ISOLATION, GENOME SEQUENCING, ASSEMBLY, ANNOTATION AND CHARACTERIZATION OF *Thermus* sp. CCB_US3_UF1 FROM ULU SLIM, PERAK, MALAYSIA

TEH BENG SOON

UNIVERSITI SAINS MALAYSIA
2011

ISOLATION, GENOME SEQUENCING, ASSEMBLY, ANNOTATION AND CHARACTERIZATION OF *Thermus* sp. CCB_US3_UF1 FROM ULU SLIM, PERAK, MALAYSIA

by

TEH BENG SOON

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

February 2011

PEMENCILAN, PENJUJUKAN GENOM, PENGHIMPUNAN, ANNOTASI DAN PENCIRIAN *Thermus* sp. CCB_US3_UF1 DARI ULU SLIM, PERAK, MALAYSIA

oleh

TEH BENG SOON

Tesis yang diserahkan untuk
memenuhi keperluan bagi
Ijazah Sarjana Sains

Februari 2011

ACKNOWLEDGEMENTS

I would like to thank my family for their love, support, and understanding through all years of study. My sincere thanks to faculty members especially my supervisor, Professor Maqsudul Alam, Dr. Jennifer Saito and Dr. Rashidah for giving me the opportunity and necessary guidance. I would like to also attribute this success to our collaborators such as Professor Shinichi Aizawa (Prefectural University of Hiroshima) for providing brilliant electron micrograph pictures, members of the sequencing team lead by Dr. Shaobin (ASGPB, Hawaii) and Dr. Zhemin (TEDA, China), Yamin (bioinformatic technician), Luqman, and all others who have supplied great assistance and a memorable experience.

TABLE OF CONTENTS

CHAPTER 3 – RESULTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	absorbance
ATP	adenosine triphosphate
AAPs	aerobic anoxygenic phototrophic bacteria
ASGPB	Advanced Studies for Genomics, Proteomics and Bioinformatics
ATCC	American Type Culture Collection
bp	base pair
BLAST	Basic Local Alignment Search Tool
c-di-GMP	Bis-(3'-5')-cyclic dimeric guanosine monophosphate
CO_2	carbon dioxide
x g	centrifugal force
Cr(VI)	Chromium(VI)
Co(III)	Cobalt (III)
cDNA	complementary deoxyribonucleic acid
°C	degree Celcius
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DMSO	dimethyl sulfoxide
DIYA	Do-It-Yourself Annotator
dsDNA	double-stranded deoxyribonucleic acid
EC	Enzyme Commission
ED	Entner-Doudoroff
EDTA	ethylene diamine tetraacetic acid
EMP	Embden-Meyerhof-Parnas
GS	genome sequencer

GL	glycolipid
g	gram
g/l	gram per litre
HSP	heat shock protein
HTH	helix-turn-helix
HKs	histidine kinases
HCl	hydrochloric acid
ISGA	Integrated Services for Genomic Analysis
Fe(III)	iron III
kb	kilo base
kb/s	kilobyte per second
kv	kilo volt
KEGG	Kyoto Encyclopedia of Genes and Genomes
Mn(IV)	Manganese(IV)
MK-8	menaquinone 8
M	molar
µl	microlitre
µg	microgram
ml	millilitre
nm	nanometre
MOPS	3-(N-morpholino) propanesulfonic acid
NCBI	National Center for Biotechnology Information
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NO_3^-	nitrate

NA	nutrient agar
NB	nutrient broth
ORF	open reading frame
OD	optical density
PS	Pathway Studio
%	percent
PCI	phenol/chloroform/isoamyl alcohol
PTA	phosphotungstic acid
PTS	Phosphoenolpyruvate:phosphotransferase
PL	phospholipid
pmol	pico mol
PCR	polymerase chain reaction
RT-PCR	reverse transcriptase polymerase chain reaction
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
s	second
σ	sigma
SNP	Single Nucleotide Polymorphism
sp	species
NaCl	sodium chloride
SDS	sodium dodecyl sulphate
S^o	sulfur
TEM	transmission electron microscope

tRNA	transfer ribonucleic acid
TPS	trehalose-phosphate synthase
TPP	trehalose-6-phosphate phosphatase
TCA	Tricarboxylic acid cycle
TBE	Tris boric acid EDTA
TE	Tris EDTA
UV	ultraviolet
U(VI)	Uranium (VI)
V	volt
w/v	weight/volume

PEMENCILAN, PENJUJUKAN GENOM, PENGHIMPUNAN, ANNOTASI DAN PENCIRIAN *Thermus* sp. CCB_US3_UF1 DARI ULU SLIM, PERAK, MALAYSIA

ABSTRAK

Thermus sp. CCB_US3_UF1, satu bacteria thermofilik telah berjaya dipencilkan dari kolam air panas di Ulu Slim, Perak, Malaysia. Genomnya mengandungi 2,243,772 pasangan bes dan satu plasmid bersaiz 19,716 pasangan bes. Terdapat sebanyak 2293 jujukan pengkodan (rangka bacaan terbuka), 2 rRNA operon, 13 gen transposase dan 48 gen tRNA untuk semua 20 asid amino dalam genom ini. Annotasi genom meramalkan bahawa kitar asid trikarboksilik adalah lengkap. Semua protein/enzim dalam laluan metabolisme karbohidrat hadir. *Thermus* sp. CCB_US3_UF1 mempunyai laluan pentosa fosfat tak-teroksida yang diperlukan untuk ATP, histidina dan koenzim sintesis. Semua gen biosintesis untuk valina, leusina dan isoleusina juga hadir menunjukkan bahawa bakteria ini boleh menghasilkan asid amino perlu untuk tujuan pertumbuhan. Beberapa enzim kitar Calvin telah dikenalpasti menunjukkan bahawa *Thermus* sp. CCB_US3_UF1 berkemungkinan boleh mengikat CO_2. Walau bagaimanapun, oleh kerana ketidakhadiran enzim utama laluan ini, ribulose 1,5-biphosphate carboxylase/ oxygenase (RUBISCO), satu laluan pengikat karbon alternatif mungkin digunakan. Pengawalaturan gen yang mengekod DnaJ-DnaK-GrpE, GroEL-GroES, HrcA represor, chaperon ikatan disulfida daripada keluarga HSP33, HtpG chaperon dan protein kecil kejutan haba daripada keluarga HSP20 telah membantu *Thermus* sp. CCB_US3_UF1 untuk beradaptasi kepada persekitaran yang bersuhu tinggi. Protease

seperti HslV, HslU, Clp and Lon boleh ditemui dalam *Thermus* sp. CCB_US3_UF1 dan juga mesofil. Kelainan tahap pengawalaturan protein tersebut merupakan strategi kelangsungan hidup thermofil pada suhu tinggi.

ISOLATION, GENOME SEQUENCING, ASSEMBLY, ANNOTATION AND CHARACTERIZATION OF *Thermus* sp. CCB_US3_UF1 FROM ULU SLIM, PERAK, MALAYSIA

ABSTRACT

Thermus sp. CCB_US3_UF1, a thermophilic bacterium has been successfully isolated from a hot spring in Ulu Slim, Perak, Malaysia. The genome consists of 2,243,772 bp and a 19,716 bp plasmid. There are 2293 predicted coding sequences (ORFs), 2 rRNA operons, 13 transposase genes, and 48 tRNA genes for all 20 amino acids in the genome. The genome annotation predicts that the tricarboxylic acid (TCA) cycle is complete. All the proteins/enzymes for carbohydrate metabolism pathways are present. *Thermus* sp. CCB_US3_UF1 employs the non-oxidative pentose phosphate pathway for the biosynthesis of ATP, histidine and coenzymes. All biosynthetic genes for valine, leucine, and isoleucine are present, implying that this bacterium can generate its own essential amino acids needed for growth. Several enzymes of the Calvin cycle were identified, indicating that *Thermus* sp. CCB_US3_UF1 may be able to fix CO_2. However, due to the absence of the key enzyme of this pathway, ribulose 1,5-biphosphate carboxylase/ oxygenase (RUBISCO), alternative carbon-fixation pathways may be utilized. Regulation of encoded genes DnaJ-DnaK-GrpE, GroEL-GroES, repressor HrcA, disulfide bond chaperone of HSP33 family, chaperone HtpG, and small heat shock protein of HSP20 family helps *Thermus* sp. CCB_US3_UF1 to adapt to a high temperature environment. Proteases such as HslV, HslU, Clp, and Lon are found in *Thermus* sp.

CCB_US3_UF1, as well as in mesophiles. The different rate of regulation of these proteins dictates the survival strategy of thermophiles in high temperature.

CHAPTER 1

INTRODUCTION

1.1 Life

Do we know the origin of life? Life on Earth began over 3.5 billion years ago. The earliest cellular life is further confirmed with the discovery of ancient microfossils called stromatolites from South Africa and Australia (Schopf, 1999). These were originated from cyanobacteria related to modern prokaryotes. It is believed that the earliest forms of life used RNA as the information molecule before proteins provided more accurate biological messages (Gilbert, 1986).

The type of life forms in a particular niche basically depends on the interplay between physical and biological factors. Physical factors (pH, temperature, pressure, oxygen level) and biological factors (diseases, competition, predation) greatly affect different adaptation strategies employed by living organisms. The history and interaction of life with the surrounding environments are well described in a roadmap published by the NASA Astrobiology team (Des Marais *et al*., 2008). This roadmap is used as the introduction of nature and distribution of habitable environments as mentioned below.

1.1.1 Nature and distribution of habitable environments

Life could exist on other planets by acquiring certain environmental requirements. In order to allow life to begin and evolve, equilibrium between liquid water source, formation of complex organic compounds, and energy sources to sustain metabolism must be achieved. Liquid water and oxygen are two essential components for the development and evolution of life on Earth.

Possible life in the solar systems could develop differently than life on Earth. The existence of life elsewhere in the Solar systems provides better understanding to explore life on Earth. The study of microorganisms from extreme environments for example has widened our belief of the potential of life on Mars and the icy moon. By studying the pre-existing phenomenon occurring on other planets, we might gain valuable insight into the origin of life.

Raw materials of life are the key to evolution in habitable environments. Organic compounds are very important in the chemical processes of an organism. Chemical processes occur in every life that leads to the synthesis of important signature biomolecules such as polymers made of amino acids, carbohydrates, and nucleotides. A good chemical system must have molecules that can interact with their environments to capture nutrients and energy, sense environment changes, and produce essential metabolic pathways for growth.

Life evolved in response to environment changes. As a result, living organisms have to perform adaptation strategies such as the development of crucial metabolic pathways to counteract with the environmental perturbations. Comparative genomic studies on environmental samples can provide insights of microbial biodiversity within communities that carry biogeochemical processes such as the use of sulphur, carbon, iron, and nitrogen compounds.

The interaction between genetics, metabolic processes, and environmental change has shaped the diversity of life on Earth. Most of the Earth's environments are colonized by microorganisms that lead to physical-chemical environment change. Microorganisms introduce their own biological processes into the environments that control the evolution of subsequent life. Microbes and viruses are perfect candidates to be used for biochemical, genetic, and genomic studies. Microbes of known

genome sequences can be used to study microbial ecosystems such as predicting environmental and evolutionary changes.

There are places on Earth that are too harsh for most life to exist. There are microbes living at temperatures of 113°C, regarded as hyperthermophiles. Some even thrive in acidic environments are resistant to radiation exposure, or live in deep-sea hydrothermal vents with high hydrostatic pressure. These organisms have evolved special adaptation strategies for survival in extreme environments. Among the strategies performed by microorganisms is the ability to form spores, repair damaged DNA, or live in a dormant stage.

Interaction between biogeochemical reactions with the crust, ocean and atmosphere has formed a huge network that is indispensable to life on Earth. These networks exist within microbial ecosystems and are affected by environmental conditions and changes. The relationship between biological and environmental processes that gives birth to certain ecosystems is poorly understood.

1.2 Extremophiles

Environmental changes have shaped different types of life forms on Earth. Some environments are too hostile for organisms to live in. There are groups of organisms known as extremophiles that are able to thrive in those environments. They come from three domains of life, Archaea, Bacteria, and Eukarya. Extremophiles have the ability to adapt to extreme conditions in terms of salinity, pH, radiation, dessication and hydrostatic pressure that would be lethal for non-extremophiles.

Organisms that can thrive at high temperature beyond 80°C are called thermophiles. Some of them are capable of growing up to 100°C; known as

hyperthermophiles. Modern study of thermophilic microorganisms was started robustly after *Thermus aquaticus* had been discovered (Brock and Freeze, 1969). *Pyrolobus fumarii*, another type of bacterium, grows at 113°C (Bloechl *et al*., 1997).

The term 'psychrophiles' was proposed by Scmidt-Nielsen in 1902 as he had identified bacteria capable of growth at 0°C. Psychrophiles are organisms that have optimal temperature for growth at about 15°C or lower, a maximal temperature for growth at about 20°C and a minimal temperature for growth at 0°C or lower (Morita, 1975).

Alkaliphiles are organisms capable of surviving in high alkaline environments. The bacterium *Streptococcus faecalis* was the first mentioned alkaliphile (Downie and Cruickshank, 1928). Extreme alkaliphiles of genera *Clostridium* and *Bacillus* were isolated from soils (Horikoshi and Akiba, 1982).

Some organisms, known as halophiles, love to live in high saline environments. Saline soils and lakes have been targeted as ideal places to study halophilic microorganisms. *Haloferax mediterranei* has been shown to have good growth at 30% NaCl (Pikuta and Hoover, 2007).

Barophiles can withstand huge hydrostatic pressure associated with great depths such as deep ocean floor. In the Black Smokers studies, microorganisms that need high temperature and pressure (Bloechl, 1997) were detected at pressure of 100 MPa (Yayanos *et al*., 1979).

Some organisms are resistant to high levels of ionizing irradiation. *Deinococcus radiodurans* is the first radioresistant bacterium to be discovered during the process of food conservation and storage (Raj *et al*., 1960). The hyperthermophilic sulphur-reducing *Thermococcus gammatolerans* survives at 30 kGy of gamma-irradiation (Edmond *et al*., 2003).

1.2.1 Thermophiles

Thermophilic microorganisms prefer living at temperatures not commonly found in nature but in hot thermal environments like hot springs and geothermal areas. Hyperthermophiles were among the first living things on this planet when the surface of Earth was hot as a result of frequent bombardment by meteorites coming from outer space. Organisms that can thrive in hot environment will survive.

In recent years, thermophilic organisms have been of biological interest and have been isolated from a variety of sources such as soil, water, and compost. Thermophilic microorganisms have been intensively studied because of its ability to survive at temperatures that normally destroy enzymes, nucleic acids, and cellular components of mesophiles.

Basically, thermophilic microorganisms are divided into three groups. They are moderately thermophilic (optimum temperature at 50-60°C), thermophilic (higher than 70°C), and hyperthermophilic (higher than 80°C). Thermophilic Archaea consists of four phyla: Crenarchaeota (Sulfolobales-Thermoproteales branch), Euryarchaeota (halophiles-Methanogens branch), Korarchaeota, and Nanoarchaota (Pikuta and Hoover, 2007).

Enzymes of hyperthermophiles are of great important because of their high thermostability as well as stability against organic solvents, detergents, and other chemical reagents. Chromosomes of hyperthermophiles are densely packed with genes that play biological role. This has shown that the earliest life forms may have small genomes (Fujiwara, 2002).

1.3 The genus *Thermus*

1.3.1 Taxonomy and Phylogeny

There are a total of eight known species in the genus *Thermus*. They are *T. thermophilus* (Oshima and Imahori, 1974; Manaia *et al*., 1994), *T. filiformis* (Hudson *et al*., 1987b), *T. aquaticus* (Brock and Freeze, 1969), *T. brockianus* (Williams *et al*., 1995), *T. oshimai* (Williams *et al*., 1996), *T. scotoductus* (Kristjánsson *et al*., 1994), *T. antranikianii*, and *T. igniterrae* (Chung *et al*., 2000).

The type strains of each of the eight known species of the genus *Thermus* was compared based on 16S rDNA sequences and shows degree of similarities to be in the range of 91.2-96.4%. *T. oshimai* is the most unrelated of the eight species in terms of 16S rDNA sequence similarity values (da Costa *et al*., 2006). *Thermus* spp are said to be closely related to the genus *Deinococcus* based on several comparative studies on 16S rRNA and protein sequences, thus forming another independent phylogenetic branch of the bacterial tree (Weisburg *et al*., 1989; Griffiths and Gupta 2004, 2007; Omelchenko *et al*., 2005). Nevertheless, the exact phylogenetic position of the *Deinococcus-Thermus* phylum remains to be questioned. This phylum was proposed to derive from the oldest groups of the Bacteria Domain, after those of *Aquifex* and *Thermotoga* based on 16S rRNA sequence comparison (Woese, 1987). A more in-depth analysis of the phylogeny of the *Deinococcus-Thermus* phylum based on the conserved orthologs could be carried out as both of the genomes are completely available (Ciccarelli *et al*., 2006).

1.3.2 Habitats

Thermus spp are isolated from hydrothermal sites where the water temperature is 55-70°C and pH 5.0-10.5 (Kristjánsson and Alfredsson, 1983; Munster *et al*., 1986; Hudson *et al*., 1989; Santos *et al*., 1989).

T. aquaticus was the first bacterium of the genus *Thermus* to be isolated from hot springs in Yellowstone National Park, United States at temperatures of 53-86°C and pH 8.0-9.0 (Brock and Freeze, 1969). Since then, more isolates have been discovered from several hydrothermal areas in Japan (Yoshida and Oshima, 1971; Saiki *et al*., 1972; Taguchi *et al*., 1982), Iceland (Pask-Hughes and Williams, 1977; Kristjánsson and Alfredsson, 1983; Hudson *et al*., 1987a), New Mexico (Hudson *et al*., 1989), New Zealand (Hudson *et al*., 1986), Artesian Basin in Australia (Denman *et al*., 1991), and the Island of São Miguel in the Azores (Santos *et al*., 1989; Manaia and da Costa, 1991).

Besides, *Thermus* strains from shallow marine hot springs have been isolated such as in Iceland (Kristjánsson *et al*., 1986), the Azores (Manaia and da Costa, 1991), and the islands of Fiji (Hudson *et al*., 1989). Interestingly, *Thermus* have been found to live in the abyssal geothermal areas such as in the Mid-Atlantic Ridge and Guaymas Basin, Gulf of California (Marteinsson *et al*., 1995; Marteinsson *et al*., 1999). These springs produce strains that can adapt to high salinity conditions (Manaia and da Costa, 1991; Tenreiro *et al*., 1997). Most of the isolates from inland hydrothermal sites do not survive at salinities above 1% NaCl (Kristjánsson *et al*., 1986; Hudson *et al*., 1989; Santos *et al*., 1989; Manaia and da Costa, 1991; Manaia *et al*., 1994).

Man-made thermal environments are also inhabited by *Thermus* spp, such as *T. aquaticus* (Brock and Freeze, 1969). *Thermus* strains also could be found in hot

water systems (Brock and Boylen, 1973), hot water taps, and thermally polluted streams (Ramaley and Hixson, 1970; Pask-Hughes and Williams, 1975; Degryse *et al*., 1978; Stramer and Starzyk, 1981).

1.3.3 Cell Structure and Lipid Composition

Bacteria of the genus *Thermus* are Gram-negative and have rod-shaped filamentous structures. *T. filiformis* does not produce short rod-shaped cells instead having a stable filamentous cell which differs from other strains (da Costa *et al*., 2006).

All strains of the genus *Thermus* that are known so far are cytochrome oxidase-positive and non-motile in liquid cultures. In terms of cell structures, electron microscopy has shown that *Thermus* spp have a cell envelope, engulfing the cytoplasmic membrane and a cell wall with a thin dense layer represents the peptidoglycan which is connected by corrugated outer layer called “cobble-stone” by invaginations (Figure 1.1) (Hensel *et al*., 1986; Brock and Edwards, 1970; Pask-Hughes and Williams, 1978).

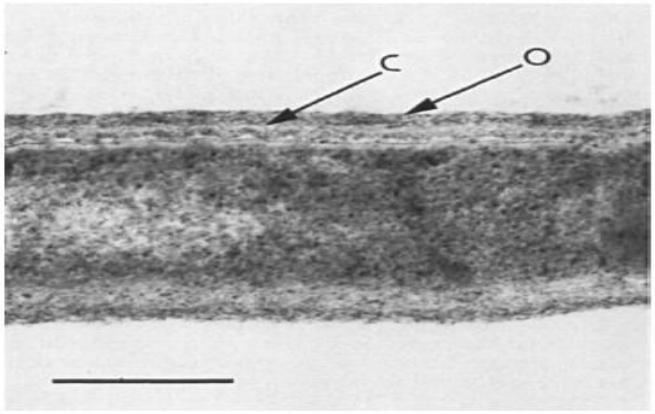

FIG. 2. Thin-section electron micrograph showing cell wall detail. The corrugated layer (C) is typical of *Thermus*, and the outer layer (O) probably serves to hold the trichome intact. Bar = 0.25 μm.

Figure 1.1 The corrugated layer of cell wall of *Thermus*. Figure from (Hudson and Morgan, 1987b), page 432

There are two unusual morphological structures called “rotund bodies” that are sometimes observed by phase contrast microscopy (Brock and Freeze, 1969; Golovacheva, 1976) (Figure 1.2). The aggregation type of structure involves a group of cells bound by the external layer of the cell envelope enclosing a large space between them (Brock and Edwards, 1970; Kraepelin and Gravenstein, 1980; Becker and Starzyk, 1984). Another type of structure with a vesicular shape rotund body is observed as a growth from the surface of a cell (Kraepelin and Gravenstein, 1980).

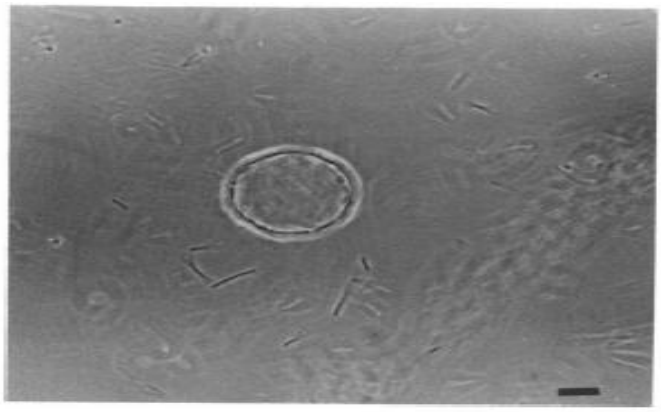

FIG. 6. *Large sphere of the kind frequently seen in T. aquaticus cultures. Bar represents 10 μm.*

Figure 1.2 The rotund body of *T. aquaticus*. Figure from (Brock and Freeze, 1969), page 296

The genus *Thermus* has a peptidoglycan containing glycylglycine as the interpeptide bridge and L-ornithine as the dibasic amino acid (Merkel *et al*., 1978; Pask-Hughes and Williams, 1978). This type of peptidoglycan is also present in the species of genera *Deinococcus* and *Meiothermus* (Hensel *et al*., 1986; Embley *et al*., 1987; Sharp and Williams, 1988). The family *Thermaceae* uses menaquinone 8 (MK-8) as their respiratory quinone (Collins and Jones, 1981; da Costa *et al*., 2001b; Miroshnichenko *et al*., 2003a; Miroshnichenko *et al*., 2003b; Sako *et al*., 2003).

The polar lipid composition of *Thermus* species includes one major glycolipid (GL-1) and one major phospholipid (PL-2). Besides, a minor glycolipid

(GL-2) and phospholipid (PL-1) are detected by thin-layer chromatography (Prado *et al*., 1988; Donato *et al*., 1990). Some of them have a major glycolipid known as diglycosyl-(*N*-acyl)glycosaminyl-glucosyldiacylglycerol which is made up of one *N*-acylated hexosamine and three hexose residues. The polar head group of major glycolipid (GL-1) may consist of three glucose residues, *N*-acylglucosamine or *N*-acylgalactosamine, two glucose residues and one galactose or one glucose and two galactose residues (Oshima and Yamakawa, 1974; Prado *et al*., 1988; Wait *et al*., 1997). The novel long-chain 1,2 diols known as 15-methylheptadecane-1,2-diol and 16-methylheptadecane-1,2-diol have been identified as major components of *T. filiformis* Tok A4 as well as GL-1 and GL-2 of *T. scotoductus* X-1 (Wait *et al*., 1997). It was reported recently that similar diols were present as major glycolipids in other *T. scotoductus* strains as well (Balkwill *et al*., 2004).

The genera *Thermus* and *Meiothermus* both have the same *iso*- and *anteiso*-branched C17:0 and C15:0 fatty acids as the dominant acyl chains. At the optimum growth temperature, most of the strains contain unsaturated branched chain fatty acids and straight chain saturated fatty acids as minor components (Donato *et al*., 1990; Nobre *et al*., 1996a). Majority of the strains have more *iso*-branched fatty acids than *anteiso*-branched fatty acids at the optimum growth temperature as well (Nobre *et al*., 1996a). *Thermus* spp such as *T. aquaticus* and *T. filiformis* also contain mild amounts of branched chain 3-hydroxy fatty acids. Interestingly, the 3-hydroxy fatty acids are only bound to galactosamine present in the glycolipids through amide bonding but are completely absent in the strains where glucosamine replaces galactosamine (Carreto *et al*., 1996).

1.3.4 Physiology

The genus *Thermus* is capable of growing at temperatures ranging between 45°C and 83°C (da Costa *et al*., 2006). Most of them have a maximum growth temperature slightly lower than 80°C (Brock and Freeze, 1969; Chung *et al*., 2000; da Costa *et al*., 2001). Interestingly, a few strains of *T. thermophilus* can grow at 80°C or above (Manaia *et al*., 1994).

Most of the *Thermus* isolates are yellow-pigmented ranging from deep to pale yellow. Majority of the isolates coming from man-made environments that survive in the dark are non-pigmented, although a small number of yellow-pigmented strains still can be found in these environments (da Costa *et al*., 2006). For example, non-pigmented and some yellow-pigmented *Thermus* strains have been isolated from abyssal hot springs (Marteinsson *et al*., 1995).

The ability of yellow-pigmented *Thermus* strains to live in thermal areas exposed to sunlight leads to the hypothesis that carotenoids would protect the cells from harsh sunlight irradiation. Non-pigmented strains have a selective advantage in dark environments because the production of carotenoids is energy-wasting and serves no functional purpose (da Costa *et al*., 2006). A carotenoid gene cluster in *T. thermophilus* strain HB27 is located in the megaplasmid known as TT27 (Henne *et al*., 2004). The wild type strain and carotenoid-underproducing mutants are less resistant to ultraviolet irradiation than the carotenoid-overproducing mutants (Hoshino *et al*., 1994; Tabata *et al*., 1994).

1.3.5 Metabolism

Thermus need carbohydrates, amino acids, carboxylic acids, and peptides as sources of carbon and energy. Most of them are aerobic because they own a respiratory metabolism. Some of them are capable of growing anaerobically using nitrate as electron acceptor and some strains even reduce nitrite as well (Hudson *et al*., 1989; Santos *et al*., 1989; Manaia *et al*., 1994; Chung *et al*., 2000).

T. scoductus strains such as NMX2 A1 and SA-01 utilize Fe(III), NO_3^{-1}, and S° as terminal electron acceptors for growth. Furthermore, these strains also reduce Co(III), Cr(VI), Mn(IV), and U(VI) (Kieft *et al*., 1999). In the presence of organic carbon sources, *T. scotoductus* strain IT-7254 is capable of oxidizing thiosulfate to sulphate (Skirnisdottir *et al*., 2001). Interestingly, the discovery of a few gene homologues related to sulphur oxidation in the genome of *T. thermophilus* HB27 has assumed that this bacterium may obtain energy from reduced sulphur compounds as well (Henne *et al*., 2004).

Thermophilic organisms tend to produce compatible solutes in salt stress environments. The compatible solutes of the thermophilic and hyperthermophilic prokaryotes are totally different from their mesophilic counterparts (Santos and da Costa, 2002). Organisms that grow at extremely high temperatures tend to synthesize di-mannosyl-di-*myo*-inositol-phosphate, diglycerol phosphate, di-*myo*-inositol-phosphate, and mannosylglyceramide as their compatible solutes. On top of that, thermophiles and hyperthermophiles tend to produce mannosylglycerate as compatible solute as well (Martins *et al*., 1997; Nunes *et al*., 1995; Silva *et al*., 1999).

The most common compatible solute present in mesophiles, trehalose, is also present in a few thermophiles and hyperthermophiles (Lamosa *et al*., 1998; Martins

et al., 1997; Silva *et al*., 1999). The main pathway for the synthesis of trehalose in bacteria involves gene *otsA*, encoding trehalose-phosphate synthase (TPS) which converts UDP-glucose and glucose-6-phosphate to trehalose-6-phosphate. Subsequently, trehalose-6-phosphate phosphatase (TPP) encoded by *otsB* dephosphorylates trehalose-6-phosphate into trehalose (Giaever *et al*., 1988). Besides that, another pathway involves trehalose synthase encoded by *treS* converts maltose to trehalose. Some bacteria have homologues of *treS* such as *Mycobacterium tuberculosis* (De Smet *et al*., 2000), *Pimelobacter* sp. (Tsusaki *et al*., 1996), *T. thermophilus* AT-62 and GK24 (Koh *et al*., 2003; Tsusaki *et al*., 1997), and *D. radiodurans* (White *et al*., 1999).

The genera *Thermus* and *Meiothermus* are unable to grow in media containing over 1% NaCl except the strains of *T. thermophilus*. Most of the *T. thermophilus* strains grow in yeast extract-containing media with 3-6% NaCl which make them halotolerant organisms. During salt stress, trace amounts of mannosylglycerate are essential to maintain the osmotic balance although trehalose is still the primary compatible solute (Nunes *et al*., 1995; Empadinhas *et al*., 2003; Silva *et al*., 2003). During osmotic stress, *T. thermophilus* strains accumulate trehalose in yeast extract-containing medium, most probably due to its uptake from the medium itself (Lamosa *et al*., 1998; Mikkat *et al*., 1997). Genetic tools have been developed for *T. thermophilus* strains that can provide better understanding of osmotic adjustment in this organism (Fernandéz-Herrero *et al*., 1995; Lasa *et al*., 1992).

1.3.6 Isolation Procedures

Thermus are easily grown by inoculating biofilms, water, or mud samples in a medium comprising Castenholz D basal salts medium (Castenholz, 1969) with yeast extract (1.0 g/l) and tryptone (1.0 g/l) as supplements. Most of the *Thermus* strains are grown in this medium simply known as *Thermus* medium (Brock and Freeze, 1969; Munster *et al*., 1986; Williams and da Costa, 1992).

Thermus and *Meiothermus* strains are also grown in basal mineral medium 162 (Degryse *et al*., 1978) with the addition of 0.25% yeast extract and 0.25% tryptone with or without solidifying agents. The other combination of media that is suitable for the growth of many strains composes of 0.4% yeast extract, 0.3% NaCl, and 0.8% polypeptone (Oshima and Imahori, 1974). The level of organic nutrients higher than 1.0% inhibits the growth of most of the strains. The organic nutrient such as hexose is inhibitory because of the acidification of the medium (da Costa *et al*., 2006).

Some strains related to *T. thermophilus* HB8 are particularly more susceptible to grow in culture medium containing organic nutrients such as trypticase or polypeptone (8.0 g/l), yeast extract (4.0 g/l), and NaCl (2.0 g/l) (Oshima and Yamakawa, 1974). An extensive list of media together with formulae suitable to grow *Thermus* is listed by (Sharp *et al*., 1995). Defined media have also proven to bring success in some experimental protocols but not for sample enrichment purposes (Silva *et al*., 2003).

1.3.7 Preservation of Strains

Majority of the strains of *Thermus* and *Meiothermus* can be kept frozen at -80°C in liquid nitrogen or *Thermus* medium containing 10-15% glycerol for

prolonged period of time without loss of viability. Strains that are stored in lyophilized form can be maintained for years. Moreover, densely grown *Thermus* colonies on plates of *Thermus* medium survive for about a month at 4°C (da costa *et al*., 2006).

1.3.8 Genetic Manipulation of *Thermus thermophilus*

1.3.8.1 Natural transformation

Natural competence system plays an important role in the development of genetic tools for *Thermus*. For six strains of *Thermus* sp., it has been shown that the natural transformation process is dependent on pH and divalent cations (Hidaka *et al*., 1994; Koyama *et al*., 1986). The availability of the complete genome sequence of *T. thermophilus* HB27 showed that at least 16 genes were involved in natural competence (Friedrich *et al*., 2001, 2002).

Among the genes, three of them (*comEA*, *comEC*, *dprA*) encode proteins identical to DNA translocator components, four pilin-like proteins (PilA1, PilA2, PilA3, PilA4), a traffic-NTPase protein (PilF), a secretin-like protein (PilQ), a leader peptidase (PilD), an inner membrane protein (PilC), and a PilM-homologue. Apart from these conserved competence proteins, another four proteins (ComZ, PilN, PilO and PilW) were discovered with no homologues in the protein data banks. Based on all these genes found a natural competence system model of *T. thermophilus* has been proposed (Averhoff, 2004). Remarkably, this system shows the highest rates of DNA incorporation ever measured (40 kb/s and cell), revealing the efficiency of the system (Schwarzenlander and Averhoff, 2006).

1.3.8.2 Bacteriophages

A few numbers of phages infecting *Thermus* sp. are reported (Cava *et al*., 2009). The tailed icosahedral dsDNA phi-YS40 was the first reported phage to infect *T. thermophilus* HB8 (Sakaki and Oshima, 1975). Filamentous phage PH75 also infects *T. thermophilus* (Pederson *et al*., 2001) and phage TS2126 infects *T. scotoductus* (Blondal *et al*., 2005). At the molecular level, the phi-YS40 phage exists as the most characterized *Thermus* phages (Naryshkina *et al*., 2006; Sevostyanova *et al*., 2007).

1.3.8.3 Plasmids and replication

Plasmids are ubiquitous in *Thermus* spp but remain cryptic as they show no significant benefits to their hosts (Munster *et al*., 1985). One of the best studied cryptic plasmid, pTT8 from *T. thermophilus* HB8, has derived the first group of shuttle *E.coli*/*Thermus* vectors (Koyama *et al*., 1990a). This 9.3 kb plasmid uses the theta mechanism as the mode of replication. It encodes eight proteins, three of them have similarities to some plasmids from mesophiles (Aoki and Itoh, 2007; Takayama *et al*., 2004).

The most commonly used plasmid known as pMK18 was obtained by isolating the minimal replication region of a 16 kb cryptic plasmid derived from *Thermus* sp. ATCC 27737 (de Grado *et al*., 1998). This minimal replicon contains a 1,798 bp region encoding a 402 amino acids replication protein, RepA. RepA is partially similar to the RepT protein encoded by the plasmid pTsp45s from *Thermus* sp. YS45 (Wayne and Xu, 1997). In addition, the ORF35 and ORF7 of plasmids pL4C and pS4C of *Thermus* sp. 4C show a small degree of similarity to RepA as well (Ruan and Xu, 2007).

1.3.9 Biotechnological Applications of *Thermus* spp

1.3.9.1 Enzymes and proteins of biotechnological interest

In general, enzymes, especially thermozymes and proteins from the genus *Thermus*, are of great interest because of their high thermal stability and co-solvent compatibility which make them to be good candidates for biocatalytical processes. Apart from that, purification involving a single step heat denaturation may be performed when mesophilic organisms are used as host. The most well-known enzyme mined from the genus *Thermus* is DNA polymerase. Other than DNA polymerase, other thermozymes from this genus are of great important too (Cava *et al*., 2009).

A protein called RecA plays vital roles in a few cellular processes such as DNA repair and recombination in bacteria (Lusetti and Cox, 2002). RecA is involved in several applications such as SNPs detections (Shigemori and Oishi, 2005), the deprotection and protection of restriction sites for genomes digestion control (Ferrin and Camerini-Otero, 1991; Szybalski, 1997), and the capture of small dsDNA fragments (Clontech cloncapture) (Shigemori and Oishi, 2004).

The genus *Thermus* produces chaperonins that help to fold other proteins (Kohda *et al*., 2000; Teshima *et al*., 1998; Witzmann and Bisswanger, 1998) as well as stabilize enzymes at lower temperature. On top of that, galactosidases from *Thermus* sp. have been extensively used in the production of lactose-free dairy product whereby at the same time they could sterilize the product during heat treatment (Pessela *et al*., 2003, 2007). Several *Thermus* sp. such as *T. caldophilus* (Park *et al*., 1999), *T. yunnanensis* (Gong *et al*., 2005), *T. aquaticus* (Smile *et al*., 1977), and *T. thermophilus* (Angelini *et al*., 2001) have produced alkaline

phosphatase which is important in the PCR product detection, primer labeling, and as a reporter in promoter probe vectors (Moreno *et al*., 2003).

Another enzyme called amylomaltase from *Thermus* is essential in the starch industry. Starch is used as a texturizer in the food industry in which the native one must be modified to enhance its low shear, low solubility in cold water and thermal resistance, poor flavour release characteristics, and elevated viscosity (Hansen *et al*., 2008; Lee *et al*., 2006, 2008; Park *et al*., 2007). DNA ligases have been found in several *Thermus* sp. such as *T. filiformis* (Kim *et al*., 1998), *T. scotoductus* (Jónsson *et al*., 1994), and *T. thermophilus* (Takahashi *et al*., 1984). Moreover, *Thermus* sp. has NAD-dependent DNA ligase which shows 1-2 orders of magnitude higher fidelity than T4 ligase (Luo *et al*., 1996; Tong *et al*., 1999). Another important protein called single stranded DNA binding proteins (SSBs) play a crucial role in protecting and binding to single stranded DNA during recombination, repair, and replication. It has been shown that the participation of SSBs in DNA replication minimize deletion mutagenesis artifacts (Chou, 1992) and expedites DNA amplification independently of the polymerase used (Dabrowski and Kur 1999; Dabrowski *et al*., 2002; Perales *et al*., 2003).

1.3.9.2 *T. thermophilus* as host for protein thermostabilization

It is ideal to use thermophilic hosts as a tool to drive the selection of thermostable recombinant protein. Two articles have shown that thermophiles are perfect candidates to be host for such selection methods. In this case, *Geobacillus stearothermophilus* was used for the thermal selection of a kanamycin nucleotidyl transferase from *Staphylococcus aureus* (Liao *et al*., 1986; Matsumura and Aiba, 1985). Both articles also reported that two identical amino acid replacements, namely

Asp80Tyr and Thr130Lys, were identified which are crucial for the thermostabilization of the protein. Furthermore, the mutation Thr130Lys was found a year ago in the kanamycin resistance gene encoded by plasmid pTB913 of a thermophilic bacillus (Matsumura *et al*., 1984).

Recombinant Hygromycin B phosphotransferase from *E. coli* was expressed and purified in *Sulfolobus solfataricus* (Cannio *et al*., 2001). In 2005, these recombinant variant proteins were successfully purified from *T. thermophilus* (Nakamura *et al*., 2005). On top of that, Bleomycin binding protein Shble thermostable mutants from *Streptoalloteichus hindustanus* were selectively isolated in *T. thermophilus* (Brouns *et al*., 2005).

1.4. Aims of this study

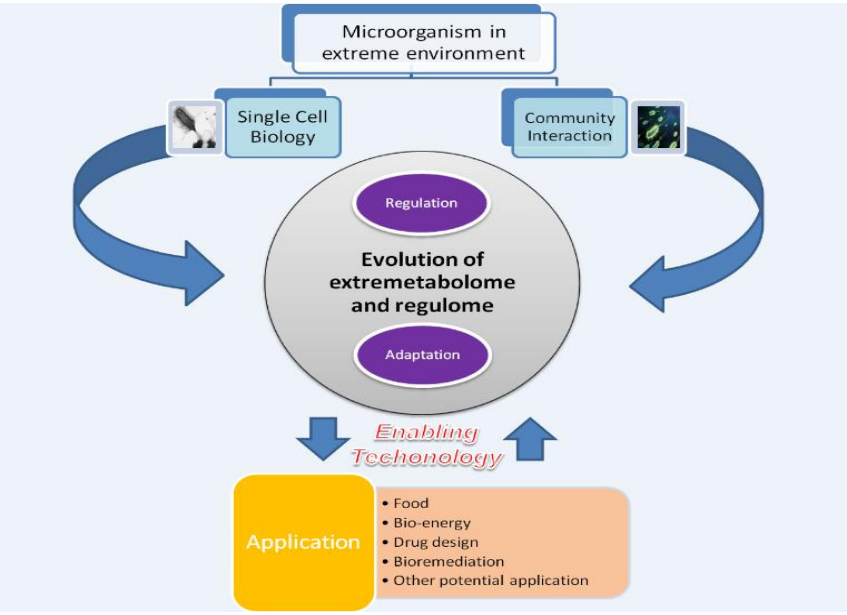

Figure 1.3 CCB@USM Extremophile Roadmap

The Centre for Chemical Biology (CCB@USM) has initiated an extremophile program to explore the biology diverse life forms in Malaysia that are too harsh to live in. Communities in a particular niche form a relationship network to counteract with environmental change. The biological process of each individual organism shows the way they adapt and evolve. The genome of *Thermus* sp. CCB_US3_UF1 serves as a platform to compare modern and primitive life on Earth. The genus *Thermus* is conserved in the deepest branch of ancient origin phylogenetic tree. The origin of life could be explored more related to ancient origin. Ancient bacteria could have evolved in response to the available environments that allow their adaptation. Different environments dictate what kind of metabolic pathways need to be encoded in the genome of the organisms. The *Thermus* genome can be a model to study the evolution of thermophile adaptation.