn versamel is.

Die genus *Thermus* is omvattend bestudeer sedert die ontdekking van *Thermus aquaticus* in 1969 deur Brock en Freeze en honderde spesies is sedertdien geïsoleer. Tot op hede is egter net agt spesies in die genus *Thermus* geldig beskryf.

Verder, onder hierdie handvol spesies van die genus *Thermus* is groot ontdekkings gemaak en beklemtoonde dele het verdere aandag aan die bestudering van hierdie genus verleen. Hul metabolisme is een van die aspekte waarna gekyk is, veral die denitrifikasierespirasie. Met die toepassing van 'n nitraat-operon het dit bekend geword dat 'n paar spesies in hierdie genus in staat is om onder sulke omstandighede te groei. In die tussentyd is ander spesies verander deur genetiese manupilasie, byvoorbeeld *T. thermophilus* HB27.

Daarna is gerapporteer dat *T. scotoductus* SA-01 die nitraat operon op 'n natuurlike wyse bevat, wat dit toelaat om in suurstof-beperkte omstandighede te groei sonder dat enige genetiese manipulasie toegepas word. Onlangs is 'n ander stam ontdek met dieselfde funksies as *T. scotoductus* SA-01 wanneer onder denitrifikasierespirasie gekweek word met die aanvulling van kaliumnitraat

Daar is ook baie filogenetiese ooreenkomste en identiese punte waargeneem tussen *T. scotoductus* SA-01 en die nuwe reeks *Thermus* sp. NMX2 A.1 stam soos dié uitgevoer met behulp van die BOX-PCR-vingerafdrukke. Vergelyking is dus uitgevoer op genoomvlak om die filogenetiese ooreenkomste te verifieer; waaruit gesien kon word dat die twee stamme tot 81,6% ooreenkomste gedeel het. Die oorblywende persentasie

verteenwoordig 'n stel van gene wat wanneer dit uniek in die nuwe reeks stamme gevind word, dit die Calvin-siklus behels.

Dit is egter verstaanbaar waarom daar dalk verskille tussen die twee stamme is aangesien die *T. scotoductus* SA-01 geïsoleer is in die diep myne van Suid-Afrika, en *Thermus* sp. NMX2 A.1 stam uit die termiese warmbronne van New Mexico in die VSA. As gevolg hiervan, dui *T. scotoductus* SA-01 ook op eilande in sy genoom wat nie in die nuwe reeks stamme gevind is nie.

'n Gedetailleerde eksperimentele prosedure is uitgevoer om die teoretiese ooreenkomste van die twee stamme met die teenwoordigheid van hul gene verder te ondersteun wat binne die nitraat operon deelneem. Hul funksies is dus ontleed met behulp van die profiel van die nitrietopsporing as eerste stap van nitraatvermindering, waar hulle 'n identiese profiel van die begin van die inkubasietydperk tot aan die einde getoon het.

Ten slotte, met die voordele van 'n verskeidenheid van bio-monsters van die verskillende diep myne van Suid-Afrika en die Northam platinummyn, is 'n soektog na moontlike *Thermus* ook uitgevoer. 'n Reeks van verskillende programme is geïmplementeer om suiwer Gram-negatiewe selle te verkry. Dit het egter nie die verwagte resultate suksesvol gegee nie; selfs al was daar talle aanduidings dat die Gram-negatiewe selle teenwoordig was. Daar kan selfs veronderstel word dat die Gram-negatiewe selle 'n genus *Thermus* was aangesien "rotund bodies" eienskappe waargeneem is gesien tussen hierdie skeurwatermonsters. Ongelukkig is hierdie riglyn nie behaal nie, maar die soektog gaan voort om die bekende *Thermus* sp. isolasies uit te brei, en eksplorasie van hul metaboliese veelsydigheid.

Sleutel woorde: Calvin siklus; Genoom; Next Generation Volgorde; Nitraat vermindering; Thermophiles; *Thermus*; *Thermus* sp. NMX2 A.1.