IDENTIFICATION, CLONING AND EXPRESSING OF DNA POLYMERASE PRODUCING THERMOPHILE FROM GEOTHERMAL WATER IN MALAYSIA

NURUL AKMAR BINTI HUSSIN

UNIVERSITI SAINS MALAYSIA
2013

IDENTIFICATION, CLONING AND EXPRESSING OF DNA POLYMERASE PRODUCING THERMOPHILE FROM GEOTHERMAL WATER IN MALAYSIA

NURUL AKMAR BINTI HUSSIN

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

July 2013

IDENTIFIKASI, PENGKLONAN DAN PENGEKSPRESSAN TERMOFIL PENGHASIL DNA POLIMERASE DARI SUMBER AIR GEOTERMA DI MALAYSIA

NURUL AKMAR BINTI HUSSIN

Tesis yang diserahkan untuk memenuhi keperluan bagi
Ijazah Sarjana Sains

July 2013

ACKNOWLEDGEMENTS

First and foremost, my deepest and greatest gratitude and thankfulness to the Allah S.W.T. for without His Grace and Mercy, I would not be able to complete this thesis.

To my main supervisor Dr. Sasidharan Sreenivasan, I would like to extend my sincere and deep gratitude for giving me the opportunity to pursue my postgraduate studies. Without his knowledge, understanding, guidance, patience and encouragement, I would not have been able to complete my M.Sc. on time. I would also like to thank my co-supervisor, Dr. Venugopal Balakrishnan for his guidance and advice.

Next, I would like to specially thank Dr. Eugene for his advice and mentoring throughout my project. It is always encouraging and exciting to work in a lab that has supportive lab mates. I would like to thank them especially Malar, for all the helps rendered throughout my work in the lab.

To INFORMM, thank you for providing a conducive and comfortable environment for me to do my research. My thanks also go to the administration staff, science officers and lab assistants who have always been pleasant and have assisted me in so many ways.

I am very grateful to Institute of Postgraduate Studies (IPS), USM for providing financial support through the USM Fellowship Scholarship.

Also thanks to Research University Grant (1001/CIPPM/815049) from Universiti Sains Malaysia and HICOE311/CIPPM/4401005 for funded this project.

To my cherished husband Mardani Abdul Halim, family, and friends, thank you for always being there to support and encourage me.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A. amylolyticus	*Anoxybacillus amylolyticus*
A. ayderensis	*Anoxybacillus ayderensis*
A. bogrovensis	*Anoxybacillus bogrovensis*
A. contaminans	*Anoxybacillus contaminans*
A. flavithermus	*Anoxybacillus flavithermus*
A. gonensis	*Anoxybacillus gonensis*
A. kamchatkensis	*Anoxybacillus kamchatkensis*
A. kestanbolensis	*Anoxybacillus kestanbolensis*
A. pushchinoensis	*Anoxybacillus pushchinoensis*
A. rupiensis	*Anoxybacillus rupiensis*
A. voinovskiensis	*Anoxybacillus voinovskiensis*
APS	Ammonium persulfate
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
B. caldotenax	*Bacillus caldotenax*
Bca	*Bacillus caldotenax*
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
Bst	*Bacillus stearothermophilus*
ddH_2O	double distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-stranded Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
Glu	Glutamic acid
Gly	Glycine
HDA	Helicase-dependant amplification
HDMS	Hexamethyldisilazane
IPTG	Isopropyl β-D-1-thiogalactopyranoside

LAMP	Loop-mediated isothermal amplification of DNA
LB	Luria bertani
Lys	Lysine
$MgCl_2$	Magnesium chloride
NA	Nutrient agar
NB	Nutrient broth
NCBI	National center for biotechnology information
OD	Optical density
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
Pwo	*Pyrococcus woesei*
RCA	Rolling circle amplification
rDNA	Ribosomal deoxyribonucleic acid
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SDA	Strand displacement amplification
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning electron microscope
SPIA	Single primer isothermal amplification
ssDNA	Single-stranded deoxyribonucleic acid
SSU	Small-subunit
Taq	*Thermus aquaticus*
TBE	Tris-borate EDTA
TEMED	Tetramethylethylenediamine

IDENTIFIKASI, PENGKLONAN DAN PENGEKSPRESSAN TERMOFIL PENGHASIL DNA POLIMERASE DARI SUMBER AIR GEOTERMA DI MALAYSIA

ABSTRAK

Mikroorganisma tahan haba bersifat aerotolerant, Gram-positif berbentuk rod panjang telah berjaya dipencilkan daripada sampel air yang diambil dari kolam air panas yang terletak di Ulu Legong Hot Spring, Kedah, Semenanjung Malaysia. Di dalam kajian ini, bacteria termofil tersebut telah berjaya dipencilkan daripada sampel sedimen dan dinamakan sebagai 3UL yang menunjukkan pertumbuhan serta beradaptasi dengan baik pada keadaan makmal pada suhu 45 hingga 80°C apabila dibandingkan dengan pencilan lain. Maka, pencilan ini telah digunakan bagi kajian lanjutan. Sepasang primer universal (F_UNI16S and R_UNI16S) telah digunakan untuk mengamplifikasi jujukan gen 16S rRNA tersebut. Keputusan daripada analisis 'jujukan gen 16S rRNA' mendapati rangkaian jujukan 3UL telah mendapat nilai persamaan tertinggi (100%) dengan *Anoxybacillus sp* DR04. Hasil dari ujikaji jujukan gen 16S rRNA untuk 3UL telah disimpan dalam 'Genbank Data Library' dan ditetapkan di bawah nombor kemasukan JQ951796. DNA genom daripada pencilan itu telah diekstrak dan digunakan untuk mengamplifikasi jujukan gen DNA polimerase I. NREAF2 and XREAR_Fxa adalah primer kehadapan dan kebelakang untuk amplifikasi jujukan DNA polymerase dengan ruang halangan *Nco*I untuk primer kehadapan dan *Xho*I untuk primer kebelakang. Panjang gen tersebut adalah 2,628 bp dan mengkodkan protein sepanjang 876 asid amino. Enzim ini mempunyai jisim molekul sebanyak 99 kDa dan menunjukkan persamaan jujukan dengan DNA polymerase I (94%) daripada *Anoxybacillus sp*., (75%) *Geobacillus sp*., dan (74%) *Bacillus sp*. Gen ini telah diekspresikan dalam *Escherichia coli* BL21 (DE3) dengan

menggunakan pET28a(+) sebagai vektor ekspressi yang mempunyai his-tag di hujung C. Kehadiran jalur dengan berat molekul menghampiri 100 kDa pada gel SDS-PAGE membuktikan bahawa pengekspressan DNA polymerase I telah berjaya dan penemuan ini telah disahkan melalui kajian western blot. Dalam kajian ini, DNA polymerase I yang baru daripada *Anoxybacillus sp*. DR04 telah berjaya dipencilkan dan dikenal pasti bagi pelbagai aplikasi.

IDENTIFICATION, CLONING AND EXPRESSING OF DNA POLYMERASE PRODUCING THERMOPHILE FROM GEOTHERMAL WATER IN MALAYSIA

ABSTRACT

Aerotolerant anaerobe, Gram-positive with long rod in shape thermophilic bacteria was successfully isolated from Ulu Legong Hot Spring, Kedah, Malaysia in this study. The thermophilic bacteria was successfully isolated from the sediment sample in this study and denoted as 3UL which showed the best growth and well adapted to the laboratory condition compared with other isolates at temperature ranged between 45 to 80^{o}C. Hence, this microbe was chosen for further study. A set of universal primers (F_UNI16S and R_UNI16S) were used to amplify the 16S rRNA gene sequences for ribotyping identification methods. Its 16S rRNA gene sequences (1454 nucleotides) showed very high homology (100%) with *Anoxybacillus sp*. DR04. The 16S rRNA gene sequence for 3UL has been deposited into "Genbank Data Library" and assigned the accession number JQ951796. Genomic DNA from the isolate was extracted and was used to amplify DNA polymerase I gene sequences. The NREAF2 and XREAR_Fxa were forward and reverse primers used for the DNA polymerase amplification with restriction sites *Nco*I for forward and *Xho*I for reverse. The gene was 2,628 bp long and encodes a protein of 876 amino acids in length. The enzyme has molecular mass of 99 kDa and showed sequence homology with DNA polymerase I (94%) from *Anoxybacillus sp*., (75%) *Geobacillus sp*., and (74%) *Bacillus sp*. The gene was over expressed in *Escherichia coli* BL21 (DE3) by using pET28a(+) as expression vector with his-tag at the C-terminus. The presence of the observed band with molecular weight approximately 100 kDa on SDS-PAGE gel indicate the expression of DNA polymerase I was successful and the interest protein

was detected by western blotting assay for verification. In this study, the new DNA polymerase I from local *Anoxybacillus* sp. DR04 strain was successfully isolated and identified for future applications.

CHAPTER 1.0: INTRODUCTION

An anthropocentric has defined many terrestrial environments present physical and chemical conditions as extreme condition and this kind of environment is colonized by special microorganisms which are adapted to the ecological niches. These organisms are called extremophiles and might be divided into five categories; thermophiles, acidophiles, alkaliphiles, halophiles, and psychrophiles. Among the extreme conditions are high temperature of hot permanent environment such as hydrothermal vents, volcanic areas and hot springs. Hot springs exist throughout the world including Malaysia and different springs have different temperature, chemical compositions and pH values. In the environment with temperature above 65^{o}C, only prokaryotes are present, but the diversity of Bacteria and Archae may be extensive (Madigan *et al*., 2009). Organisms able to growth at such high temperature are defined as thermophiles; whose growth temperature optimum exceeds 45^{o}C while those whose growth temperature optimum exceeds 80^{o}C are called hyperthermophiles. These organisms do not only survive in but might even thrive in boiling water (Vieille and Zeikus, 2001).

Most research on the microbe of hot springs has concentrated on cultivating and isolating extreme thermophilic and acidophilic strains during a decade ago (Belkova *et al*., 2007). Thermophiles have provide many thermostable enzymes that we used today and these enzymes already occupy a prominent position in modern biotechnology, optimizing or even replacing processes that already exist. The majority of the industrial enzymes known to date have been derived from bacteria and fungi. Enzymes from thermophiles are thermostable and expected to be a powerful tool in biotransformation processes (Maugini *et al*., 2009). These enzymes

have several advantages such as more stable toward organic solvents and detergents; can exhibit higher activity at elevated temperatures and also more resistant to proteolytic attacks (Hamid *et al*., 2003).

DNA polymerases are enzymes that responsible for genome replication and existing in all living things (Kim *et al*., 2002). They are a group of enzymes equipped with different functions that involved in DNA replication and repair (Sandalli *et al*., 2009). A variety of DNA polymerases have been isolated, identified, cloned and sequenced from thermophilic, hyperthermophilic and mesophilic bacteria (Kim *et al*., 2002; Choi *et al*., 2004). The amino acid sequences of these DNA polymerase have been aligned and partial homologous regions have been identified (Choi *et al*., 2004; Akhmaloka *et al*., 2008).

DNA polymerases have been extensively used in molecular biology such as DNA amplification and DNA sequencing by polymerase chain reaction (Kim *et al*., 2002; Akhmaloka *et al*., 2008). Thermostable DNA polymerases are our focusing in this study. Most of the thermostable DNA polymerases have been isolated from *Thermus aquaticus*, a thermostable bacterium, known as *Taq* polymerase (Roayaei and Galehdari, 2008). These enzymes are able to withstand the protein-denaturing conditions at high temperature which is required during PCR and replaced the DNA polymerase from *E. coli* originally used in PCR (Roayaei and Galehdari, 2008). The increasing number of applications utilizing PCR and isothermal amplification has generated increasing demands for thermostable DNA polymerase. The use of thermostable DNA polymerase has made PCR applicable for to a large variety of molecular biology problems concerning DNA analysis (Roayaei and Galehdari, 2008).

PCR requires thermocycler machine to control the temperature during DNA amplification and this characteristics made its application in the field limited. Therefore, several isothermal amplification techniques have been developed without using thermocycler machines. Because they are isothermal amplification, any reaction can takes place at any single temperature whether at high temperature or low temperature depending on the properties of polymerase being used (Demidov, 2005; Gill *et al*., 2008). Some of the best known isothermal amplification methods that utilize DNA polymerases are rolling circle amplification (RCA), loop-mediated isothermal amplification (LAMP), single primer isothermal amplification (SPIA) and helicase-dependent amplification (HDA).

There are two DNA polymerases that efficiently working with isothermal amplification methods, which are *Bst* DNA polymerase and *Phi29* DNA polymerase (Demidov, 2002; Yoshimura *et al*., 2006). Yoshimura *et al*., 2006 has proved that *Bst* DNA polymerase is the most suitable for RCA reaction because an enough amount of amplified DNA had been obtained in a short time compared with other polymerases. Isothermal amplification methods can conveniently performed at room temperature but some country their room temperature are higher than other country which are unfavorable for some mesophilic DNA polymerases to react.

Hence, this study will be carried out to isolate and identify of thermophilic bacteria producing thermostable DNA polymerase from Ulu Legong Hot Spring. Thus, this identified DNA polymerase will be cloned and expressed in order to produce the thermostable DNA polymerase that can be used for isothermal amplification.

1.1 Objectives of the Project

1.1.1 General objective:

1. Identification and isolation of Isothermal DNA polymerase producing thermophile from local hot spring.

1.1.2 Specific objectives:

1. Isolation and identification of thermophilic bacteria from Ulu Legong Hot Spring.

2. Cloning and expression of isothermal DNA polymerase from the isolated thermophile.

CHAPTER 2: LITERATURE REVIEW

2.1 Geothermal Areas

Geothermal areas have been divided into two categories, high temperature areas and low temperature areas. The high temperature areas are located within the volcanic zone with the heat above 200°C (Vesteinsdottir, 2008), especially in the top 1000 meters of the earths crust. These areas are mostly mud pots, sulphur pots and fumaroles (Gudmundsson *et al*., 2010). Water from high temperature areas is usually rather sour, acidic and are more of dissolved chemicals than in the low temperature areas (Vesteinsdottir, 2008). The low temperature areas are at the top 1000 meters under the surface with temperature below 150°C. The main characteristics of the low temperature areas are clear water pools and springs with temperature from 20-100°C. The hot water of these areas is mostly alkaline with pH between 8 and 10 and their chemical content is most of the times like the contents of freshwater (Vesteinsdottir, 2008; Cid-Fernandez, *et al*., 2007).

A unique and distinct collection of plants, animals and microorganisms can be found in geothermal areas and because of their adaptations to extreme temperatures and toxic environments; some of these species are very valuable to science. The plants, animals and microorganisms that live in geothermal areas must be able to survive extremes of temperature, acidity and alkalinity, turbidity, and toxicity. The unique thermophilic properties of the microorganisms in hot-springs are mostly attracting many scientists and researchers. They harvest the samples of the microorganisms and grow them in laboratory conditions in order to study them for

potential use in many applications including medicines, food products and the mining of precious metals.

2.2 Ulu Legong Hot Spring, Kedah

Hot springs exist throughout the world, but they are mainly abundant in the western United States, New Zealand, Iceland, Japan, Italy, Indonesia, Central America, and central Africa. The world's largest single concentration of hot springs is in Yellowstone National Park, Wyoming (USA). Many of hot springs have constant temperature, not varying less than 1-2°C over many years even though the temperature between the hot springs can vary greatly, ranging from 20°C - 100°C depending on location and pressure (Madigan *et al*., 2009). In Malaysia, there are about 45 known hot springs; most of them are either under developed or totally undeveloped for open to public people (Samsudin, *et al*., 1997). From the geological study, hot springs in both Peninsular and East Malaysia are originated from the deep-lying groundwater of the earth crust which moved forwards the surface as a result of magmatic heat and pressure. The groundwater seeping through the fractures and crevices in the earth's crust is heated by contact with the hot granitic rock and emerges as hot springs (Samsudin *et al*., 1997). Samsudin *et al*. (1997) has indicates that most of the hot springs commonly occur at low lying areas in various geographic environments which include swampy areas, river beds and bedrock surfaces. However, few of the hot springs is found situated within the area of sedimentary rock and close to the granite body.

Ulu Legong Hot Spring is one of the interesting places in Baling, Kedah. Located in Mukim Siong, Baling, this is approximately at 66 km from Sungai Petani town and 22 km from Baling town. Apart from seeking relaxation, people with ailments and skin problems go there to seek therapeutic treatment and to enjoy the

natural hot mineral waters by immersing themselves in one of the available five hot spring pools. The water content has high properties of sulphur at the best temperature at about 30°C and 60°C.

2.3 Thermophilic Microorganisms

For several decades, thermophilic bacteria have attracted the interest of many scientists due to their biotechnological potential in addition to scientific curiosity. In particular, phenotypic and genotypic characterization of thermophilic bacteria has been done for many geothermal areas in different regions in the World, including Turkey (Adiguzel *et al*., 2009), South Shetland archipelago (Llarch *et al*., 1997), Iceland (Vesteinsdottir, 2008), Mexico (Minana-Galbis *et al*., 2010), and Japan (Sugimori *et al*., 1991). Thermophiles are organisms whose growth temperature optimum exceeds 45°C and hyperthermophiles are organisms whose growth temperature optimum exceeds 80°C (Vieille and Zeikus, 2001). Most of thermophiles can be found in hot springs and other thermal environments. From previous findings, prokaryotic organisms are to grow at far higher temperatures than are eukaryotes and the most of thermophilic prokaryotes are certain species of *Archaea* (Madigan *et al*., 2003). Thermophiles also have been found in artificial thermal environment such as hot water heaters as long as the temperature is favorable habitat for them to grow. Thermophilic species are found in most bacterial genera. In general appearance they resemble their mesophilic counterparts, ferment similar carbohydrates, utilize similar nitrogen sources, and have similar oxidative pathways. They can exist as aerobes, anaerobes, or as facultative aerobes; autotrophic and heterotrophic (Singleton *et al*., 1973).

Many studies have shown that enzyme and proteins of thermophiles are much more heat-stable than are those from mesophiles and function optimally at high temperature. The enzymes in thermophiles are more stable at high temperature because they often differ very little in amino acid sequences of protein as compared with mesophiles which catalyze the same reaction (Madigan *et al*., 2003; Vesteinsdottir, 2008). In addition, an increase in the proportion of charged residues and improved electrostatic interactions are among the most consistent mechanisms for increasing protein thermal stability (Kumar *et al*., 2001). An increase in Ionic bonds between basic and acidic amino acids, and increase in occurrence of hydrophobic residues with branched side chain in proteins have also contributed in heat-stability (Kumar *et al*., 2001; Madigan *et al*., 2003). The cell membrane of thermophiles is made up of saturated fatty acids. The fatty acid provides a hydrophobic environment for the cell and keeps the cell rigid enough to live at elevated temperatures (Haki *et al*., 2003).

Other than that, the DNAs of thermophiles are other aspects of heat stability which contains a reverse DNA gyrase which produces positive super coils in the DNA. The positive super coils will raises the melting point of the DNA (the temperature at which the strands of the double helix separate) to at least as high as the organisms' maximum temperature for growth (Madigan *et al*., 2003). In addition, hyperthermophilic and mesophilic enzymes have typically linear Arrhenius plots, suggesting that functional conformations of their enzyme remain unchanged throughout their perspective temperature ranges (Vieille and Zeikus, 2001).

2.4 Thermophilic enzymes

Thermophiles and hyperthermophiles are interesting for more than just basic biological reasons. These organisms offer some major advantages for industrial and biotechnological processes, many of which can be run more rapidly and efficiently at high temperatures. Microbial enzymes already occupy a prominent position in modern biotechnology, optimizing or even replacing processes that already exist. The majority of the industrial enzymes known to date have been derived from bacteria and fungi. In general, enzyme from thermophiles and hyperthermophiles are more stable than enzyme from mesophiles and expected to be a powerful tool in biotransformation processes (Maugini *et al*., 2008). These enzymes are more resistant to proteolytic attacks, more stable toward organic solvents and detergents and display higher activity at elevated temperatures (Hamid *et al*., 2003). Many studies dedicated to the comprehension of molecular basis of the adaptation to high temperature especially in the field of molecular and physiological properties of extremophiles.

Most enzymes characterized from hyperthermophiles are optimally active at temperatures close to the host organism's optimal growth temperature, usually 70 to 125°C but sometimes are optimally active at temperatures far above the host organism's optimum growth temperature especially extracellular and cell-bound hyperthermophilic enzymes (i.e., saccharidases and proteases). For example, *Thermococcus litoralis* amylopullulanase is optimally active at 117°C, which is 29°C above the organism's optimum growth temperature of 88°C (Haki *et al*., 2003 and Vieille and Zeikus, 2001). Intracellular enzymes (such as xylose isomerases) are usually optimally active at the organism's optimal growth temperature usually less than extracellular enzymes purified from the same host. Only a few enzymes have

been described that are optimally active at 10 to 20°C below the organism's optimum growth temperature. While most hyperthermophilic enzymes are intrinsically very stable, some intracellular enzymes get their high thermostability from intracellular factors such as salts, high protein concentrations, coenzymes, substrates, activators, or general stabilizers such as thermamine (Vieille and Zeikus, 2001).

Surveys of collection of hyperthermophilic and thermophilic proteins pointed out that the most common differences of amino acid composition to the mesophilic counterparts involve charged residues. Maugini (2009) have indicated that, at the subunit interfaces, Arg and Glu content markedly increases while Lys does not vary significantly. Asp and Asn content diminishes. This pattern is in common to the hyper- and thermophilic interfaces. Decrease of Gly frequency suggests that the polypeptide chain is more rigid at the interface of extemophilic proteins. Maugini (2009) suggested in his statistics, there were a net increase at the interfaces of both types of extremophilic interfaces of Phe content which, on the contrary, decreases at the monomer level.

2.4.1 The advantages of thermostable enzymes

Thermostable enzymes are more stable and active at temperatures which are even higher than optimal temperatures for the growth of microorganisms. Therefore thermostable enzymes are more popular and become interest for industrial and biotechnological uses. There are a few advantages in using thermostable enzymes compared to thermolabile enzymes. As the temperature of the process is increased, the rate of reaction increases for example, an increase 10°C in temperature approximately doubles the reaction rate, which in turn decrease the amount of

enzyme needed (Zamost *et al*., 1991). The ability to withstand to high temperatures will increase a longer half-life to the thermostable enzyme. This is useful in systems such as glucose isomerase, which is used at high temperatures (50-65°C) in immobilized reactors for periods up to 12 months (Zamost *et al*., 1991). Conducting biotechnological and industrial processes at elevated temperature (above 60°C) may reduce the risk of microbial contamination by common mesophiles (Haki *et al*., 2003; Zamost *et al*., 1991). In addition, by applying high temperatures in industrial enzyme processes may also be helpful in mixing, causing a decrease in the viscosity of liquids (Zamost *et al*., 1991). Allowing a higher operation temperature has also a significant influence on the bioavailability and solubility of organic compounds and thereby provides efficient bioremediation (Haki *et al*., 2003).

The ability to clone the genes from thermophiles into mesophilic production strains has been use widely in protein engineering. Therefore, study for understanding of thermostability and thermo-activity has been increases. Nowadays, there are several thermostable enzymes available commercially. The protease enzyme is used to remove protein-based stains, food industry, leather softening, hydrolysis of protein, production of aspartame and other peptides. Example of thermostable protease is the protease pyrolysin, from *Pyrococcus furiosis,* which has the greatest thermostability of any reported protease, showing a half-life of 3600 min at 98 (Zamost *et al*., 1991). Amylolytic enzymes are important in starch industry for hydrolysis and modifications in this useful raw material. These enzymes are α-amylases, glucoamylases or β-amylases and isoamylases or pullulanases. Thermostable amylolytic enzymes were isolated from diversified source such as *Bacillus sp*. to meet the requirements of the starch industry. Termamyl and Fungamyl are two well-known amylolytic enzymes which are now available commercially

(Haki *et al*., 2003). Xylanases are secreted by a variety of bacteria, fungi, and yeast. This enzyme is used to breakdown hemicellulosic materials. It has a great application in the pulp and paper industry. However, the search for thermophiles xylanase with higher yield of enzyme and the desired characteristics is still in pursued (Zamost *et al*., 1991; Haki *et al*., 2003).

Thermostable DNA polymerases are our focusing in this study. Thermostable DNA polymerases, such as *Taq* DNA Polymerase was first isolated from thermophilic bacterium *Thermus aquaticus* YT-1 and has been considered as the key element in the development of the polymerase chain reaction (PCR) (Saiki *et al*., 1988; Mullis *et al*., 1986). In earlier PCR procedures, *Escherichia coli* DNA polymerase were utilized (Mullis *et al*., 1986). However, these enzymes lost their enzymatic activities at elevated temperatures and thus, adding a new polymerase enzyme after each cycle following the denaturation and primer hybridization steps was necessary. This process made the thermal cycling a time-consuming and costly procedure. Therefore, with existence of *Taq* polymerase have become beneficial for PCR technologies. However, the multiple applications of the PCR technology make use of two major properties of these DNA polymerases: processivity and fidelity. Although, *Taq* polymerases with 5'-3' exonuclease activity, a 3'-5' exonuclease acitivity (proofreading activity) was not detected (Chien *et al*., 1976). The *Taq* polymerase synthesizes DNA faster (but with a higher error rate) than do enzymes with 3'-5' proofreading acitivity. *Taq* DNA polymerase's high processivity make it the enzyme of choice for sequencing or detection procedures. When high fidelity is required proofreading enzymes (such as Vent and Deep Vent polymerases) are preferred.

Currently, there are many nucleic acid isothermal amplification methods that widely used in research, forensics, and medicines. A larger variety of DNA polymerases can be performed in isothermal assays compared to PCR which relies on only thermostable enzymes (Nelson et al., 2002; Roayaei et al., 2008; Gill *et al*., 2008). Examples of isothermal amplification methods are RCA, LAMP, SPIA, and HDA. Until today, *Bst* DNA polymerase and *Phi29* DNA polymerase have been extensively used in isothermal amplification methods (Demidov, 2002; Yoshimura *et al*., 2006; Gill *et al*., 2008).

2.5 *Anoxybacillus sp.*

The first representative of the genus *Anoxybacillus*, *A. pushchinoensis* was described by Pikuta *et al*. (2000) as strictly anaerobic and an emended description of the species was published later on (Pikuta *et al*., 2003) according to which this species should be considered as aerotolerant anaerobe and the genus *Anoxybacillus* should be emended to aerotolerant anaerobes and facultative anaerobes. In the next few years, new representatives of the genus *Anoxybacillus* have been described and it comprises eleven species at the time of writing of this thesis: *A. pushchinoensis* (Pikuta *et al*., 2000), *A. flavithermus* (Pikuta *et al*., 2000), *A. gonensis* (Belduz *et al.*, 2003), *A. contaminans* (De Clerck *et al*., 2004), *A. ayderensis* (Dulger *et al*., 2004), *A. kestanbolensis* (Dulger *et al*. 2004), *A. voinovskiensis* (Yumoto *et al*., 2004), *A. kamchatkensis* (Kevbrin *et al*., 2005), slightly acidophilic species *A. amylolyticus* (Poli *et al*., 2006), strict aerobe *A. rupiensis* (Derekova *et al*., 2007), and *A. bogrovensis* (Atanassova *et al*., 2008). Although the name of the genus *Anoxybacillus* means ‘‘without oxygen *Bacillus*’’, according to the authors (Pikuta *et al.*, 2000), most of the species described grow well aerobically and even for some

species anaerobic growth was registered only under certain conditions (Yumoto *et al*. 2004). The genus *Anoxybacillus* includes Gram-positive, sporeforming rods, alkaliphilic or alkalitolerant, thermophilic and aerotolerant or facultative anaerobes. *Anoxybacillus flavithermus ssp. Yunnanensis ssp*. nov -Organic-solvent-tolerant bacteria are a relatively new subgroup of extremophiles. They are able to overcome the toxic and destructive effects of organic solvents on account of their unique adaptive mechanisms (Dai *et al*., 2011).

2.6 16S ribosomal RNA (16S rRNA)

16S rRNA is a part of the 30S small subunit of prokaryotes ribosomes with approximately 1500 bp in length and now frequently used for taxonomic purposes for bacteria. The 16S rRNA gene is also designated 16S rDNA, and the terms have been used interchangeably (Clarridge, 2004). The comparison of 16S rRNA sequences is a great tool for tracing phylogenetic relationships between bacteria, and to identify bacteria from various sources, such as environment or clinical specimens. 16S rDNA sequence for identification is of significance because (i) its ribosomal SSU are highly conserved nucleotide sequences among bacteria and includes regions with genus- or species-species variability and exist universally; (ii) almost all bacteria possess 16S rRNA, often existing as a multigene family, or operons; (iii) its function has not changed over time; and (iv) 1500 bp of the 16S rRNA gene is large enough for informatics purposes (Mignard, *et al*., 2006; Janda, *et al*., 2007). The variable regions of DNA sequences form the basis of phylogenetic classification of microbes (Harris, *et al*., 2003). The 16S rRNA gene can also be compared with 16S rRNA gene of archaebacteria and 18S rRNA gene of eucaryotes (Clarridge, 2004). The most potential use of 16S rRNA gene sequence informatics is to provide genus

and species identification for isolates that do not fit any recognized biochemical profiles.

Nowadays, this technology is used in clinical laboratories for routine identifications, especially for slow-growing, unusual or fastidious bacteria (Mignard, *et al*., 2006). The bacteria grow slow in the laboratory due to stringent growth requirements, or may not grow because of prior empirical treatment of patients with antimicrobial agents (Harris, *et al*., 2003). 16S rRNA sequencing is also used as a method of detecting pathogens in normally sterile clinical specimens, or for detecting species that cannot be cultured (Mignard, *et al*., 2006).

2.7 DNA polymerase

DNA polymerase plays the central role in the processes of life as it plays leading roles in cellular DNA replication and repair that is present in all living things. It carries the weighty responsibility of duplicating genetic information. Each time a cell divides, DNA polymerase duplicates its entire DNA, and the cell passes one copy to each daughter cell. In this way, the integrity of all organisms can be accomplished by DNA polymerase. DNA polymerase synthesizes DNA with extraordinary fidelity and efficiency to guarantee proper transfer of genetic information from parent to progeny. DNA polymerase from *Escherichia coli* was first isolated by Kornberg and colleagues in the 1950s (Kornberg *et al*., 1992). Nowadays, more than 100 DNA polymerases from various organisms have been isolated and studied including thermophile and archae. Their deduced amino acid sequences have been compared and characterized. DNA polymerase can be classified into six families: A, B, C, D, X, and Y (Steitz, 1999; Sandalli *et al*., 2009). The DNA polymerases that share sequence homology with *E.coli* DNA polymerase I, II, and III

have been classified into the A, B, and C families, respectively (Uemori *et al*., 1993; Sandalli *et al*., 2009). DNA polymerases in family A possibly the most extensively studied such as the *E.coli* DNA Pol I and *Thermus aquaticus* DNA Pol I, whose amino acid sequences and crystal structures are known (Steitz, 1999).

The DNA polymerases have been used extensively in molecular biological research and thermostable DNA polymerases, such as *Taq* DNA polymerase were widely used in the polymerase chain reaction (PCR) (Saiki *et al*., 1988; Mullis *et al*., 1986). After that, many of the DNA Pol from other *Thermus* strains was studied. A typical Pol I protein consists N-terminal domain with a 5′→3′ exonuclease activity, a central domain with a 3′→5′exonuclease activity (or proofreading) and a C-terminal domain with DNA polymerase activity. However, DNA polymerases I, like bacteriophage T5 and T7 DNA polymerases, the N-terminal domain which contains 5′ nuclease activity is found as separate polypeptides. *Taq* and other DNA polymerases from the *Thermus* genus possess 5'-3' exonuclease (nick translation) activity but lack 3'-5' exonuclease proofreading activity of the *E. coli* homologue (Chien *et al*., 1976). On the other hand, a highly thermostable Pol I from the hyperthermophiles contains all three function of the *E. coli* Pol I (Perler *et al*., 1996). The enzyme has 3'-5' exonuclease activity dependent proofreading activity which is required for error correction during the polymerization. Several thermostable DNA polymerases with proofreading activity, such as Pfu, Vent, Deep Vent and Pwo have also been studied and introduce for high-fidelity PCR amplification (Lundberg *et al*., 1991; Frey *et al*., 1995).

A few moderately thermostable DNA polymerases have been isolated and purified from thermophilic *Bacillus* species (Akhmaloka *et al*., 2006; Perler *et al*., 1996). *Bst* DNA polymerase was isolated from *B. stearothermophilus* (Stenesh *et al*.,

1972; Kaboev *et al*., 1981; Sellman *et al*., 1992). *Bca* DNA polymerase was isolated and cloned from *B. caldotenax* (Sellman *et al*., 1992; Uemori *et al*., 1993). *Bst* DNA polymerase has been used for DNA sequencing. The polymerases I from different species also exhibit differences in other biochemical properties like specific activity, dideoxyribonucleotide triphosphate (ddNTP) sensitivity, strand displacement synthesis, and RNA-dependent DNA synthesis.

Most of the native enzymes are synthesized at very low levels by the thermophilic microorganisms, therefore, they are cumbersome to purify. Most of the thermostable DNA polymerases were produced in a biologically active form in *E.coli* expression system (Blóndal *et al*., 2001; Kim *et al*., 2002; Choi *et al*., 2004; Shin *et al*., 2005; Kim *et al*., 2007). However several problems persist, such as error-prone amplification and unwanted amplification at low temperatures. New and improved thermostable DNA polymerases are needed. Table 2.1 shows the list of current DNA polymerase in the market.

Table 2.1: Properties of the current DNA polymerase enzymes

	5'-3' Exonuclease	3'-5' Exonuclease	Strand displacement	Thermal stability	Primary applications
Bst DNA polymerase, large fragment	–	–	++++	+	Strand displacement applications
Deep Vent DNA polymerase	–	+++	++	++++	PCR (high fidelity)
E. coli DNA polymerase I	+	++	–	–	Nick translation
Klenow Fragment DNA polymerase I	–	++	–	–	Polishing ends
Phi29 DNA polymerase	–	++++	+++++	–	Strand displacement applications
T4 DNA polymerase	–	++++	–	+	Polishing ends, 2nd strand synthesis
T7 DNA polymerase	–	++++	+	–	Site-directed mutagenesis
Taq DNA polymerase	+	–	–	++	PCR (routine)
Vent DNA polymerase	–	++	++	+++	PCR (routine, high fidelity)
Pfu DNA polymerase	–	++++	–	++	PCR (routine, high fidelity, site-directed mutagenesis)

2.8 Polymerase chain reaction (PCR)

PCR is the in vitro enzymatic synthesis and amplification of specific DNA sequences. PCR technology began with the discovery of the first DNA polymerase around 1955. The enzyme was purified in 1958 (Lehman *et al*., 1958) but automation and modern PCR technology was not developed until 1983. American chemist Kary Mullis was struck by an idea when he was driving along a monotonous stretch of dark road one April weekend in 1983. Later to earn him the Nobel Prize: the principle of the polymerase chain reaction (Mullis, 1990). The basic PCR principle is simple. As the name implies, it is chain reaction: One DNA molecule is used to produce two copies, then four, then eight and so forth. This result in the exponential accumulation of the specific target fragment, approximately 2^n, where *n* is the number of cycles. This continuous doubling is accomplished by specific proteins known as thermostable polymerases, enzymes that are able to string together individual DNA building blocks to form long molecular strands. The reaction begins by mixing the polymerases with DNA (template), two suitable oligonucleotide primers, buffers and nucleotides in a tube and placed in the PCR machine (Saiki *et. al*., 1988).

Klenow fragment of *Escherichia coli* DNA polymerase I used to catalyze the extension of the annealed primer function best at 37°C in which they originate. Below this temperature the enzyme's activity declines steeply, above this temperature it is quickly destroyed. In PCR, however, the denaturation step requires the heat to separate the newly synthesized strands of the DNA in order to permit the primers to anneal to them. This is done by raising the temperature to around 95°C. As a result, fresh newly enzyme must be added during each cycle – a time

consuming, a tedious and error-prone process if several samples are amplified simultaneously (Saiki *et al*., 1988).

A solution was found in hot springs. Certain microorganisms thrive in such hot pools under the most inhospitable conditions, at temperatures that can reach 100°C. For example, thermostable *Thermos aquaticus* (*Taq*) polymerase can now replace the *E.coli* DNA polymerase that can survive extended incubation at 95°C (Mullis, 1990).

2.9 Isothermal assay

There is an increasing need for quantitative technologies suitable for molecular detection in a variety of settings for applications including pathogen detection or host gene of interest. Although PCR has been widely used by researchers, it requires thermocycler to cycle the temperature during amplification or elaborate methods for detection of the amplified product and this characteristic has been limited its application in the field (Demidov, 2005; Gill *et al*., 2008). Novel developments in molecular biology of DNA synthesis *in vivo* demonstrate the possibility of amplifying DNA in isothermal temperature without the need of a thermocycling apparatus (Karami *et al*., 2011). Therefore, isothermal assay can be run at any single temperature depending on the properties of polymerase being used.

There are several isothermal nucleic acid amplifications, such as rolling-circle amplification (RCA), loop-mediated isothermal amplification of DNA (LAMP), single primer isothermal amplification (SPIA), and helicase-dependant amplification (HDA) (Gill *et al*, 2008; Schweitzer and Kingsmore., 2001; Karami *et al*., 2011). All are very sensitive and compatible with many detection techniques, such as fluorescence, chemiluminescence, or gel electrophoresis (Schweitzer and

Kingsmore, 2001). However, these isothermal amplification technologies have advantages or weaknesses that limit their use in some aspects of molecular biology like PCR (Gill *et al*., 2008).

2.9.1 Rolling Circle Amplification (RCA)

RCA is an isothermal method that generates multiple copies of small, single stranded, circular DNA probes (Demidov, 2005). Linear RCA uses a single primer and results in the monotonous rolling out of long, repeated sequences of DNA with a gradual accumulation of RCA products (Demidov, 2005). The linear RCA is used in signal amplification on microarrays and detection of different DNA/RNA, protein, and other biomarkers (Nallur, *et al*., 2001; Marciniak *et al*., 2008; Gill *et al*., 2008). Exponential RCA or geometric RCA uses a pair of primers and results in a discrete set concatemeric double-stranded DNA (dsDNA) fragments (Demidov, 2002). The process allows amplification of circular DNA directly from cells or plaques, generating, or cloning (Reagin *et al*., 2003; Gill *et al*., 2008).

RCA is probably holds a distinct position in DNA diagnostics among other single-temperature amplification techniques due to its robustness and simplicity (Schweitzer and Kingsmore, 2001; Demidov, 2005; Gill *et al*., 2008). The success of RCA is dependant on strand displacement activity of DNA polymerases being used. As compared with RCA, all other isothermal methods of signal, probe, or target DNA amplification require prior assay optimization (Demidov, 2005; Gill *et al*., 2008). RCA is the most flexible and adaptable amplification methodology featuring merely few drawbacks but RCA assays also require certain caution to avoid possible contamination or false positives. Although RCA is described as isothermal

amplification systems, it requires an initial heat denaturation step if colony or plaque are being used as starting material (Reagin *et al*., 2003).

2.9.2 Loop-Mediated Isothermal Amplification (LAMP)

Loop-mediated isothermal amplification is a novel technique that amplifies DNA with high specificity, efficiency, and rapidity under isothermal condition (Notomi *et al*., 2000; Karami *et al*., 2011). The LAMP method requires a set of four or six specific designed primers and a DNA polymerase with strand displacement activity to produce a product with stem-loop structures (Gill *et al*., 2008; Karami *et al*., 2011). This method results in forming of white precipitate by magnesium pyrophosphate which will allows easy distinction of whether nucleic acid was successfully amplified (Gill *et al*., 2008). However, gel electrophoresis, real-time turbidimetry, and fluorescence probes can also be used for detection of LAMP products (Gill *et al*., 2008; Karami *et al*., 2011).

2.9.3 Single primer isothermal amplification (SPIA)

This method uses a single chimeric primer for amplification of DNA (SPIA) and RNA (Ribo-SPIA) (Karami *et al*., 2011). SPIA employs a single, target-specific chimeric primer composed of DNA at the 3' end and RNA at its 5' end, RNase H, and a DNA polymerase with a strong strand displacement activity (Gill *et al*., 2008). This method of amplification can be used for global genomic DNA amplification and for the amplification of specific genomic sequences and synthetic oligonucleotide DNA targets (Gill *et al*., 2008).

2.9.4 Helicase-Dependent Amplification (HDA)

This method is based on the unwinding activity of a DNA helicase ((Karami *et al*., 2011). This process utilizes a helicase to separate double stranded DNA to generate single-stranded templates for in vitro amplification of a target nucleic acid (Gill *et al*., 2008). Then, DNA polymerases extend the sequence-specific primers that hybridize with the template for amplification of the target sequence. HDA eliminate the need for thermo cycling equipment and initial heat denaturation thus, can be performed at any single temperature (Vincent *et al*., 2004). The results can be detected using gel electrophoresis, real-time format, and enzyme-linked immunosorbent assay (ELISA) (Gill *et al*., 2008; Karami *et al*., 2011).

CHAPTER 3.0: MATERIALS AND METHODS

3.1 Isolation and Identification of thermophiles

3.1.1 Sampling of mix water, biomats and sediments from hot spring

The spring samples were collected from Ulu Legong hot spring, Kedah in Malaysia. Water, biomats and sediments samples were collected from the main pool located in Ulu legong hot spring. There were two methods for sampling; first method, at each site, 1 mL of each sample recovered by sterile syringe was transferred to each of the bottles, consisting of 10 mL nutrient broth in a 30 ml sterilizes universal bottles. The sampling was conducted in triplicate. In the second method, the water, biomats, and sediments samples were collected using 3 bottles of 1 Liter sterile thermos flasks, respectively. These thermo flasks were transported to the laboratory and used without delay for inoculation in the nutrient broth medium. One milliliter sample from each water, biomats and sediments sample were added to 10 mL nutrient broth in universal bottles. The temperature and pH value of spring waters were measured at that time. All the samples in the universal bottles were incubated in water bath at relative temperature for 7 days. Growth was followed by measuring the increase in turbidity at 600 nm. Then, the culture was streaked onto a nutrient agar plate. Isolation of pure culture was done by using spread plate method and streak plate method recommended by Rath and Subramanyam (1998). Figure 3.1 shows the map of location of Ulu Legong Hot Spring located in Kedah, Malaysia and Figure 3.2 shows the pool of Ulu Legong Hot Spring with various sources of water, biomats, and sediments samples collected for the isolation of thermophiles for this