In silico ANALYSIS OF PLANT BIOMASS-DEGRADING GENES AND THERMOSTABLE ENZYMES FROM A MALAYSIAN HOT SPRING MICROBIOME USING TARGETED METAGENOMIC APPROACH

LEE LI SIN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

In silico ANALYSIS OF PLANT BIOMASS-DEGRADING GENES AND THERMOSTABLE ENZYMES FROM A MALAYSIAN HOT SPRING MICROBIOME USING TARGETED METAGENOMIC APPROACH

LEE LI SIN

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Lee Li Sin

Registration/Matric No: SGR150042

Name of Degree: Master of Science (Except Mathematics & Science Philosophy)

Title of Dissertation: *In silico* Analysis of Plant Biomass-Degrading Genes And Thermostable Enzymes From A Malaysian Hot Spring Microbiome Using Targeted Metagenomic Approach

Field of Study: Genetics and Molecular Biology

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation

In silico ANALYSIS OF PLANT BIOMASS-DEGRADING GENES AND THERMOSTABLE ENZYMES FROM A MALAYSIAN HOT SPRING MICROBIOME USING TARGETED METAGENOMIC APPROACH

ABSTRACT

Biomass decomposition using thermophilic enzymes has attracted attention due to their high reaction speed, thermostability, and decreased risk of contamination. Exploitation of efficient thermostable glycoside hydrolases (GHs) and auxiliary activities (AA) enzymes could accelerate the industrialisation of biofuels and biochemicals. Yet, the full spectrum of thermophiles and their enzymes that are important for biomass degradation at high temperatures are not well studied. In this work, a Malaysian hot spring located within a wooded area with fallen foliage which formed a thick layer of biomass bed under heated water represents a good resource for the discovery of microbial biomass decay communities. Here, the hypervariable regions of bacterial and archaeal 16S rRNA genes were sequenced on Illumina MiSeq platform using total community DNA extracted from the hot spring. Data suggested that 25 phyla, 58 classes, 110 orders, 171 families, and 328 genera inhabited this hot spring. Members of the genera *Acidimicrobium*, *Aeropyrum*, *Caldilinea*, *Caldisphaera*, *Chloracidobacterium*, *Chloroflexus*, *Desulfurobacterium*, *Fervidobacterium*, *Geobacillus*, *Meiothermus*, *Melioribacter*, *Methanothermococcus*, *Methanotorris*, *Roseiflexus*, *Thermoanaerobacter*, *Thermoanaerobacterium*, *Thermoanaerobaculum*, and *Thermosipho* were among the main thermophiles which harbour various GHs that may be involved in cellulose and hemicellulose breakdown. Among these communities, a variety of previously under-studied genera was found, revealing substantial diversity in foliage-rich hot springs.

Keywords: biofilm, biofuel, biomass degradation, cellulase, hot spring, thermophile

In silico ANALISIS GEN PENGURAIAN BIOJISIM TUMBUHAN DAN ENZIM TERMOSTABLE DARI MICROBIOME KOLAM AIR PANAS MALAYSIAN MENGGUNAKAN PENDEKATAN METAGENOMIK SASARAN

ABSTRAK

Penguraian biojisim yang menggunakan enzim thermophilic telah menarik perhatian disebabkan oleh kelajuan reaksi yang tinggi, thermostability, dan mengurangkan risiko pencemaran. Eksploitasi glycoside hydrolases (GHs) dan enzim auxiliary activities (AA) yang thermostable dan cekap boleh mempercepatkan perindustrian biopetrol dan biokimia. Namun, thermophiles dan enzim-enzimnya yang penting untuk degradasi biojisim pada suhu tinggi tidak dikaji dengan terperinci. Dalam kajian ini, kolam air panas Malaysian terletak di dalam kawasan yang berhutan dengan kejatuhan dedaun telah membentuk lapisan tebal katil biojisim di bawah air panas. Ia merupakan sumber yang baik untuk penemuan mikrob degradasi biojisim. Di sini, jujukan di bahagian hypervariable 16 rRNA gen untuk bakteria dan arkea telah dilaksanakan dengan menggunakan MiSeq platform Illumina yang memerlukan komuniti DNA yang diekstrak dari mikrob kolam air panas tersebut. Data mencadangkan 25 phyla, 58 kelas, 110 order, 171 keluarga, dan 328 genera mendiami kolam panas ini. Ahli-ahli dari genus *Acidimicrobium*, *Aeropyrum*, *Caldilinea*, *Caldisphaera*, *Chloracidobacterium*, *Chloroflexus*, *Desulfurobacterium*, *Fervidobacterium*, *Geobacillus*, *Meiothermus*, *Melioribacter*, *Methanothermococcus*, *Methanotorris*, *Roseiflexus*, *Thermoanaerobacter*, *Thermoanaerobacterium*, *Thermoanaerobaculum*, dan *Thermosipho* adalah antara thermophiles utama yang mengandungi pelbagai GHs yang mungkin terlibat dalam pecahan selulosa dan hemiselulosa. Antara komuniti tersebut, pelbagai genus yang tidak dikaji dengan terperinci sebelum ini telah ditemui, dan mendedahkan kepelbagaian mikrob dalam kolam air panas yang kaya dengan dedaun.

Kata kunci: biofilm, biofuel, penguraian biojisim, selulase, kolam air panas, thermophile

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Associate Professor Dr. Chan Kok Gan FASc and Associate Professor Dr. Geok Yuan Annie Tan, for their valuable guidance, useful comments, remarks, advice, and engagement throughout the learning process of this master programme. They inspired me greatly to work on this project. Their willingness to motivate me contributed tremendously to my project. Apart from that, I would like to express my sincere appreciation to my collaborator, Dr. Goh Kian Mau from Universiti Teknologi Malaysia, for his constant guidance and encouragement. I would also like to thank him for showing me some examples that related to the topic of my project. Furthermore, I would also like to take this opportunity to thank my lab manager, Ms. Yin Wai Fong, for her excellent financial management and coordination for all laboratory matters.

Besides that, I would like to thank Faculty of Science and High Impact Research Institute for providing me with a good environment and facilities to complete this project. Furthermore, it would be my pleasure to thank the examiners for providing me useful comments in completing this project, and Dr. Yong Kien Thai from Rimba Ilmu, Institute of Biological Sciences, Faculty of Science, University of Malaya for assistance in plant identification analysis. In addition to that, I would also like to thank all the postgraduate students in High Impact Research Institute which provide me valuable information as the guidance of my project.

Finally, an honourable mention goes to my family members and friends for their understandings and supports on me in completing this project. Without helps of the particulars that mentioned above, I would face many difficulties while doing this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

%	:	Percent
<	:	Less than
>	:	More than
×	:	Times
×g	:	Times gravity
≤	:	Less than or equal
≥	:	More than or equal
°C	:	Degree celcius
α	:	Alpha
β	:	Beta
μm	:	Micrometer

ABBREVIATIONS

AA	:	Auxiliary activity
ADF	:	Acid detergent fibre
ADL	:	Acid detergent lignin
AGE	:	Agarose gel electrophoresis
BacDive	:	The Bacterial Diversity Metadatabase
bp	:	Base pairs
CAZy	:	Carbohydrate-Active enZymes
cm	:	Centimetre
CTAB	:	Cetyltrimethylammonium bromide
DNA	:	Deoxyribonuclei acid

dsDNA	:	Double-stranded deoxyribonuclei acid
et al.	:	Latin word of “et alii” which means “and other”
g	:	Gram
GBS	:	Great Boiling Spring
GH(s)	:	Glycoside hydrolase(s)
HS	:	High sensitivity
i.e.	:	Latin term of “id est” which means “that is”
M	:	Molar
mg	:	Milligram
min	:	Minute(s)
mL	:	Milliliter
mM	:	Millimolar
n.a	:	Not applicable
NCBI	:	National Center for Biotechnology Information
NDF	:	Neutral detergent fibre
ng	:	Nanogram
NGS	:	Next generation sequencing
OTUs	:	Operational taxonomic units
PBS	:	Phosphate-buffered saline
PCoA	:	Principal coordinates analysis
PCR		Polymerase chain reaction
pg	:	Picogram
PICRUSt	:	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
psi	:	Pounds per square inch
QIIME	:	Quantitative Insights into Microbial Ecology

qPCR	:	Real-time PCR
SDS	:	Sodium dodecyl sulfate
sec	:	Second(s)
SK-Y	:	Y-shaped Sungai Klah
spp.	:	Species
TBE	:	Tris-boric acid ethylenediaminetetraacetic acid
TCA	:	Citrate cycle
TCU	:	True colour unit
v/v	:	Volume/volume
w/v	:	Weight/volume
YNP	:	Yellowstone National Park

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Lignocellulolytic biomass is one of the main sources for second generation biofuel production due to its renewable, sustainable, and abundance (Xia et al., 2013). Plant biomass constitutes mainly lignin, cellulose, and hemicellulose (Van Dyk & Pletschke, 2012). Biomass decomposition enzyme cocktails from fungi have becoming a crucial source in saccharification step (Walton et al., 2011). Presently, most commercial enzyme cocktails for large scale industrial usage are products of various fungi. Despite several companies offer these cocktails, there are limitations of these enzymes to meet the growing demand for the economic viability as these enzyme preparations are highly similar in activities and composition as the strains of fungi are similar (Banerjee et al., 2010a). In the initial development of fungi based enzyme cocktails, many of these commercial cellulase enzyme cocktails are ineffective in releasing free glucose from the pre-treated biomass without complementary of β-glucosidase (Banerjee et al., 2010a; Qing & Wyman, 2011). As many fungal enzymes are glycosylated, heterologous expression generally results abnormal glycosylation that adversely affect enzyme properties (Banerjee et al., 2010a). As such, the insufficient supplies of active and thermostable enzymes that efficiently deconstruct lignocellulosic biomass represent a major bottleneck for the industrial-scale conversion of lignocellulosic materials to biofuel (Chang & Yao, 2011).

Thermophilic prokaryotes with lignocellulolytic activities offer several advantages over mesophiles. The ability of thermophiles and its enzymes (thermozymes) to remain active at high temperature can improve the biomass decomposition reactions. Operate the decomposition process at higher temperature facilitates the downstream recovery due to the alcoholic products (i.e., bioethanol) are easily vaporised at high temperature (≥50°C) (Vishnivetskaya et al., 2015). Apart from that, conducting

bioprocessing at higher temperature minimises feedstock-derived contamination. The process is more cost-effective as cooling after pre-treatment process could be avoided (Taylor et al., 2009).

There are several efforts to mine new thermostable enzymes or whole thermophilic cells for lignocellulolytic degradation. Bacteria class namely Clostridia and its order Thermoanaerobacterales are widely studied for lignocellulosic ethanol production (Chang & Yao, 2011). Besides, *Deinoccocus-Thermus* spp. (Wu et al., 2015), *Geobacillus* spp. (Brumm et al., 2015b; Brumm et al., 2015a), *Melioribacter* spp. (Rakitin et al., 2015), and *Thermoanaerobacterium* spp. (Currie et al., 2014) are well known for deconstructing biomass at high temperature (Bhalla et al., 2013). Furthermore, *Thermotoga* (Yu et al., 2016), *Rhodothermus* (Keshk, 2016), and *Caldicellulosiruptor* (Peng et al., 2015) have shown to produce hyperthermostable cellulases with an optimum hydrolysis temperature above 80°C. A β-glucosidase from thermophilic *Anoxybacillus* sp. was shown to be glucose-tolerant and this characteristic is important to avoid product inhibition during saccharification process (Chan et al., 2016). Recently, the genome of a new thermophilic *Rhodothermaceae* strain RA isolated from a hot spring was sequenced, and several protein sequences for biomass degradation were identified (Goh et al., 2016).

Thermophilic bacteria and certain archaea offer a 'blue ocean' opportunity for harnessing novel lignocellulolytic degradation strains or enzymes. Traditionally, in order to clone a recombinant enzyme, the microorganism is isolated, the expression strain is selected, gene sequence is identified, and then the gene is cloned. This process could be costly and time consuming. In addition, very small fraction (<10%) of environmental microorganisms could be cultivated using current isolation and culturing techniques; consequently, this has severely reduced the discovery of microbial resources and novel

enzymes with superior qualities, which could be exploited for biomass deconstruction. Besides, easily cultivable thermophiles may not represent those that dominate natural microbiota (Urbieta et al., 2015). Compared to culture-dependent studies, culture-independent studies represent more inclusive approach for biodiversity analysis (Urbieta et al., 2015). Rather than applying conventional culture-dependent approaches for gene discovery, this project is mainly focused on the gene mining using culture-independent approach involving 16S rRNA gene amplicon based sequencing from Malaysian hot spring (60-90°C, mean 68°C, pH 8.6). Interestingly, thick layer of plant litter (mainly foliage) forms biomass bed that submerged and decomposed in this hot spring. This study provides insight on the list important thermophiles and their thermozymes for degradation of biomass at high temperature. These novel genes may provide the key to the solution of the biofuel production bottleneck. Moreover, their potential use in biomass deconstruction cocktails is worth investigating.

The objectives of this study are to:

i. To identify thermophilic plant litter microbiome from a Malaysian hot spring.

ii. To conduct *in silico* analysis of the microbiome for plant biomass-degrading genes and thermostable enzymes.

CHAPTER 2: LITERATURE REVIEW

Thermophiles are able to tolerate high temperatures or require high temperatures for their survival and growth. According to Brock (2012), thermophile is "a microorganism capable of living at temperatures at or near the maximum for the taxonomic group of which it is a part." Their optimum growth temperature can be ranged from 50°C to 121°C. These microorganisms have been classified into moderate thermophiles (optimally at 50°C or higher), extreme thermophiles (65-79°C), and hyperthermophiles (>80°C) (Wagner & Wiegel, 2008). Although many phylogenetically distinct groups of thermophiles have been isolated using culture-dependent method, unforeseen diversity of thermophiles survived in high temperature environments have been noticed via several techniques such as 16S rDNA amplification, direct microscopic examination, and other culture-independent methods. Hence, this shows that there are many more non-culturable thermophiles yet to be found (Satyanarayana et al., 2013).

Natural biotopes (i.e. terrestrial geothermal and volcanic areas, and deep-sea hydrothermal vents) for the survival of thermophiles are distributed everywhere. These habitats are relatively high-temperature than the man-made. Many extremophiles and hyperthermophiles have been discovered from these habitats by culture-dependent and culture-independent methods. Geothermal and volcanic areas consist of terrestrial fumaroles, terrestrial hot springs, and geysers. Others natural biotopes include geothermally heated oil and petroleum reservoirs, and sun-heated soils/sediments (Satyanarayana et al., 2013).

Many extremophiles or hyperthermophiles have been isolated from terrestrial hot springs (Brumm et al., 2015b; Brumm et al., 2015a; Goh et al., 2016; Hamilton-Brehm et al., 2010; Losey et al., 2013). Various culture-dependent and culture-independent studies

have been performed to study their diversity in hot springs (Chan et al., 2015; Gladden et al., 2011; Gladden et al., 2012; Mhuantong et al., 2015; Park et al., 2012; Peacock et al., 2013; Vishnivetskaya et al., 2015; Xia et al., 2012; Xia et al., 2014). Generally, less than 1% of the extant microbial "species" can been cultured and most of the microorganisms cannot grow in conventional isolation media (Satyanarayana et al., 2013).

2.1 Metagenome-derived thermozymes

Extremophiles or hyperthermophiles are robust microorganisms which are able to produce thermostable enzymes, and adapt changes in environmental conditions, such as pH and temperature (Barnard et al., 2010). There are many advantages in operating bioprocesses under high temperature (>50°C) (Georgieva et al., 2008; Taylor et al., 2009). For instance, usage of antibiotics in industrial processes to prevent biological contamination can result in higher cost and negative impacts to the environment (Skinner & Leathers, 2004; Taylor et al., 2009). However, high temperature can be used instead of antibiotics to eliminate or minimize microbial contamination (Mirete et al., 2016). In industrial processes, thermozymes possess certain advantages over their mesophilic counterparts as high temperature could speed up chemical reaction rate (Sarmiento et al., 2015). Moreover, high temperature reduces energy input by avoiding cooling step between pre-treatment and hydrolysis of lignocellulosic feedstock, and it increases solubility and efficient mixtures of substrate (Lynd et al., 2008; Taylor et al., 2009; Turner et al., 2007). Some thermozymes are also highly resistant to denaturing agents and organic solvents (Fan et al., 2011; Roh & Schmid, 2013). In addition, high temperature facilitates downstream product recovery whereby thermozymes are easily separated from heat-labile proteins during purification steps as reported by Pessela et al. (2004).

2.2 Enzymes required to degrade lignocellulose

A wide variety with different specificities of enzymes are needed to decompose different layers/components of lignocellulose feedstocks. Generally, there are three types of enzymes are required to hydrolyse cellulose into glucose monomers which known as endo-1,4-β-glucanases (EC 3.2.1.4), exo-1,4-β-glucanases (EC 3.2.1.91) or cellobiohydrolase (EC 3.2.1.176), and β-glucosidases (EC 3.2.1.21) (Bayer et al., 1998; Demain et al., 2005; Schwarz, 2001; Wilson, 2009; Wilson, 2011). Endoglucanases act randomly on internal glucosidic linkages, in the amorphous portion of cellulose, releasing oligosaccharides with several polymerization degrees. Cellobiohydrolases degrade cellulose by removing cellobiose molecules; they can act on the crystalline portion of cellulose and attack from the reducing and non-reducing ends of the glucose chain (van den Brink & de Vries, 2011). Exoglucohydrolases are responsible for removal of glucose monomers from the non-reducing ends of cyclodextrins. Finally, β-glucosidases hydrolyse cellobiose into glucose and also remove glucose monomers from non-reducing ends of small cyclodextrins (Maitan-Alfenas et al., 2015). Table 2.1 shows a brief overview of different types of enzymes required to degrade complex lignocellulose substrates.

Table 2.1: Types of enzymes required to degrade lignocellulose to monomers (Parisutham et al., 2014)

Enzymes	Cleavage target
Lignin degradation	
Lignin peroxidase	Oxidatively degrades lignin
Aryl-alcohol oxidase	Oxidation of aromatic alcohol of lignin
Laccase	Degrades phenolic portion of lignin
Glyoxal oxidase	Oxidation of glyoxal
Cellobiose dehydrogenase	Oxidize cellobiose and cellodextrins to aldonolactones
Hemicellulases	
Endoxylanase	Hydrolyzes glycosidic bond of glucuronoxylan or arabinoxylan
β-Xylosidase	Hydrolyze xylobiose or xylooligosaccharide from xylan
Acetyl xylan esterase	Deacetylates the backbone glycosyl unit of xylan
Feruloyl esterase	Hydrolyzes feruloyl esters of arabinan/arabinoxylan, rhamnogalacturonan or xyloglucan
Glucuronoyl esterase	Demethylates xylan backbone of glucuronoarabinoxylan

Table 2.1, continued.

Arabinofuranosidase	Removes arabinose from arabinoglucuronoxylan or arabinoxylan
Galactosidase	Removes galactose from galactomannan or pectin
Glucuronidase	Removes glucuronoyl or its methyl ester from xylan
Mannanase	Degrades mannosyl polymer of glucomannas or galactomannas
Xyloglucan hydrolase	Degrades glucan of xyloglucan
Cellulases	
Cellobiohydrolase	Hydrolyzes reducing or non-reducing end of cellulose chain
Endoglucanase	Hydrolyzes random internal glycosidic bonds of cellulose
β-Glucosidase	Cleaves non-reducing end of cellobiose and cellodextrin
Phospho-β-glucosidase	Cleaves phosphorylated cellobiose and cello-dextrins

2.3 Plant cell wall degradation using auxiliary activities

Auxiliary Activities (AA) acts as catalytic modules in plant cell wall degradation (Levasseur et al., 2013). AA has been classified as a new family of enzymes that cleave complex carbohydrates with the potential ability to aid the glycoside hydrolases, polysaccharide lyases, and carbohydrate esterases gain access to the cellulose and hemicellulose encrusted in the plant cell wall (Levasseur et al., 2013). Therefore, AA category is a large class of modules which targets on variety of catalytic reactions or multiple substrates. Up to date, AA class is split into 13 families, and includes subfamilies for families AA1, AA3, and AA5 (Table 2.2). Subfamilies are designated using the family name plus a suffix indicating the subfamily (Levasseur et al., 2013).

Table 2.2: Subfamily division and known activities in the AA families (Levasseur et al., 2013)

(Sub) families	Known activities	EC number	Mechanism	References
AA1	Multicopper oxidase			
AA1_1	Laccase	EC 1.10.3.2	Catalyses the one-electron oxidation of phenolics, aromatic amines, and other electron-rich substrates via the reduction of oxygen to water.	(Baldrian, 2006)
AA1_2	Ferroxidase	EC 1.10.3.2	Catalyses the oxidization of iron II to iron III.	(Levasseur et al., 2013)
AA1_3	Laccase-like multicopper oxidase	EC 1.10.3.2	n.a	n.a
AA2	Class II peroxidase			
	Manganese peroxidase	EC 1.11.1.13	Oxidizes Mn(II) to Mn(III) which in turn oxidizes a variety of phenolic model compounds able to degrade and/or modify lignin polymers.	(Levasseur et al., 2013)
	Lignin peroxidase	EC 1.11.1.14	Oxidizes phenolic aromatic substrates and catalyses oxidative cleavages of C–C bonds and ether bonds in high-redox-potential non-phenolic aromatic substrates.	(Levasseur et al., 2013)
	Versatile peroxidase	EC 1.11.1.16	Oxidizes Mn(II), phenolic, and non-phenolic substrates.	(Levasseur et al., 2013)
AA3	GMC oxidoreductase			
AA3_1	Cellobiose dehydrogenase	EC 1.1.99.18	Uses electron acceptors to oxidize soluble cellodextrins to the corresponding lactones.	(Henriksson et al., 2000; Turbe-Doan et al., 2013)
AA3_2	Aryl-alcohol oxidase / Glucose oxidase	EC 1.1.3.7 / 1.1.3.4	Aryl-alcohol oxidase catalyses the oxidative dehydrogenation of several aromatic and aliphatic polyunsaturated alcohols with an α-carbon primary hydroxyl group, with the concomitant reduction of O_2 to H_2O_2; glucose oxidase oxidizes the hydroxyl group at the C1 position of sugars.	(Levasseur et al., 2013)
AA3_3	Alcohol oxidase	EC 1.1.3.13	Catalyses the oxidation of primary aliphatic alcohols to their respective aldehydes.	(Levasseur et al., 2013)
AA3_4	Pyranose oxidase	EC 1.1.3.10	Catalyses the C2-oxidation of $_D$-glucopyranose and structurally analogous mono- and di-saccharides to the corresponding sugars while reducing O_2 to H_2O_2.	(Giffhorn, 2000)
AA4	Vanillyl alcohol oxidase	EC 1.1.3.38	Catalyses the conversion of a wide range of phenolic compounds bearing side chains at the para-position of the aromatic ring.	(de Jong et al., 1992)
AA5	Radical-copper oxidase			
AA5_1	Glyoxal oxidase	EC 1.2.3.15	Oxidizes simple aldehydes to the corresponding carboxylic acids.	(Whittaker et al., 1996)
AA5_2	Galactose oxidase	EC 1.1.3.9	Catalyses the oxidation of a wide range of carbohydrates (including galactose) but also primary alcohols into the corresponding aldehydes, with the reduction of O_2 into H_2O_2.	(Whittaker, 2003)
AA6	1,4-Benzoquinone reductase	EC 1.6.5.6	Involves in a quinone redox cycle that generates extracellular Fenton reagents.	(Brock et al., 1995; Jensen Jr et al., 2002; Lee et al., 2007)

Table 2.2, continued.

(Sub) families	Known activities	EC number	Mechanism	References
AA7	Glucooligosaccharide oxidase	EC 1.1.3.-	Oxidizes the C1 hydroxyl groups of β-1,4-linked sugars through two half-reactions featuring the oxidation of the reducing sugar to the lactone and spontaneous hydrolysis to the corresponding acid; oxidizes a variety of carbohydrates ($_D$-glucose, maltose, lactose, cellobiose, malto- and cello-oligosaccharides) with the concomitant reduction of O_2 to H_2O_2.	(Fan et al., 2000; van Hellemond et al., 2006)
AA8	Iron reductase domain	-	Stimulates the reduction of Fe(III) by acting as an electron sink or by directly reducing Fe(III); generation of highly reactive hydroxyl radicals ($OH^{\bullet}$) via Fenton's reaction.	(Eastwood et al., 2011; Henriksson et al., 1993)
AA9	Lytic polysaccharide monooxygenase (GH61)	-	Oxidizes the C1 carbon of the glucose ring structure but may also be able to oxidize C4 and C6.	(Beeson et al., 2011; Bey et al., 2013; Phillips et al., 2011)
AA10	Lytic polysaccharide monooxygenase (CBM33)	-	Cleaves polysaccharide chains in crystalline chitin and cellulose to produce aldonic acids.	(Forsberg et al., 2011; Vaaje-Kolstad et al., 2005)
AA11	Lytic polysaccharide monooxygenase	-	Cleavage of chitin chains with oxidation of C-1 has been demonstrated for a AA11 LPMO from *Aspergillus oryzae*.	(Hemsworth et al., 2014)
AA12	pyrroloquinoline quinone-dependent oxidoreductase	-	Activity was demonstrated for the CC1G_09525 protein of *Coprinopsis cinerea*.	(Matsumura et al., 2014)
AA13	Lytic polysaccharide monooxygenase	-	Cleavage of starch with oxidation of C-1 at the site of cleavage has been demonstrated for the LPMO encoded by gene NCU08746 from *Neurospora crassa*.	(Vu et al., 2014)

Note: Latest update on September 2017.

2.4 Biotechnological applications of thermophiles and thermozymes

The adaptation of thermophiles to the harsh habitats (i.e. hot springs and deep-sea hydrothermal vents) explains the high genomic and metabolic flexibility of microbial diversity in these ecosystems which causes them and its thermostable proteins highly suitable for some industrial and biotechnological applications (Badhai et al., 2015). Therefore, mining of novel biocatalysts from extremophiles has become an important field. In recent years, many thermozymes such as novel thermostable polymerases (Schoenfeld et al., 2013), beta-galactosidases (Wang et al., 2014), esterase (Fuciños et al., 2014), and xylanases (Shi et al., 2013), have been discovered and characterized to explore a new horizon in biotechnology.

2.4.1 Bioenergy – Biofuels

Plant biomass is the most plentiful renewable biomass on earth and is considered as an attractive raw material of bioenergy and bio-based chemicals (de Gonzalo et al., 2016). The biotechnological conversion of lignocellulose into different carbohydrates, including glucose, is the basis for the generation of bioethanol, carbohydrates and aromatic products (Asgher et al., 2014; Kawaguchi et al., 2016; Ragauskas et al., 2014). In order to optimize the use of plant biomass through biorefining, lignin degradation has become a key target in recent years. Efficient and cost-effective methods for selective lignin degradation are in high demand (de Gonzalo et al., 2016).

As lignin is naturally recalcitrant to biological degradation, the production of bioethanol from lignocellulosic biomass always involves harsh methods such as thermo-mechanical and chemical pre-treatments to enhance material accessibility, and followed by chemical-based or enzyme-catalysed hydrolysis to release soluble sugars (Balat, 2011; Mussatto et al., 2010). Although chemical hydrolysis is best suited to industrial-scale

continuous processing of biomass, none of these methods have been commercialized due to cost-ineffective of pre-treatment processes, and production of microbial growth inhibitory compounds (i.e. phenols, organic acids, furfurals and/or hydroxylmethyl furfurals) in most thermal and acid pre-treatment processes which remarkably inhibit fermentation (Balat, 2011; Taherzadeh & Karimi, 2007; Yang & Wyman, 2008). Hence, enzymatic hydrolysis of polysaccharides has been viewed as the key of cost effective in long term production of bioethanol (Zhang et al., 2012). It is more selective towards sugar production as it is very mild process, high yields, and the maintenance costs are low compared to chemical-based hydrolysis (Kuhad et al., 1997; Saini et al., 2015).

2.5 Diversity analysis of thermophiles

Conventional cultivation methods have been important in characterizing the biochemical and physiological properties of pure cultures, or in developing whole-cell applications. Nevertheless, the number of cells that can be visualised using microscopy are higher than the numbers of colonies grown on plates (Urbieta et al., 2015). This explains that we underestimated the diversity existed in the environment using culture dependent method. Moreover, easily cultivable thermophiles may not represent those that dominate natural microbiota (Urbieta et al., 2015). Thus, culture-independent methods involving 18S or 16S rRNA gene sequencing represent a more comprehensive approach for biodiversity analysis. The universality of the 16S rRNA genes makes them an ideal target for phylogenetic analysis and taxonomic classification (Olsen et al., 1986).

Next-generation sequencing (NGS) also known as second-generation sequencing, has accelerated the 16S rRNA-gene based biodiversity studies as more samples can be analysed at a lower sequencing cost (Urbieta et al., 2015). Partial 16S rRNA gene sequences is often used for the biodiversity analysis as the read length of most NGS

sequencer platforms is relatively short, (Urbieta et al., 2015). Generally, most 16S rRNA genes contains nine conserved regions (C1-C9) and nine variable regions (V1-V9) (Urbieta et al., 2015). Thus, degenerate or non-degenerate primers are designed according the targeted conserved regions, which also known as targeted metagenome (Urbieta et al., 2015).

Biodiversity analysis using partial 16S rRNA-based metagenome sequences is not restricted only for microbial diversity study in hot springs but also with coastal environments, soil samples, municipal wastewater treatment plants, tongue coating microbiomes, and the mouse gut (Cai et al., 2013; Fierer et al., 2012; Inskeep et al., 2013; Jiang et al., 2012; Lee et al., 2010; Somboonna et al., 2012). Although the strategies of targeted metagenomics can be used to infer the taxonomic diversity of the community (16S rRNA gene profiling) or particular aspects of its functional diversity, a broader view of functional diversity, i.e., a more exhaustive answer to the question "what are they doing," is provided by shotgun metagenomics (DeCastro et al., 2016).

2.6 Functional analysis of thermophiles

Currently, advanced development of high throughput sequencing methods has increased the understanding of microbial ecology which could enhance the widespread of biotechnological applications from personalized medicine to bioenergy (Cho & Blaser, 2012; Suen et al., 2010). As an example, metagenomics has been used to identify the potential enzymes for use in biofuels production. Many novel enzymes have been characterized using this approach, including amylolytic enzymes, β-glucosidase for bioethanol, endoglucanase, lignases, lipolytic enzymes for biodiesel, and xylanase (Xing et al., 2012).

The taxonomic composition and phylogenetic diversity of environmental samples are often characterized using gene markers such as the 16S rRNA gene of bacteria and archaea. Nevertheless, metabolic or other functional capabilities of the microorganisms cannot be directly identified based on the marker gene studies or focus on one or a few universal genes only (Langille et al., 2013). In turn, shotgun metagenomic aims to extract all the genes from a community and generate detailed metabolic and functional profiles. However, the sequencing cost is extremely expensive for shotgun metagenomic sequencing across many samples (Langille et al., 2013).

Despite the fact that marker gene and shotgun sequencing approaches differ in the type of information generated, phylogeny and biomolecular function are strongly, if imperfectly, correlated. Phylogenetic trees based on 16S marker gene closely resemble clusters obtained on the basis of shared gene content (Konstantinidis & Tiedje, 2005; Segata & Huttenhower, 2011; Snel et al., 1999; Zaneveld et al., 2010), and researchers frequently infer properties of uncultured microorganisms from cultured relatives. For instance, the genome of a *Bacteroides* spp. might reasonably be inferred to contain many genes encoding glycoside hydrolase activity, based on the commonality of these activities in sequenced *Bacteroides* isolates (Xu et al., 2007). This interrelation is in turn closely related to the pan- and core-genomes of each phylogenetic subtree (Collins & Higgs, 2012), in that larger and more strongly conserved core genomes result in more confident linkages of genes with clades.

Conversely, a clade's core genome consists of genes its members can be expected with high probability to carry in their genomes. The correlation between phylogeny and functional attributes depends on factors including the complexity of the trait (Martiny et al., 2013), but the overall degree of correlation suggests that it may be fruitful to predict

the functions encoded in a microorganism's genome on the basis of functions encoded in closely related genomes (Langille et al., 2013).

Some 16S studies have extended these intuitions to infer the functional contribution of particular community members by mapping a subset of abundant 16S rDNA gene sequences to their nearest sequenced reference genome (Barott et al., 2012; Morgan et al., 2012; Muegge et al., 2011). The accuracy of such approaches has not been evaluated, but the correlation between gene content and phylogeny (Chaffron et al., 2010; Konstantinidis & Tiedje, 2005; Zaneveld et al., 2010) (excepting special cases such as laterally transferred elements and intracellular endosymbionts with reduced genomes) suggests that it may be possible to approximately predict the functional potential of microbial communities from phylogeny (Langille et al., 2013).

CHAPTER 3: MATERIALS AND METHODS

3.1 Materials

3.1.1 Equipment and Instruments

Instruments used during course of this study included -20°C freezer (Liebherr, UK); 2100 Bioanalyzer (Agilent Technologies, USA); 4°C chiller (Thermo Scientific, USA); -80°C freezer (Gaia Science, Singapore); agarose gel electrophoresis (AGE) (Biorad, USA); autoclave machine (Hirayama, USA); centrifuge machine (Eppendorf, North America); eco™ real-time PCR system (Illumina, USA); eco-spin microcentrifuge (Elmi, Latvia); gel documentary image analyzer (UVP, USA); ice maker (Nuove Technologie Del Freddo, Italy); laminar flow cabinet (Esco Technologies, USA); milli-Q® integral water purification system (Merck, Germany); Illumina MiSeq sequencer (Illumina, USA); nanodrop spectrophotometer (Thermo Scientific, USA); pH meter (Sartorius, Germany); polymerase chain reaction (PCR) T100 thermal cycler (Biorad, USA); Qubit® 2.0 fluorometer (Invitrogen, USA); sonicator (Branson Ultrasonics, USA); thermomixer (Eppendorf, North America); water bath (Benchmark, USA); and weighing machine (Sartorius, Germany).

Equipment used in this study included laboratory glassware (beakers, conical flasks, glass beads, measuring cylinders, Schott's bottles); microcentrifuge tubes (1.5 mL and 2.0 mL); syringe (Terumo, USA); syringe filter (0.22 μm pore size) (Sartorius, Germany); polypropylene tubes (15 mL and 50 mL); pipettes (Eppendorf, North America); and pipette tips (Eppendorf, North America).

3.1.2 Commercial kits

The commercial kits used in the study are described in Table 3.1.

Table 3.1: Commercial kits used in this study

Kit (Manufacturer)	Application
Agilent High Sensitivity kit (Agilent Technologies, USA)	Sizing and quantitation of DNA sequencing libraries
Agencourt AMPure XP System (Beckman Coulter, USA)	Purify the extracted metagenome libraries and DNA sequencing libraries
KAPA SYBR Fast qPCR Master Mix (KAPA Biosystems, USA)	Real-time PCR (qPCR) amplification
Nextera XT Index Kit (Illumina, USA)	Append sequencing adapters and indexes to DNA sequencing libraries
2x KAPA HiFi HotStart ReadyMix Kit (KAPA Biosystems, USA)	Amplification of next-generation sequencing libraries prepared for Ilumina sequencing
MiSeq Reagent Kit V2 (2 × 250 base pairs) and V3 (2 × 300 base pairs) (Illumina, USA)	For next-generation sequencing
Qubit® dsDNA HS Assay Kit (Life Technologies, USA)	Quantify the concentration of dsDNA ranging from 10 pg/μL to 100 ng/μL

3.1.3 Commercial reagents

The commercial reagents used in this study are 1 kb DNA marker (Promega, USA); elution buffer (Qiagen, USA); GelStar™ Nuclei Acid Gel Stain (Lonza, USA); lysozyme (Sigma-Aldrich, USA); phosphate-buffered saline (Merck, Germany); proteinase K (Qiagen, USA); RNase A (Qiagen, USA); and Tris-EDTA buffer (Epicentre, USA).

3.1.4 Chemical reagents

All the chemical reagents used in this study are analytical grade which purchased from Merck, Germany; Thermo Scientific, USA; Sigma-Aldrich, USA; Promega Ltd, USA; Amresco, USA; and Invitrogen Corp., USA. Solvents used in this work were supplied by Merck, Germany and Thermo Scientific, USA.

3.1.5 Buffer solutions

3.1.5.1 Modified cetyltrimethylammonium bromide (CTAB) lysis buffer

The modified CTAB lysis buffer was prepared by 100 mM Tris-HCl, 100 mM EDTA, 100 mM K_2HPO_4, 1.5 M NaCl, and 1% w/v CTAB. The buffer was filter-sterilized with minisart syringe filter (0.22 μm pore size).

3.1.5.2 Tris-boric acid ethylenediaminetetraacetic acid (TBE)

The 10× TBE stock solution was prepared by mixing 108 g Tris base, 55 g boric acid, and 50 mL 0.5 M $Na_2EDTA.2H_2O$ in 1 L of deionized water with pH adjusted to 8.0 before subjected to autoclave at 15 psi, 121°C for 15 min.

3.1.6 Stock solution

3.1.6.1 Preparation of 10% (w/v) sodium dodecyl sulfate (SDS) solution

Briefly, 100 mL of 10% SDS solution was prepared by weighing 10 g SDS in 250 mL Schott's bottle. Then, 80 mL deionized water was added and mixed well. The volume was adjusted to 100 mL with deionized water. The solution was filter-sterilized with minisart syringe filter (0.22 μm pore size) before use.

3.1.7 Agarose gel electrophoresis (AGE)

The extracted metagenome was separated by AGE using 1% (w/v) agarose gel. Firstly, 5 μL of metagenome samples were mixed with 1 μL of 6× bromophenol blue loading dye onto an agarose gel along with 1 kb DNA marker. The loaded gel was subjected to electrophoresis in 1× TBE buffer at 75 V until the loading dye front approached about 1 cm from the edge of the gel (approximately 60 minutes). The gel was pre-stained with 0.5× GelStar™ Nuclei Acid Gel Stain (Lonza, USA). The stained agarose gel was visualized and digitally photographed with gel documentary image analyzer (UVP, USA).

3.1.8 Polymerase Chain Reaction (PCR) master mix for 16S rRNA gene amplification and DNA libraries purification

PCR and DNA libraries purification were carried out in accordance to Illumina 16S rRNA Metagenomic Sequencing Library Preparation Guide. The thermocycling conditions were programmed and performed according to the manufacturer's instructions.

Table 3.2: Amplicon PCR master mix preparation using region of interest-specific primers with overhang adapters attached

Component	Volume (µL)
Microbial DNA (5 ng/µl)	2.5
Amplicon PCR Forward Primer 1 µM	5
Amplicon PCR Reverse Primer 1 µM	5
2x KAPA HiFi HotStart ReadyMix	12.5
Total	**25**

Table 3.3: Thermocycling conditions for amplicon PCR

Step		Temperature	Duration
Initial denaturation		95°C	3 min
25 cycles	Denaturation	95°C	30 sec
	Annealing	55°C	30 sec
	Extension	72°C	30 sec
Final extension		72°C	5 min
Hold		4°C	-

Table 3.4: Amplicon PCR clean-up

Component	Volume (µL)
10 mM Tris pH 8.5	52.5 µl per sample
AMPure XP beads	20 µl per sample
Freshly Prepared 80% Ethanol (EtOH)	400 µl per sample
Amplicon PCR product	25 µl per sample

Table 3.5: Index PCR master mix preparation with dual indices and Illumina sequencing adapters using Nextera XT Index Kit (Illumina, USA)

Component	Volume (µL)
Amplicon PCR product	5
Nextera XT Index Primer 1 (N7xx)	5
Nextera XT Index Primer 2 (S5xx)	5
2x KAPA HiFi HotStart ReadyMix	25
PCR Grade water	10
Total	**50**

Table 3.6: Thermocycling conditions for index PCR

Step		Temperature	Duration
Initial denaturation		95°C	3 min
8 cycles	Denaturation	95°C	30 sec
	Annealing	55°C	30 sec
	Extension	72°C	30 sec
Final extension		72°C	5 min
Hold		4°C	-

Table 3.7: Index PCR clean-up

Component	Volume (µL)
10 mM Tris pH 8.5	27.5 µl per sample
AMPure XP beads	56 µl per sample
Freshly Prepared 80% Ethanol (EtOH)	400 µl per sample
Amplicon PCR product	50 µl per sample

3.1.9 Eco™ real-time PCR (qPCR) master mix for library quantification

Eco™ qPCR was carried out in accordance to Eco™ Real-Time PCR System User Guide. The thermocycling conditions were programmed and performed according to the manufacturer's instructions.

Table 3.8: Eco™ qPCR master mix preparation for library quantification

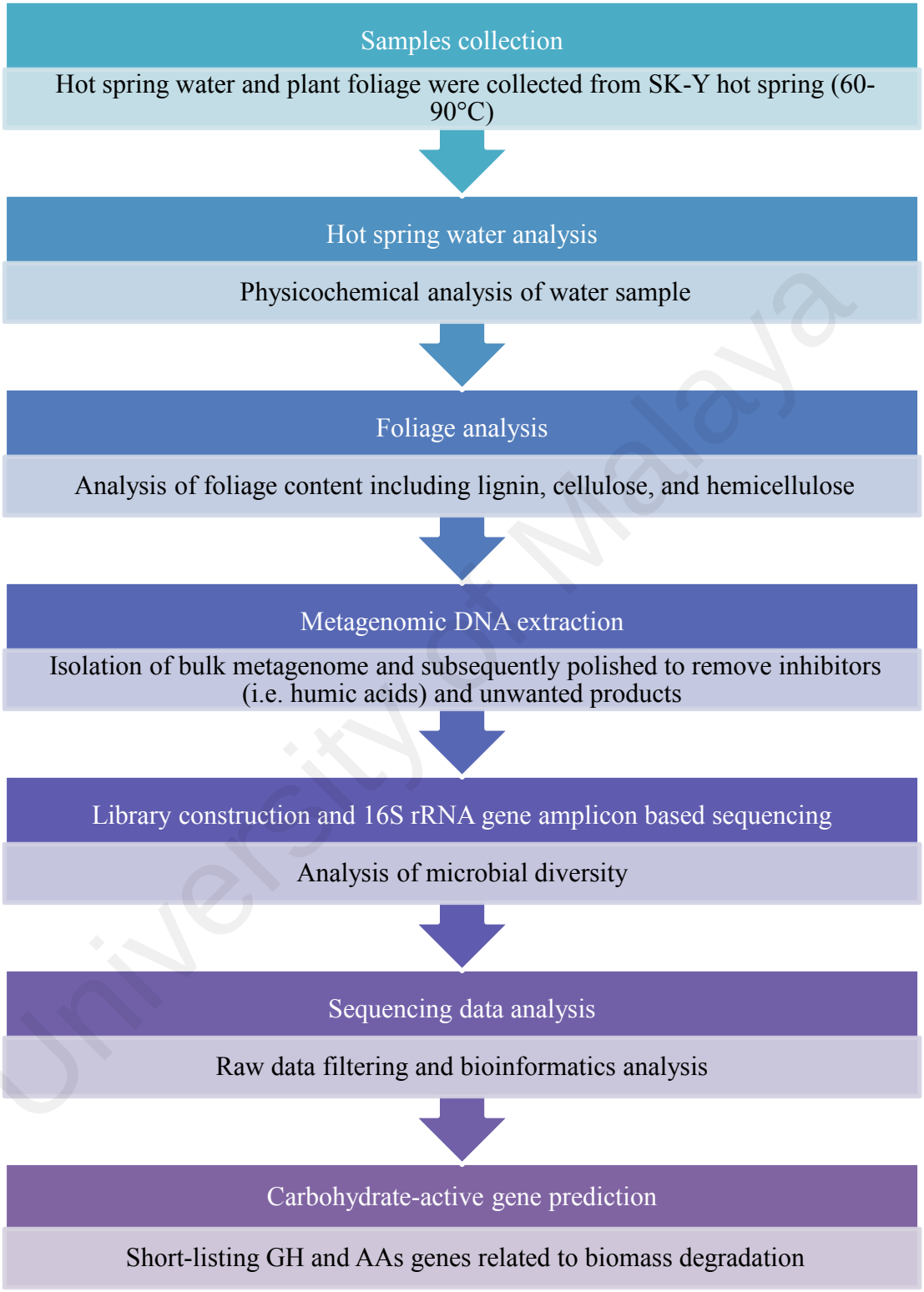
Component	Volume for a 20µL well reaction	Volume for a full 48 well plate (x50)
Kapa SYBR Fast qPCR Master Mix	10 µL	500 µL
Primer Premix	2 µL	100 µL
PCR grade water	4 µL	200 µL
Total	**16** µL	**800** µL

Table 3.9: Thermocycling conditions for Eco™ qPCR

Step	Temperature	Duration	
Initial activation/denaturation	95°C	5 min	
Denaturation	95°C	30 sec	35 cycles
Annealing / extension / data acquisition	60°C	45 sec[2]	

3.2 Methods

3.2.1 Flow chart



3.2.2 Sample collection and water analysis

SK-Y hot spring (3°59′50.50″N, 101°23′35.51″E) is located near to Sungkai, Perak, in Malaysia. Previously, Chan et al. (2015) have conducted microbial diversity analysis of the main water source in SK hot spring. In this study, as the shape of the sampling drainage is 'Y-shaped', this enclosed site is therefore named as SK-Y hot spring (hereinafter named as SK-Y hot spring), in order to differentiate current work to the earlier study on the SK hot spring (Chan et al., 2015). SK-Y hot spring is located approximately 10 metres away to the SK hot spring previously reported by Chan et al. (2015).

Sampling was performed on 24 March 2016. Temperature and pH of the SK-Y were measured using thermometer and pH paper, respectively. A clean stainless water sampling dipper was used to collect water samples without any foliage at four different spots with approximately 5 metres between sampling locations. Water was stored in sterile glass Schott bottles and immediately transported to laboratory within 2.5 hours and stored at 4°C for overnight. On the following day, water analysis was conducted by MyTest Lab Sdn Bhd (Malaysia) using American Public Health Association (APHA) standard protocols. At least 20 pieces of submerged foliage with no apparent biofilm were collected with sampling dipper and transferred to polypropylene Ziploc bags using a tweezer. Submerged foliage with green biofilm was collected and stored separately in 1 L Schott bottles. Degraded foliage was collected at the base of SK-Y with green biofilm and non-degraded plant litters were carefully removed in situ.

Freshly picked foliage samples from trees grown along SK-Y were collected and stored in sterile polyethylene bags. The content of lignin, cellulose, and hemicellulose of these foliage samples were analysed at Malaysian Agricultural Research and

Development Institute (MARDI), a local service provider for conducting analysis of acid detergent lignin (ADL), acid detergent fibre (ADF), and neutral detergent fibre (NDF), using a modified protocol based on Van Soest and Wine (1967). The plant identification was analysed by Dr. Yong Kien Thai from Rimba Ilmu, Institute of Biological Sciences, Faculty of Science, University of Malaya, Malaysia.

3.2.3 Metagenomic DNA extraction

In order to study microbial diversity of SK-Y hot spring, metagenome extraction of the following samples was performed: (i) pooled water of four sites at equal volume ratio (SK-Y hot spring water), (ii) submerged foliage with no apparent biofilm (labelled as non-decay), (iii) green biofilm found on submerged foliage (labelled as green biofilm), and (iv) degraded foliage (labelled as decay) collected at the base of SK-Y hot spring. The top layer of submerged plant litter bed is always covered with green biofilm (soft in texture) (Figure 3.1C). Apparently, foliage are decomposing underneath at the base of the bed (Figure 3.1E).

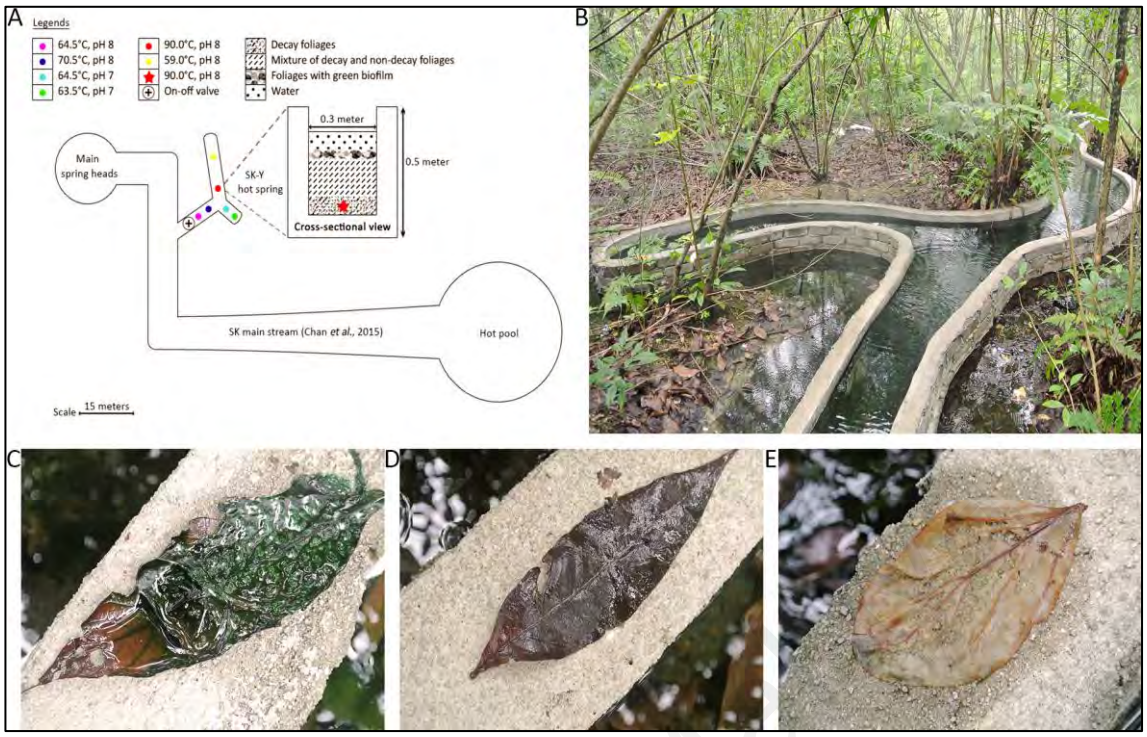


Figure 3.1: Y-shaped Sungai Klah (SK-Y) hot spring and types of samples. (A) Illustration of sampling site for water and foliage, (B) SK-Y hot spring, (C) Foliage with green biofilm, (D) Non-degraded foliage with no apparent biofilm, and (E) Degraded foliage. (Figure with permission by Lee et al. (2018) is licensed under CC BY 3.0)

Briefly, 4 litres of pooled water was filtered through a filter membrane with 0.22 µm pore size (Sartorius, Göettingen, Germany). Then, membrane was kept with 10 mL of autoclaved 1× concentration of phosphate-buffered saline (PBS, pH 7.4) (137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM phosphate buffer) (Merck, Darmstadt, Germany) containing sterile glass beads and shaken vigorously for 5 minutes. Next, the membrane was withdrawn and the leftover liquid was centrifuged at 17,000×g for 2 minutes at 4°C. The supernatant was discarded and the pellet was resuspended using Tris-EDTA buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) (Epicentre, Madison, WI, USA) and modified cetyltrimethylammonium bromide (CTAB) lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 100 mM K_2HPO_4, 1.5 M NaCl, and 1% w/v CTAB) (Murray & Thompson, 1980; Zhou et al., 1996). Subsequently, enzymatic lysis was carried out by incubating at 37°C for overnight in the presence of 10 mg mL^{-1} of lysozyme (Sigma-Aldrich, Saint Louis, MO, USA) and gently swirled at 20 minutes interval. The solution

was then boiled at 90°C for 1 hour (Murray & Thompson, 1980). Subsequently, SDS was added to the final concentration of 1% (w/v), followed by the supplementation of 20 mg mL-1 of proteinase K (Qiagen, Valencia, CA, USA), incubated at 60°C for 2 hours with gentle shaking at 15 minutes interval. RNA treatment was performed with addition of 100 mg mL^{-1} RNase A (Qiagen, Valencia, CA, USA) and incubated at 37°C for 30 minutes. Proteins were removed by washing the DNA pellet with phenol/chloroform/isoamyl alcohol (25:24:1). The resulting DNA pellet was precipitated with 0.6 volume of isopropanol followed by 70% (v/v) ethanol and was rehydrated with 60 µl of elution buffer (Qiagen, Valencia, CA, USA) (Manjula et al., 2011).

Briefly, 100 g of foliage samples (separately for foliage with no apparent biofilm, foliage with green biofilm, or degraded foliage) were placed inside a 500 mL autoclaved glass bottle containing sterile PBS buffer, pH 7.4 (137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM phosphate buffer) (Merck, Darmstadt, Germany) with 0.05% (v/v) Tween 20 (4 mL PBS per 1 g leaf, pH 7.4), and sonicated (Branson Ultrasonics, Danbury, CT, USA) in water bath for 1 minute at 25°C. The preparations were then hand-shaken vigorously for 30 seconds, and the leaves debris were discarded while remaining liquid was aliquoted into a 50 mL tube and centrifuged at 14,800×g for 10 minutes at 4°C. The pellet was subjected to the aforementioned conventional metagenome extraction.

To improve the purity of extracted metagenomic DNA libraries, inhibitors such as humic acids were removed using the Agencourt AMPure XP System (Beckman Coulter, Brea, CA, USA). The quality and yield of purified metagenome were examined using 1% w/v agarose gel electrophoresis, a Nanodrop™ 1000 spectrophotometer

(Thermo Scientific, Waltham, MA, USA), and a Qubit® 2.0 Fluorometer (Invitrogen, Merelbeke, Belgium).

3.2.4 Library construction and 16s rRNA gene amplicon based sequencing

The Illumina 16S rRNA Metagenomic Sequencing Library Preparation Guide was followed for preparing metagenome libraries. Two sets of primers were used to target on bacteria and archaea hypervariable 16S rRNA conserved regions (Klindworth et al., 2012): (a) 16S rRNA V3 and V4 bacterial amplicon PCR forward primer 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG-3'; (b) 16S rRNA V3 and V4 bacterial amplicon PCR reverse primer 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C-3'; (c) targeted archaeal amplicon PCR forward primer 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CAG CMG CCG CGG TAA-3'; and (d) targeted archaeal amplicon PCR reverse primer 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTA CNV GGG TAT CTA ATC C-3'. The underlined and non-underlined sequences refer to Illumina adapter overhang nucleotide sequences and locus-specific primers for the regions to be targeted, respectively. The amplicons were then subjected to a series of library quantification steps to accurately quantify NGS sample libraries. Accurate quantification is extremely critical to optimise the output of Illumina's MiSeq system. The amplicons were quantified with Qubit® dsDNA HS Assay Kit (Life Technologies, USA) on a Qubit® 2.0 Fluorometer, library size selection based on Agilent Technologies 2100 Expert Bioanalyzer using Agilent High Sensitivity kit (Agilent Technologies, Santa Clara, CA, USA), and absolute quantification of the number of amplifiable molecules in a library using Eco™ Real-Time PCR System with KAPA SYBR Fast qPCR Master Mix (KAPA BioSystems, Boston, MA, USA) prior to sequencing. Next, metagenome libraries were sequenced using paired-end sequencing on

an Illumina MiSeq sequencer (Illumina, San Diego, CA, USA) with MiSeq Reagent Kit V2 (2 × 250 base pairs) and V3 (2 × 300 base pairs) for archaea primers amplicons and bacteria primers amplicons, respectively.

3.2.5 Sequence analysis

The raw sequence reads generated by Illumina sequencing were processed in CLC Genomic Workbench 7.0 (CLC Bio, Aarhus, Denmark). Adapter sequences were trimmed and reads were filtered to ensure an average PHRED score of 20. Paired-end reads were merged (mismatch cost = 2; gap cost = 3; maximum unaligned end mismatches = 0; minimum score = 8) in CLC Genomics Workbench 7.0. The assembled reads were then undergone chimera filtering and microbial taxonomic classification using QIIME pipeline (Caporaso et al., 2010). The QIIME pipeline was followed with default parameters using version 1.9.1 unless otherwise noted. Briefly, the steps included are removing chimera sequences, picking OTUs based on open reference clustering approach using UCLUST tool, and taxonomic assignment using BLAST against the National Center for Biotechnology Information (NCBI) 16S Microbial database with e-value of 0.001. The NCBI database was selected due to its diverse and huge database with the capability to provide greater depth of information during taxonomic profiling compared to RDP, GreenGenes or SILVA databases (Chan et al., 2015). All samples were randomly subsampled to the same sequencing depth prior to further analysis. Microbial diversity was assessed with rarefraction analysis, the number of observed OTUs per sample, Shannon-Wiener, and Simpson using QIIME. Beta diversity measurement between all the samples were calculated using Unifrac distance (Lozupone & Knight, 2005), implemented in QIIME. Principal coordinates analysis (PCoA) was performed on the weighted UniFrac distance matrix which accounts for communities' membership and relative abundances of OTUs.

3.2.6 Metabolic potential analysis

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Langille et al., 2013) was used to predict the functional profiles of the microorganisms present in SK-Y hot spring. An OTU table was generated using the closed-reference OTU picking method in QIIME Version 1.9.1 (Caporaso et al., 2010), and was taxonomically assigned using UCLUST with GreenGenes Version 13_05 as the reference database. Next, the OTU table was analysed using the PICRUSt pipeline according to the metagenome inference workflow, as described by the developers, to generate putative KEGG Ortholog abundances for each sample within the OTU table. Data from the PICRUSt pipeline were statistically evaluated with the STAMP bioinformatics package (Parks et al., 2014).

3.2.7 Carbohydrate-active gene prediction

After taxonomic assignment using QIIME, the taxa with relative abundance of ≥0.85% were individually checked against the complete genomes information available in CAZy to determine the numbers and types of glycoside hydrolase families for these taxa. Later, the taxa consisted of high GH families were subjected to BacDive (The Bacterial Diversity Metadatabase) search to determine metadata such as taxonomy, morphology, physiology, culture and growth conditions, and isolation, sampling and environmental information in each taxon.

CHAPTER 4: RESULTS

4.1 General sampling site descriptions

More than a dozen of hot spring features are present in SK hot spring park and the main hot spring at this area is a small stream with multiple spring heads. Chan et al. (2015) have performed earlier the 16S rRNA gene amplicon and shotgun sequencing for samples taken from the SK main stream (SK hot spring) (Chan et al., 2015). Despite the SK hot spring is located within a wooded area, plant litters are not accumulated in the stream as the water flows rapidly. Approximately 10 metres away to the SK hot spring, a man-made drainage with a total length of approximately 30 metres and 0.5 metre depth was built to trap heated spring water (Figure 3.1A). As the shape of the drainage is 'Y-shaped', this site is therefore named as SK-Y hot spring, in order to differentiate current work to our earlier study on the SK hot spring (Chan et al., 2015). The temperature at the SK-Y spring head was approximately 90°C, but was lower 60-70°C (mean 68°C) adjacent to water surface or further away from the spring head. The pH for SK-Y hot spring was between 7.5-8.6. The most interesting feature for this hot spring is not only high in temperature, but SK-Y hot spring is accumulated with fallen plant litters (mainly foliage) that formed a bed of foliage (thickness: approximately 20 cm). The average size for fresh fallen foliage is about 15 cm (length) × 6 cm (width) (Figure 4.1).

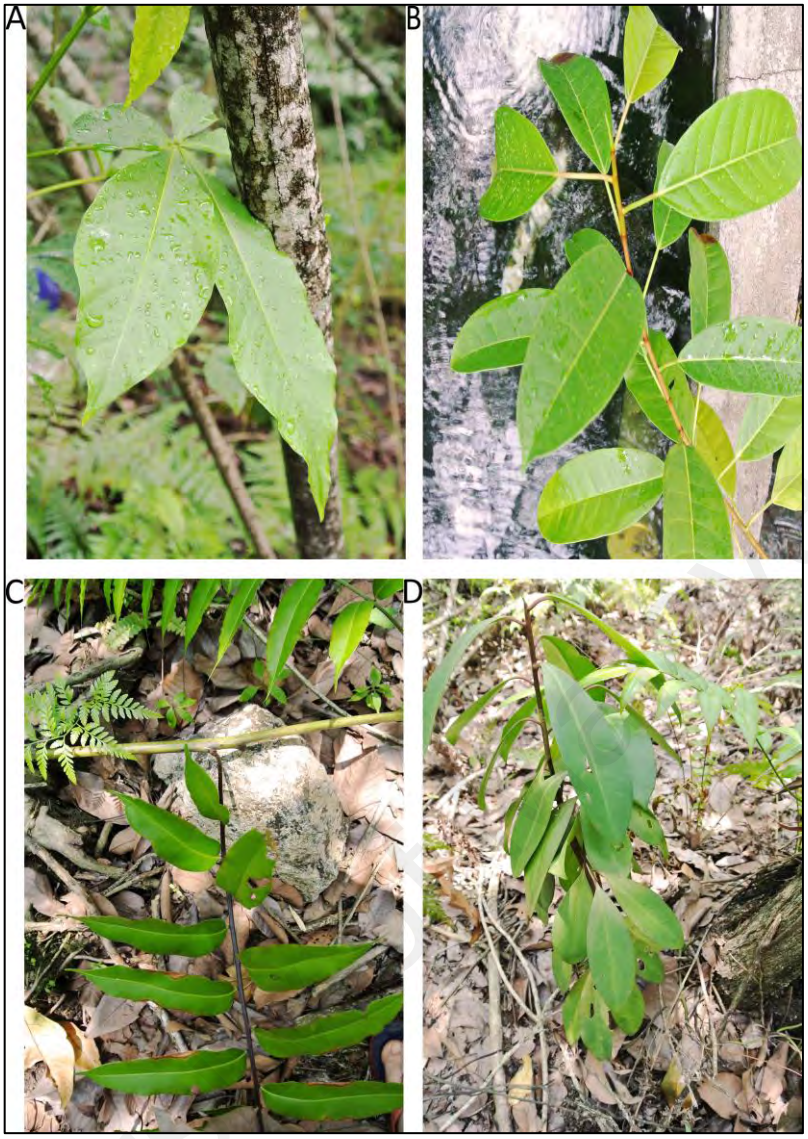


Figure 4.1: Picked foliage found in SK-Y hot spring. (A) *Vitex* sp., (B) *Ficus* sp., (C) *Stenochlaena* sp., and (D) *Adenanthera* sp.. (Figure with permission by Lee et al. (2018) is licensed under CC BY 3.0)

4.2 Physicochemical analysis of water

Water analysis was carried out to study the physiochemical condition of SK-Y hot spring (Appendix A). The temperature and pH during sampling was 68°C and pH 8.6, respectively. The colour of the water was 68 TCU (true colour unit), slightly lower than the SK stream (75 TCU) (Chan et al., 2015). 0.96 mg L^{-1} and 0.65 mg L^{-1} of aluminium and iron, respectively were detected in SK hot spring but these metal ions were not detected in SK-Y hot spring water sample. SK-Y hot spring has higher fluoride (6 mg L^{-1}), nitrate (0.29 mg L^{-1}), and zinc (0.17 mg L^{-1}) content to that of SK hot spring (1.1 mg

L^{-1}, <0.1 mg L^{-1}, <0.02 mg L^{-1}, respectively). The sulphur and sulphide content in SK-Y hot spring were 0.5 mg L^{-1} and 12.3 mg L^{-1}, respectively. Other metals such as mercury, cadmium, chromium, lead, manganese, nickel, silver, aluminium, barium, and strontium were below the quantifiable limits.

Table 4.1: Water analysis of the pooled SK-Y hot spring water sample (Lee et al., 2018)

Test parameter	Method	Unit	SK-Y	SK	WHO
Physical					
Colour ADMI	APHA 2120 F	TCU	68	75	15
Turbidity [d]	APHA 2130 B	NTU	ND (<1)	130	5
Chemical					
pH	APHA 4500 - H^+ B	-	8.6	8.2	6.5−8.5
Aluminium (Al)	APHA 3111 D	mg L^{-1}	ND (<0.1)	0.96	0.2
Ammonia (N)	APHA 4500 – NH_3 C / NH_3 B	mg L^{-1}	ND (<0.1)	ND (<0.1)	1.5
Ammonical nitrogen	APHA 4500 – NH_3 C / NH_3 B	mg L^{-1}	ND (<0.1)	<0.2	-
Anionic Detergent (MBAS) [d]	APHA 5540 C (Mod.)	mg L^{-1}	ND (<0.5)	ND (<0.2)	-
Arsenic (As)	APHA 3114 B	mg L^{-1}	ND (<0.001)	0.07	0.01
Barium (Ba)	APHA 3111 D	mg L^{-1}	ND (<0.05)	ND (<0.02)	0.7
Borate (H_3BO_3) [d]	APHA 4500 B-C	mg L^{-1}	ND (<1)	-	-
Boron (B)	APHA 4500 - B B	mg L^{-1}	ND (<0.10)	0.06	2.4
Bicarbonate ($CaCO_3$) [d]	APHA 2320 B	mg L^{-1}	97.7	-	-
Biocides (Total) [d]	In House	mg L^{-1}	ND (<0.1)	ND	-
Bromide (Br^-) [d]	By Ion Chromatography	mg L^{-1}	ND (<1.0)	-	-
Cadmium (Cd)	APHA 3111 B	mg L^{-1}	ND (<0.003)	ND (<0.002)	0.003
Carbon Chloroform Extract (CCE) [d]	In House	mg L^{-1}	ND (<0.5)	ND	0.3
Calcium (Ca) [d]	APHA 3111 B	mg L^{-1}	0.03	-	-
Chloride (Cl^-)	APHA 4500 - CI^- B	mg L^{-1}	0.5	2	300
Chloroform [d]	APHA 6210 B	mg L^{-1}	ND (<0.3)	ND	0.3
Chromium (Cr III)	In-house Method III [b]	mg L^{-1}	ND (<0.01)	-	0.05
Chromium (Cr VI)	APHA 3500 - Cr B	mg L^{-1}	ND (<0.01)	ND (<0.02)	0.05
Copper (Cu)	APHA 3111 B	mg L^{-1}	0.08	ND (<0.02)	2
Cyanide (CN^-)	In-house Method I [a]	mg L^{-1}	ND (<0.01)	ND (<0.02)	-
Fluoride (F^-)	APHA 4500-F^- D	mg L^{-1}	6	1.1	1.5
Formaldehyde (CH_2O)	In-House Method V [c]	mg L^{-1}	ND (<0.01)	<0.1	-

Table 4.1, continued.

Test parameter	Method	Unit	SK-Y	SK	WHO
Free chloride residual (Cl_2)	APHA 4500- Cl B	mg L^{-1}	0.3	<0.1	5
Hardness ($CaCO_3$) [d]	APHA 2340 B	mg L^{-1}	0.6	<1	500
Iron (Fe)	APHA 3111 B	mg L^{-1}	ND (<0.05)	0.65	0.3
Lead (Pb)	APHA 3111 B	mg L^{-1}	ND (<0.01)	ND (<0.02)	0.01
Magnesium (Mg) [d]	APHA 3111 B	mg L^{-1}	0.12	0.5	-
Manganese (Mn)	APHA 3111 B	mg L^{-1}	ND (<0.01)	ND (<0.02)	0.1
Mercury (Hg)	APHA 3112 B	mg L^{-1}	ND (<0.001)	ND (<0.001)	0.006
Mineral oil [d]	APHA 5520 F Hydrocarbons	mg L^{-1}	ND (<0.3)	ND (<0.2)	-
Nickel (Ni)	APHA 3111 B	mg L^{-1}	ND (<0.02)	ND (<0.02)	0.07
Nitrate nitrogen [d]	APHA 4500 - NO^{3-} B	mg L^{-1}	0.29	<0.1	-
Nitrite (NO_2^-)	Palin Test Kits, Method: AK109	mg L^{-1}	ND (<0.1)	<0.1	3
Phenol (C_6H_5OH)	APHA 5530 B & APHA 5530 D	mg L^{-1}	ND (<0.002)	ND (<0.002)	-
Phosphate (PO_4^{3-})	APHA 4500 - P C	mg L^{-1}	0.84	0.2	-
Potassium (K) [d]	APHA 3111 B	mg L^{-1}	2.5	-	-
Selenium (Se)	APHA 3114 B	mg L^{-1}	ND (<0.001)	ND (<0.005)	0.04
Silver (Ag)	APHA 3111 B	mg L^{-1}	ND (<0.05)	ND (<0.02)	-
Sodium (Na) [d]	APHA 3111 B	mg L^{-1}	34.5	27	50
Strontium (Sr) [d]	APHA 3120 B	mg L^{-1}	ND (<0.6)	-	-
Sulfate (SO_4^{2-})	APHA 4500 - SO_4^{2-} E	mg L^{-1}	6.4	8	250
Sulfur (S) [d]	Test by ICP-OES	mg L^{-1}	0.5	3.9	-
Sulphide	APHA 4500 - S^{2-} F	mg L^{-1}	12.3	0.2	-
Total nitrogen	APHA 4500 - N_{org} B	mg L^{-1}	ND (<0.1)	5.6	-
Zinc (Zn)	APHA 3111 B	mg L^{-1}	0.17	ND (<0.02)	5
Other					
Acidity	APHA 2310 B	mg L^{-1}	15.8	<1	-
Alkalinity	APHA 2320 B	mg L^{-1}	51.7	76	-
Biochemical oxygen demand (BOD) 5 days at 20°C	APHA 5210 B / APHA 4500 - O G	mg L^{-1}	6	5	-
Biochemical oxygen demand (BOD) 5 days at 60°C	APHA 5210 B / APHA 4500 - O G	mg L^{-1}	4	10	-
Biochemical oxygen demand (BOD) 5 days at 80°C	APHA 5210 B / APHA 4500 - O G	mg L^{-1}	3	5	-
Chemical oxygen demand (COD)	APHA 5220 C	mg L^{-1}	17	35	-

Table 4.1, continued.

Test parameter	Method	Unit	SK-Y	SK	WHO
C:N Ratio (TOC/TN) [d]	APHA 5310 D / APHA 4500 - N_{org} B	mg L^{-1}	0.5	-	-
Dissolved oxygen	APHA 4500 - O G	mg L^{-1}	3.0	6.2	
Total organic carbon (TOC) [d]	APHA 5310 D	mg L^{-1}	0.502	9.04	-
Bacteriological					
Total coliform count [d]	APHA 9221 B	MPN (per 100 mL)	ND (<2)	ND (<1.1)	ND
Escherichia coli [d]	APHA 9221 F	MPN (per 100 mL)	ND	ND (<1.1)	ND

Abbreviations:
SK-Y: Y-shaped Sungai Klah hot spring; SK: Sungai Klah main stream hot spring; WHO: World Health Organization; APHA: Standard Methods of American Public Health Association for the Examination of Water & Wastewater, 21st Edition, 2005; ADMI: American Dye Manufacturer's Institute; TCU: true color unit; NTU: Nephelometric Turbidity Unit; MPN: most probable number; Mod.: modified method; ND: not detectable

Method reference(s):

[a] In-house Method 1: Based on APHA 4500 - CN- C & E and Merck Method 14429
[b] In-house Method III: Based on APHA 3500 - Cr B /APHA 3111 B
[c] In-house Method V: Based on Macherey Nagel Nanocolor Formaldehyde 8
[d] Not SAMM Accredited

4.3 Analysis of foliage lignocellulose content

Vitex, *Ficus*, *Stenochlaena*, and *Adenanthera* are the main plant genera grown adjacent to SK-Y hot spring. Most of these plants are young trees with height of approximately 2-4 metres. The average percentage of lignin, cellulose, and hemicellulose for fresh foliage randomly picked from the trees were analysed using neutral detergent fibre (NDF; hemicellulose, cellulose, and lignin), acid detergent fibre (ADF; cellulose and lignin), and acid detergent lignin (ADL; fractions of plant cell walls) (Van Soest & Wine, 1967) (Table 4.2). The percent of lignin varied from 3.0%-16.7%, while hemicellulose contents were 2.9%-4.5% among the foliage samples. *Vitex* and *Stenochlaena* foliage samples were relatively higher in cellulose content than that of *Ficus* and *Adenanthera* samples. Lignocellulosic contents of these plant genera were not found in literatures, except for the one reported by Codron et al. (2007) for *Vitex* sp. foliage, which stated that foliage contained 6.0% lignin, 8.1% hemicellulose, and 7.2% cellulose.

Table 4.2: Approximate composition (as a percentage) of various foliage samples (Lee et al., 2018)

Genus	Lignin (%)	Hemicellulose (%)	Cellulose (%)
Vitex	9.9	3.9	10.2
Ficus	8.7	2.9	7.0
Stenochlaena	16.7	4.3	11.3
Adenanthera	3.0	4.5	3.8

4.4 16S rRNA NGS data analysis

Metagenomic DNA extracts that represent microbiome for water, non-decay, green biofilm, and decay were undergone 16S rRNA gene amplicon sequencing using primer pairs specific for bacteria and archaea. After quality filtration and adapters trimming of raw reads, high quality assembled reads were analysed using Quantitative Insights Into Microbial Ecology (QIIME) pipeline (Table 4.3). Average of 14929 and 25 observed OTUs for bacteria and archaea, respectively were generated from four samples. These data were processed at a rarefaction depth of 819322 and 12175 sequences per bacteria and archaea samples, respectively. Simpson and Shannon-Wiener indexes indicated the highest species richness and evenness in the bacterial diversity of SK-Y hot spring water sample, while the lowest was found in the archaeal diversity of non-decay sample. The rarefaction curves of all the samples did not reach saturation, indicating that there is high level of diversity in the systems (Figure 4.2).

Table 4.3: Summary of assembled data obtained from metagenomes of foliage and water microbiota (Lee et al., 2018)

Dataset		SK-Y Hot Spring Water		Green Biofilm		Non-decay		Decay	
		Bacteria	Archaea	Bacteria	Archaea	Bacteria	Archaea	Bacteria	Archaea
Number of reads		510983	1153627	248958	1455515	288476	1437865	651444	826285
Sequence length (bp)	Minimum	200	200	264	203	264	200	200	200
	Average	294	266	269	263	269	263	329	261
	Maximum	429	422	275	390	275	390	430	320
Observed OTUs		11704	31	16331	26	16153	21	15529	21
Shannon		7.867	2.195	7.450	0.961	7.318	0.574	7.169	1.402
Simpson		0.989	0.533	0.982	0.242	0.982	0.141	0.981	0.405

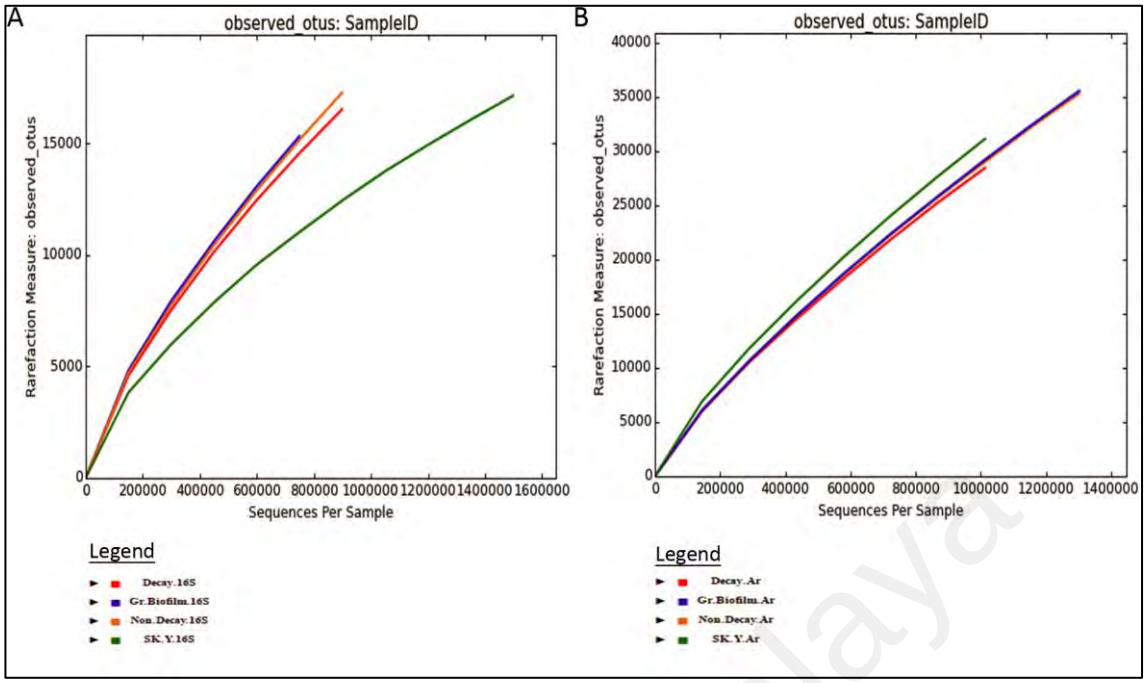


Figure 4.2: Alpha rarefraction plot based on observed OTUs. (A) Bacterial diversity and (B) Archaeal diversity. (Figure with permission by Lee et al. (2018) is licensed under CC BY 3.0)

To assess the microbial phylogenetic beta diversity, weighted Unifrac distance was used to indicate the extent of the phylogenetic similarities among the microbial communities (Lozupone et al., 2007). PCoA using weighted UniFrac revealed that the bacterial communities (Figure 4.3A) of the samples are grouped into three distinct clusters. Overall, bacterial communities were phylogenetically more similar between the green biofilm and non-decay samples, while revealed a clear difference between the SK-Y hot spring water, decay, and cluster of green biofilm and non-decay samples. Besides, the same pattern was also noticed in archaeal communities (Figure 4.3B).

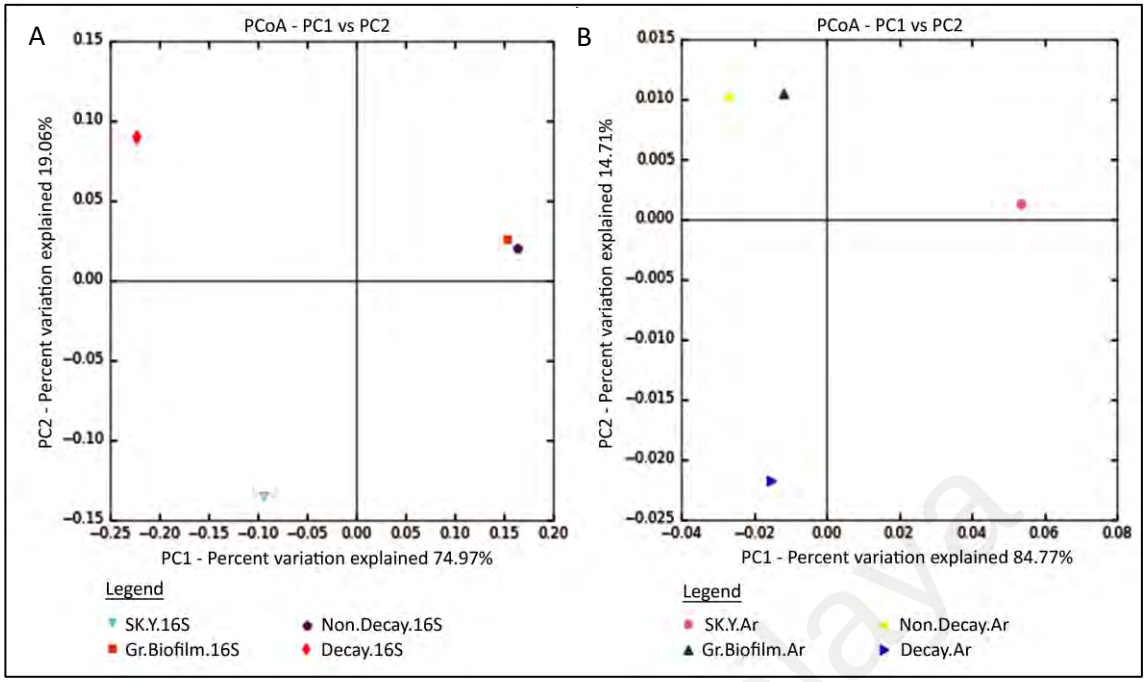


Figure 4.3: Ordination plots derived from Principal Coordinates Analysis (PCoA) of phylogenetic beta diversity metrics using weighted UniFrac algorithm between bacterial and archaeal community composition of SK-Y hot spring. (A) PCoA of bacterial diversity and (B) PCoA of archaeal diversity. (Figure with permission by Lee et al. (2018) is licensed under CC BY 3.0)

4.5 Bacterial diversity analysis

Unless specified, the values shown in this subsection is the average number for four samples. Generally, taxonomic assignment of bacteria in the four samples could be classified into 25 phyla, 58 classes, 110 orders, 171 families, and 328 genera. Ten most abundant phyla contributed up to 75.5% of the total bacterial diversity included Proteobacteria (14.7%), followed by Chloroflexi (12.8%), Firmicutes (10.8%), and Cyanobacteria (8.4%) (Figure 4.4A). Minor phyla present in SK-Y hot spring samples included Chlamydiae, Gemmatimonadetes, Elusimicrobia, Lentisphaerae, and Deferribacteres which each of them was ≤0.08% of total populations. OTUs affiliated to Armatimonadetes, Chlorobi, Spirochaetes, and unclassified candidate phyla were also detected at lower percentage. On the other hand, decay sample consisted of higher percentage of Acidobacteria (9.3%), Aquificae (4.6%), and Thermotogae (5.2%) than the other three samples. Photosynthetic phyla, for instances, Chloroflexi and Cyanobacteria

are present in lower percentage in decay sample than those present in green biofilm. Broad range of Proteobacteria class including Alpha-, Beta-, Gamma-, Delta-, and Epsilonproteobacteria were observed in all samples.

The following classes contributed to almost 69% of total population in degraded foliage microbiota: Actinobacteria, Aquificae, Deinococci, Bacilli, Clostridia, Ignavibacteria, Planctomycetia, Alphaproteobacteria, Thermodesulfobacteria, Thermotogae, unclassified Acidobacteria, and other unclassified classes or blast hits. All these classes are also present in green biofilm, non-decay, and water samples, but may appear as different percentages. Major classes associated with green biofilm but present in lower percentage in decay sample are those of Blastocatellia, Bacteroidia, Cytophagia, Caldilineae, Chloroflexia, Cyanobacteria, Betaproteobacteria, Deltaproteobacteria, and Gammaproteobacteria.

A total of 171 families were observed across all the samples with Thermaceae (4.5%), Caldilineaceae (3.8%), Chloroflexaceae (3.6%), and Roseiflexaceae (3.6%) represent the most abundant groups. Apart from that, Fervidobacteriaceae (5.0%) was found to be the predominant family in decay sample. For bacterial classified genera, the four samples in SK-Y hot spring shared most of the OTUs with different abundancy (Figure 4.4B) which were dominated by *Caldilinea* (3.8%), *Meiothermus* (3.8%), *Chloroflexus* (3.6%), *Roseiflexus* (3.6%), *Thermoanaerobaculum* (3.1%), *Melioribacter* (2.4%), *Geobacillus* (2.2%), *Desulfurobacterium* (1.8%), *Thermosipho* (1.3%), *Thermoanaerobacterium* (1.0%), *Fervidobacterium* (1.0%), *Acidimicrobium* (1.0%), *Chloracidobacterium* (1.0%), and *Thermoanaerobacter* (0.9%).

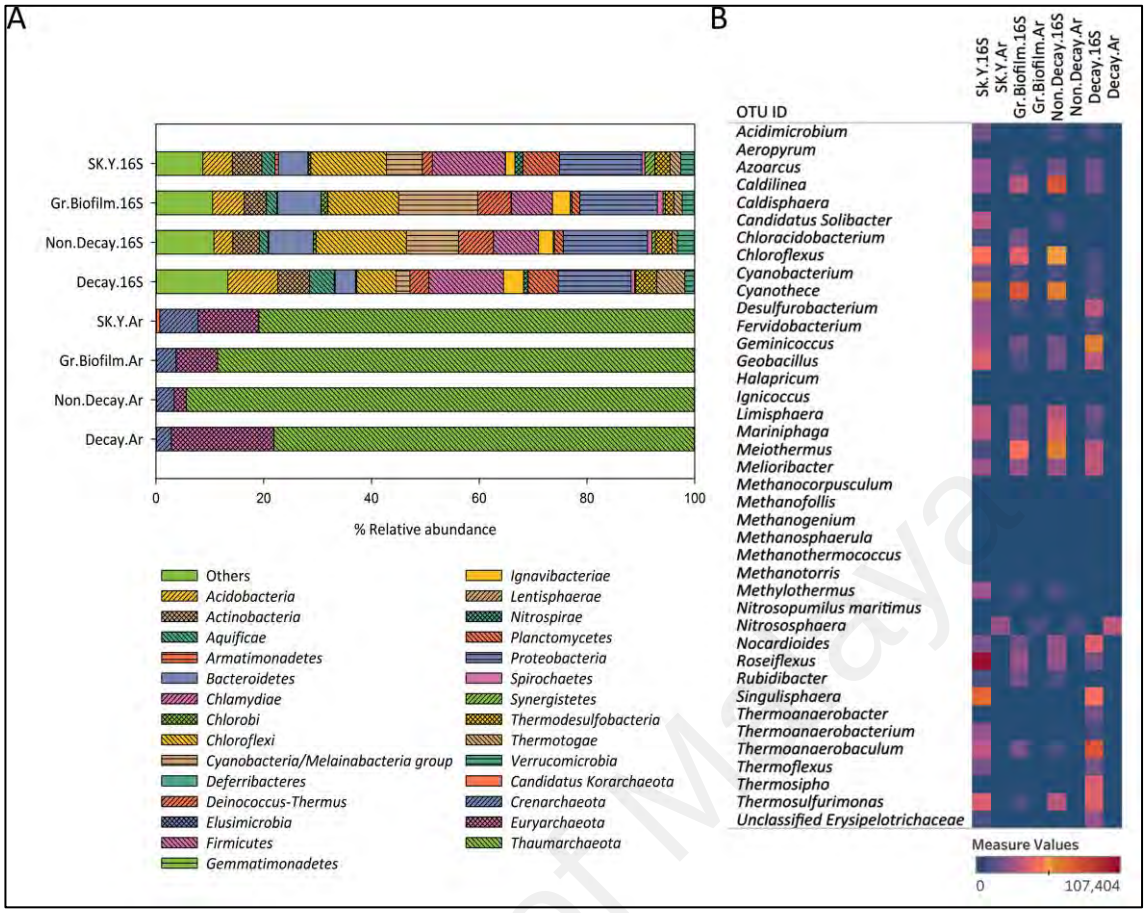


Figure 4.4: (A) Relative abundance at bacterial and archaeal phyla in SK-Y hot spring and (B) Heat map representation of abundance bacterial (≥0.85%) and archaeal genera in SK-Y hot spring. (Figure with permission by Lee et al. (2018) is licensed under CC BY 3.0)

4.6 Archaeal diversity analysis

In this work, an archaea specific primer pair was used to target locus-specific sequences of archaea in the four samples taken from SK-Y hot spring. Despite the primers were designed specifically for archaea, only 0.9%-3.8% of the OTUs identified in these four samples were affiliated to archaea. From this, archaea was deduced to be present in a relatively smaller portion in SK-Y hot spring, and further analysis using shotgun metagenome sequencing is needed to validate this assumption. For subsequent analysis for archaea, all OTUs related to bacteria were excluded. The most prevalent phylum of archaeal diversity across all samples was Thaumarchaeota (85.5%). In SK-Y hot spring water, green biofilm, and decay samples, Euryarchaeota was the second most abundant archaeal phylum, followed by Crenarchaeota. Nevertheless, Crenarchaeota was the

second dominant phylum in the non-decay sample, followed by Euryarchaeota. Within Thaumarchaeota, the most abundant class in all samples was Nitrososphaeria with *Nitrososphaera* (84.0%) as dominant taxa. Furthermore, the second most abundant class in SK-Y hot spring water, green biofilm, and decay samples was Methanomicrobia that was mainly represented by *Methanofollis*. Thermoprotei was the second dominant class in non-decay sample with *Aeropyrum* as the second abundant genus. Genus level assignment revealed that *Methanocorpusculum* (3.6%), *Nitrosopumilus* (1.5%), *Methanogenium* (0.4%), *Ignicoccus* (0.3%), *Methanotorris* (0.3%), *Methanosphaerula* (0.3%), *Halapricum* (0.2%), and *Methanothermococcus* (0.2%) genera were present across all the samples (Figure 4.4B). In contrast, *Caldisphaera* genus was found in all the samples except for non-decay sample.

4.7 PICRUSt analysis

In this study, PICRUSt classified the metagenome into six main functional modules: (1) cellular processes and signaling; (2) environmental information processing; (3) genetic information processing; (4) human diseases; (5) metabolism; and (6) organismal systems (Figure 4.5). The microbial metabolism of carbohydrates and the total numbers of carbohydrate-active genes existing in this SK-Y hot spring community were estimated. A great proportion of genes are related to carbohydrate metabolism (Figure 4.6), suggesting the presence of a high diversity of biomass degrading communities in the degrading foliage in this hot spring. Methane and sulfur metabolism gene modules were also identified by the PICRUSt pipeline, which indicate the regulation of methane and sulfur (Figure 4.7) in the geochemical cycle.

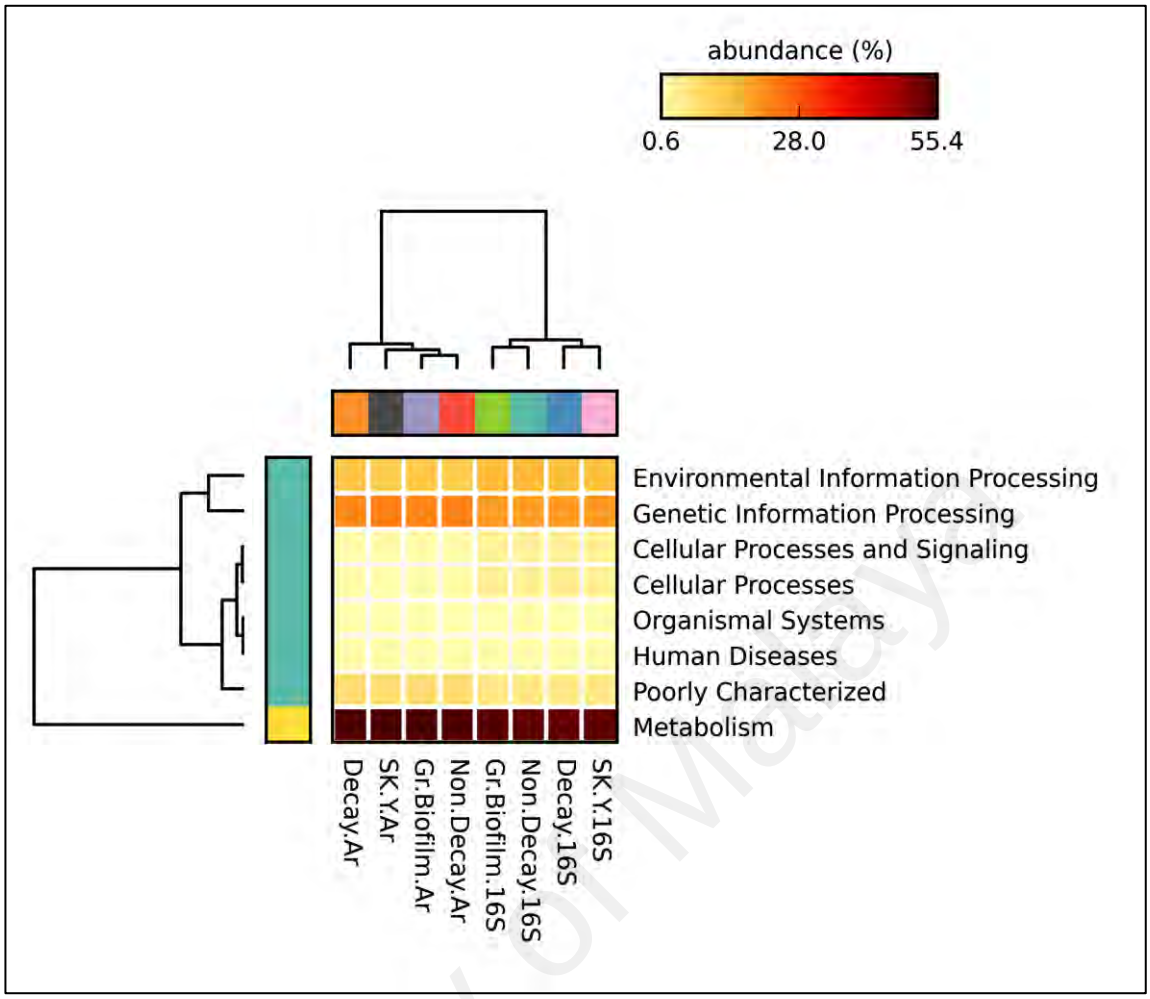


Figure 4.5: Hierarchical clustering heat map generated from functional abundance profile reflecting the distribution of metagenome into six major modules. (Source: Personal collection)

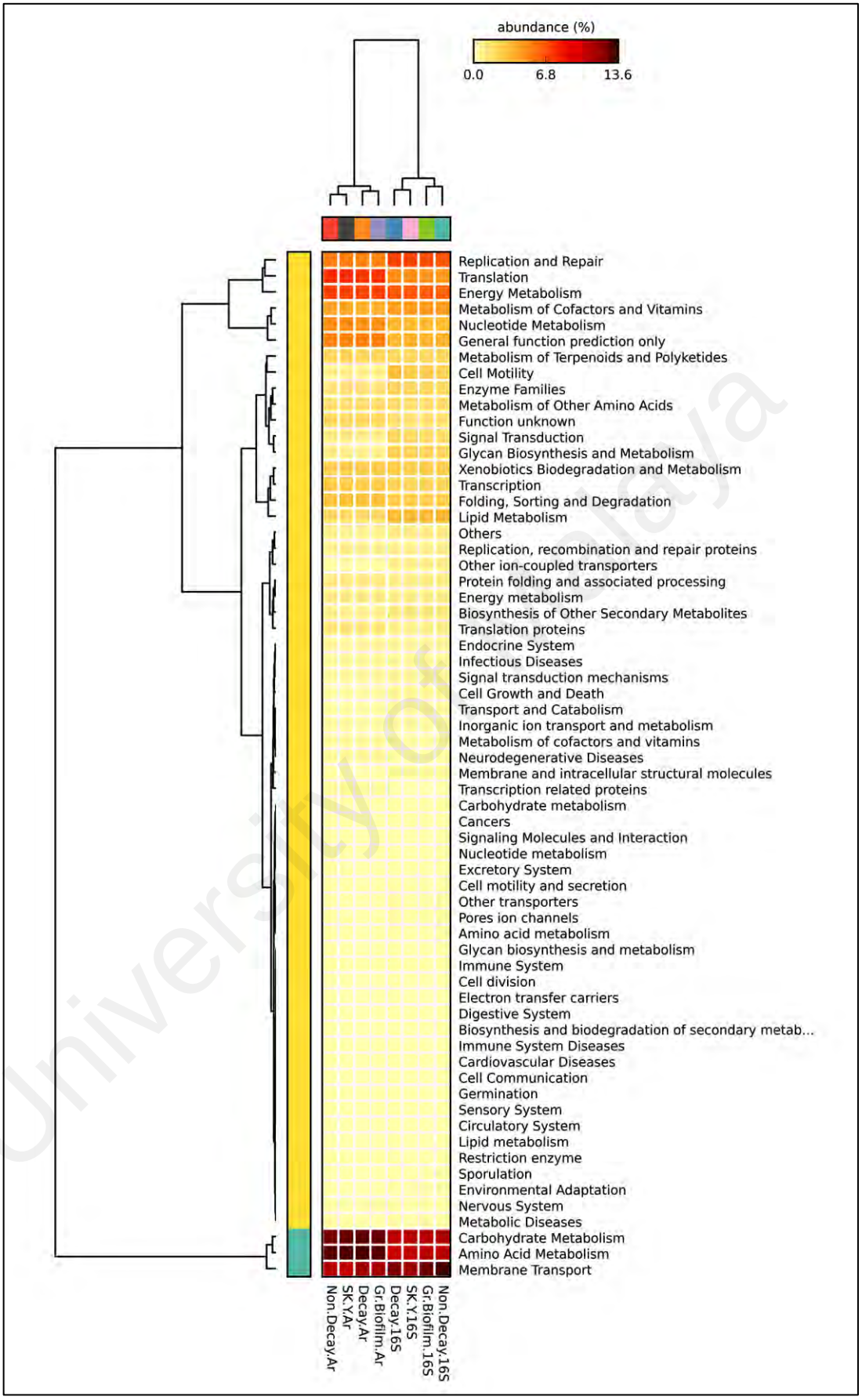


Figure 4.6: Hierarchical analysis of the community functions from bacterial and archaeal data sets classified by KEGG Orthology. (Source: Personal collection)

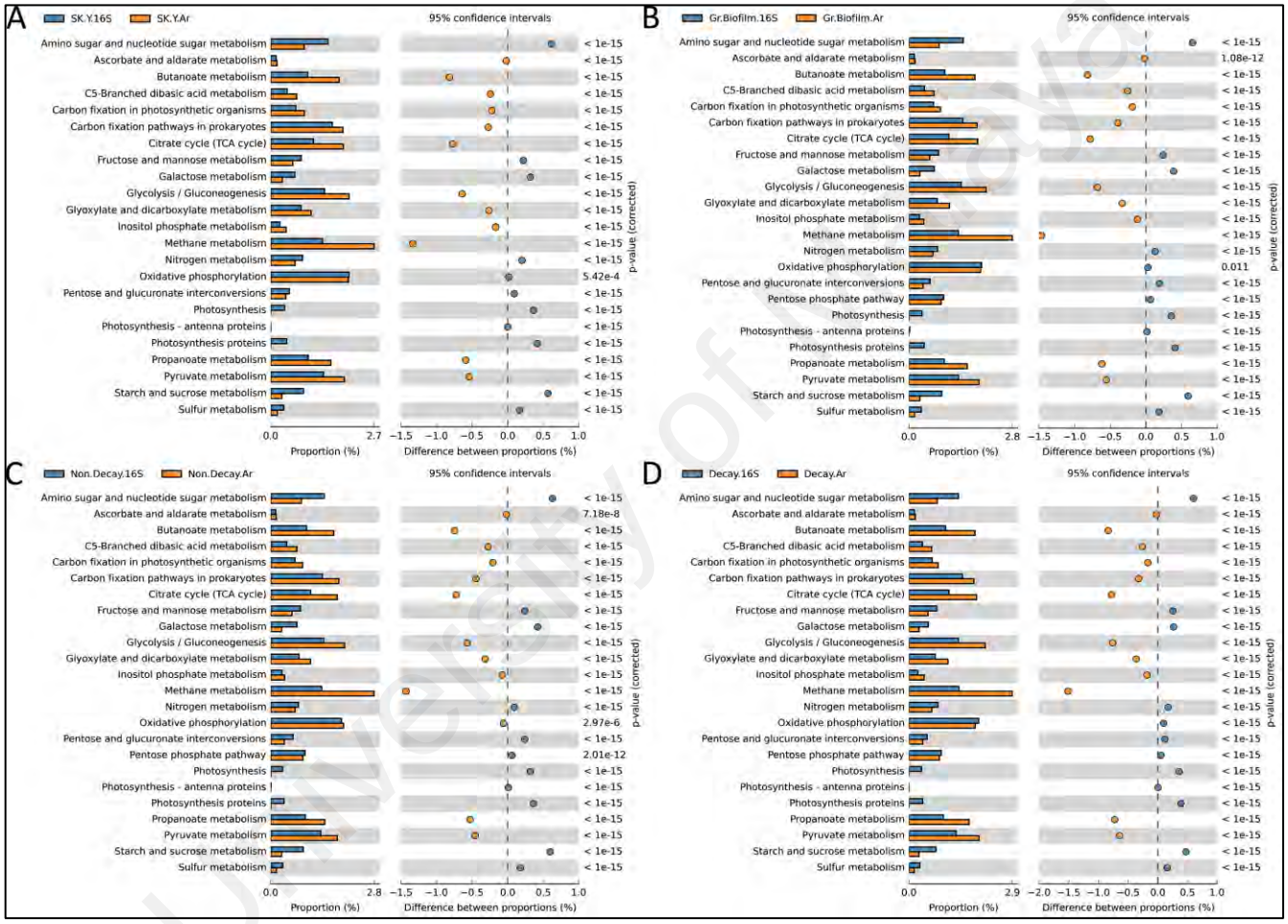


Figure 4.7: Relative abundance of genes in SK-Y hot spring samples for selected functional KEGG pathways inferred from 16S rRNA gene data using PICRUSt. (A) SK-Y hot spring water sample, (B) Green biofilm sample, (C) Non-decay sample, and (D) Decay sample. (Source: Personal collection)

4.8 Thermophiles and thermozymes involved in foliage degradation

In this study, V3-V4 region of the 16S rRNA gene was amplified on the Illumina MiSeq platforma and majority of detected OTUs (approximately 500 base pairs) are confidently assigned to genus level at blast e-value of 0.001. Region V3-V4 was selected due to its ability to distinguish all bacterial species to the genus level except for closely related Enterobacteriaceae in various environmental communities (Chakravorty et al., 2007; Claesson et al., 2010). As this region is less suitable for species identification due to a higher degree of sequence conservation compared to the other hypervariable regions, genera that are ≥0.85% relative abundance in at least one SK-Y hot spring sample were shortlisted and searched for related literatures and databases. Representative strains of these genera that have complete genomes information are summarised in Table 4.4. The following thermophilic bacteria genera have abundance of carbohydrate-active genes (GH groups of enzymes): *Caldilinea*, *Meiothermus*, *Chloroflexus*, *Roseiflexus*, *Thermoanaerobaculum*, *Melioribacter*, *Geobacillus*, *Desulfurobacterium*, *Thermosipho*, *Thermoanaerobacterium*, *Fervidobacterium*, *Acidimicrobium*, *Chloracidobacterium*, and *Thermoanaerobacter*. For archaeal genera, *Aeropyrum*, *Caldisphaera*, *Methanotorris*, and *Methanothermococcus* are potent lignocellulosic biomass degraders, yet the numbers of total types of GHs (GH 15, 31, 38, 57, 99, and 122) for these archaeal genera are relatively fewer to most bacteria (more than 60 GHs were mapped to the assigned genera) (Table 4.4). Comparison of major biomass degraders found in this study and selected literatures are shown in Table 4.5.

Table 4.4: List of important genera and thermozymes involved in high temperature cellulolytic degradation

Genus[a]	Bacteria[b] (Accession no.)	OGT[c], °C	Opt. pH[c]	GH Groups[d]	Place of Isolation	References
Acidimicrobium	*A. ferrooxidans* DSM 10331 (NC_013124)	45	1.7	5,13,15,18,23,30,39	Hot spring	(Clum et al., 2009)
Aeropyrum	*A. camini* SY1 = JCM 12091 (NZ_BBBZ00000000); *A. pernix* K1 (BA000002)	85-90	7.0-7.8	99,122	Deep-sea hydrothermal vent chimney; marine solfataric vent	(Kawarabayasi et al., 1999)
Caldilinea	*C. aerophila* DSM 14535 = NBRC 104270 (NC_017079)	55	7.0	1,2,3,4,13,16,18,20,23,29,31,33,36,38,39,42,43,51,63,65,77,78,99,116, 127,130	Hot spring; deep hot aquifer	n.a
Caldisphaera	*C. lagunensis* DSM 15908 (NC_019791)	75	4.0-4.5	15,31,38,57	Acidic hot spring	n.a
Chloracidobacterium	*Candidatus C. thermophilum* B (NC_016024; NC_016025)	44-58	5.5-9.5	1,3,13,15,23,27,57,73,77	Hot spring	(Costas et al., 2012)
Chloroflexus	*Chloroflexus* sp. Y-400-fl (NC_012032); *C. aurantiacus* J-10-fl (NC_010175); *C. aggregans* DSM 9485 (NC_011831)	48-55	7.5-8.2	1,2,3,5,9,13,15,16,18,23,31,32,35,38,39,51,57,65,77,78,94,114	Hot spring	n.a.
Desulfurobacterium	*D. thermolithotrophum* DSM 11699 (NC_015185)	60-75	6.0	23,73	Deep-sea hydrothermal vent chimney	(Göker et al., 2011)
Fervidobacterium	*F. islandicum* AW-1 (NZ_CP014334); *F. nodosum* Rt17-B1 (NC_009718); *F. pennivorans* DSM 9078 (NC_017095); *F. pennivorans* DYC (CP011393)	65-75	6.8-7.4	1,3,4,5,13,16,20,23,30,31,35,36,38,42,57,65,94,130	Hot spring mixture of water and mud; branch of wood in hot spring; filter sediment in a district heating pipe	(Lee et al., 2015)

Table 4.4, continued.

Genus[a]	Bacteria[b] (Accession no.)	OGT[c], °C	Opt. pH[c]	GH Groups[d]	Place of Isolation	References
Geobacillus	*G. kaustophilus* HTA426 (NC_006510); *Geobacillus* sp. 12AMOR1 (NZ_CP011832); *Geobacillus* sp. C56-T3 (NC_014206); *Geobacillus* sp. GHH01 (NC_020210); *Geobacillus* sp. JF8 (CP006254); *Geobacillus* sp. JS12 (NZ_CP014749); *Geobacillus* sp. LC300 (NZ_CP008903); *Geobacillus* sp. WCH70 (CP001638); *Geobacillus* sp. Y4.1MC1 (CP002293); *Geobacillus* sp. Y412MC52 (NC_014915); *Geobacillus* sp. Y412MC61 (NC_013411); *G. stearothermophilus* 10 (NZ_CP008934); *G. stearothermophilus* X1 (CP008855); *G. subterraneus* KCTC 3922 (CP014342); *G. thermodenitrificans* NG80-2 (NC_009328); *G. thermoglucosidasius* C56-YS93 (NC_015660); *G. thermoglucosidasius* DSM 2542 (NZ_CP012712); *G. thermoleovorans* CCB_US3_UF5 (NC_016593); *G. thermoleovorans* KCTC 3570 (NZ_CP014335)	45-70	6.8-7.3	1,2,3,4,5,10,13,18,23,27,31,32,36,37,38,39,42,43,51,52,53,65,67,70,73,78,94,127,130	Compost; mud (60 metres underground); hot spring; flax plants; sugar beet juice from extraction installations; hot-gas well; coatings inside tube; formation water of the oil field; evaporated milk; deteriorated canned food; water sludge sample of a hydrothermal geyser outlet	(Brumm et al., 2015b; Brumm et al., 2015a; Brumm et al., 2016; Feng et al., 2007; Shintani et al., 2014; Takami et al., 2004a; Takami et al., 2004b; Wiegand et al., 2013; Wissuwa et al., 2016; Yao et al., 2013)
Meiothermus	*M. silvanus* DSM 9946 (NC_014212); *M. ruber* DSM 1279 (NC_021081)	45-55	7.2-8.2	1,3,4,10,13,15,23,26,31,36,38,42,43,57,63,77,114,125,130	Hot spring (biofilm, water); geothermally heated soil	(Sikorski et al., 2010; Thiel et al., 2015)
Melioribacter	*M. roseus* P3M (NC_018178)	47	7.2-7.8	1,2,3,5,9,10,13,16,20,23,26,27,28,29,30,31,35,43,47,50,51,53,55,67,77,88,92,94,97,105,106,115,125,127,130	Microbial mat from the wooded surface of a chute under the flow of thermal water coming out of an oil exploration well	(Kadnikov et al., 2013)

Table 4.4, continued.

Genus[a]	Bacteria[b] (Accession no.)	OGT[c], °C	Opt. pH[c]	GH Groups[d]	Place of Isolation	References
Methanothermococcus	*M. okinawensis* IH1 (NC_015636)	65	6.5-7.0	15,57	Hydrothermal vent chimney; heated sea-shore	n.a
Methanotorris	*M. igneus* Kol 5 (NC_015562)	70-85	6.5-7.2	15,57	Black smoker; marine hydrothermal system	n.a
Roseiflexus	*R. castenholzii* DSM 13941 (NC_009767); *Roseiflexus* sp. RS-1 (NC_009523)	48	7.5	1,2,3,4,5,13,15,18,20, 23,29,36,38,39,51,57, 73,77,78,94	Bacterial mat in hot spring	n.a
Thermoanaerobacter	*T. brockii* subsp. *finnii* Ako-1 (NC_014964); *T. italicus* Ab9 (NC_013921); *T. kivui* DSM 2030 (NZ_CP009170); *T. mathranii* subsp. *mathranii* str. A3 (NC_014209); *T. pseudethanolicus* ATCC 33223 39E (NC_010321); *Thermoanaerobacter* sp. X513 (NC_014538); *Thermoanaerobacter* sp. X514 (NC_010320); *T. tengcongensis* MB4 (NC_003869); *T. wiegelii* Rt8.B1 (NC_015958)	60-70	6.9-7.4	1,2,3,4,5,10,13,15,18, 20,23,26,28,30,31,32, 36,38,43,51,52,65,66, 67,94, 105,127,130	Cow dung compost; hot spring; acidic volcanic steam outlet; continuously stirred tank reactor for hydrogen production	(Hess et al., 2014)
Thermoanaerobacterium	*T. saccharolyticum* JW/SL-YS485 (CP003184); *T. thermosaccharolyticum* DSM 571 (NC_014410); *T. thermosaccharolyticum* M0795; *T. xylanolyticum* LX-11 (NC_015555)	55-65	6.0-7.8	1,2,3,4,5,10,11,13,15, 16,18,23,25,26,28,29, 31,32,36,39,42,43,51, 52,53,65,66,67,81,84, 94,105,116,120,127,1 30	Hydrothermal vent; hot spring; corn-canning waste; salt lagoon sediment; tartrate infusion of grape residue; pond sediment; cow dung compost; fruit juice waste products	(Delver et al., 1996)
Thermoanaerobaculum	*T. terrenum* ATCC BAA-798 (NC_013525; NC_013526)	60	7.2-7.4	1,2,3,4,5,10,13,15,29, 32,36,38,39,42,43,51	Hot spring	(Kiss et al., 2010)

Table 4.4, continued.

Genus[a]	Bacteria[b] (Accession no.)	OGT[c], °C	Opt. pH[c]	GH Groups[d]	Place of Isolation	References
Thermosipho	*T. africanus* TCF52B (NC_011653); *T. melanesiensis* BI429 (NC_009616)	60-75	6.3-7.2	1,2,3,4,13,20,23,31,36, 38,57,73,85,130	Hydrothermal vent; spring; production water from an oil well; *Riftia pachyptila* sheath scraping	(Antoine et al., 1997; Nesbø et al., 2009)

Notes:
[a] These genera were found to be ≥0.85% in SK-Y hot spring
[b] Representative strain(s) with complete genome(s) information
[c] Optimum growth temperature (OGT) and optimum pH were obtained from The Bacterial Diversity Metadatabase (http://bacdive.dsmz.de)
[d] List of GH families obtained from Carbohydrate-Active EnZymes database (http://www.cazy.org)

Table 4.5: Comparison of the potential thermophilic biomass degraders in different experimental setups (Lee et al., 2018)

Source	Potential biomass degraders	Analysis approaches	Temp., (°C)	pH	Biomass substrates/source	References
Submerged foliage and hot spring water of SK-Y hot spring, Perak, Malaysia	*Acidimicrobium, Aeropyrum, Caldilinea, Caldisphaera, Chloracidobacterium, Chloroflexus, Desulfurobacterium, Fervidobacterium, Geobacillus, Meiothermus, Melioribacter, Methanothermococcus, Methanotorris, Roseiflexus, Thermoanaerobacter, Thermoanaerobacterium, Thermoanaerobaculum, Thermosipho*	Cultivation-independent	60-70	7.5-8.6	Plant litter	This study
Sediments from hot spring, Xiamen, China	*Geobacillus*, *Thermus*, *Bacillus*, *Anoxybacillus*	Enrichment	50-80	7.0	Sugarcane bagasse	(Zhao et al., 2017)
Mixture of water and sediment from SK main stream hot spring, Perak, Malaysia	*Aciduliprofundum, Caloramator, Hydrogenobacter, Ignavibacterium, Melioribacter, Methanocaldococcus, Methanocella, Methanothermus, Methylacidiphilum, Thermodesulfovibrio, Thermotoga, Thermus*	Cultivation-independent	50-110	7.0-9.0	Scattered plant litter	(Chan et al., 2015)
Soil contacting regions of a bagasse pile at Phu Khieo Bio-Energy Chaiyaphum province, Thailand	Actinobacteria, Bacteroidetes/Chlorobi, Chlamydiae/Verrucomicrobia, Chloroflexi, Fibrobacteres/Acidobacteria, Firmicutes, Planctomycetes, Proteobacteria	Cultivation-independent	50	n.a	Sugarcane bagasse	(Mhuantong et al., 2015)
Vegetated area of Obsidian Pool (site OBP 10), Yellowstone National Park	*Anaerobacter, Caldicellulosiruptor, Caloramator, Clostridium, Thermoanaerobacter*	Cultivation-independent and enrichment	55-85	5	*In situ* sampling (*Juncus tweedyi*); enrichment (Avicel, xylan, switchgrass, *Populus*)	(Vishnivetskaya et al., 2015)
Anaerobic digestion sludge collected from Shek Wu Hui wastewater treatment plant, Hong Kong, China	Anaerolineales, Bacteroidales, Clostridiales, Methanobacteriales, Methanosarcinales, Thermotogales	Enrichment	55	6.0-7.0	Microcrystalline cellulose with glucose	(Xia et al., 2014)

Table 4.5, continued.

Source	Potential biomass degraders	Analysis approaches	Temp., (°C)	pH	Biomass substrates/source	References
Sediment and water column of Great Boiling Spring, Nevada	Archaeoglobales, Desulfurococcales, *Dictyoglomus, Thermotoga*	*In situ* enrichment	74-85	5	Ammonia fiber explosion-treated corn stover and aspen shavings	(Peacock et al., 2013)
Switchgrass-adapted bacterial consortia	*Paenibacilli* spp., *Rhodothermus marinus, Thermobispora bispora, Thermomicrobia* sp., *Thermus thermophilus*	Enrichment	60	n.a	Microcrystalline cellulose	(Park et al., 2012)

CHAPTER 5: DISCUSSION

5.1 Microbial diversity analysis in SK-Y hot spring

The 16S rRNA gene amplicon sequencing for microbial diversity elucidations have been widely used in various engineered or environmental samples (Chan et al., 2015; Hess et al., 2011; Mhuantong et al., 2015; Singh et al., 2014; Vishnivetskaya et al., 2015). Culture independent studies at hot springs in Yellowstone National Park (YNP) have a long research history where clonal library was initially used (Brock, 1967; Thiel et al., 2016), and lately microbial diversity in Octopus and Mushroom Springs at YNP were revisited using high throughput NGS (Thiel et al., 2016).

Although SK main stream and SK-Y hot springs are approximately 10 metres away, the dominated microbial diversity between them are dissimilar (Table 4.5), probably due to several factors such as physiochemical or geochemical structure, temperature, dissolved oxygen level, and the quantity of plant litters. Yet, SK and SK-Y hot springs have rich microbial diversity as both sites were dominated by various genera. Often, microbial alpha diversity are higher in circumneutral or slight alkaline hot springs than those of acidic sites (Sharp et al., 2014). Furthermore, high taxonomical and functional diversity are likely driven by the abundance of micronutrients and organic materials, thus SK-Y hot spring is an important site to discover thermophilic species with robust cellulolytic capabilities.

SK-Y hot spring was studied in this work because it represents a natural 'biomass degrading bioreactor' which containing abundant of foliage embedded in the hot spring water. The top surface of the submerged foliage bed was covered by green biofilm. The microbial diversity analysis shows that Cyanobacteria (14.7%), Proteobacteria (14.4%), and Chloroflexi (13.1%) were the major three phyla that contributed to the green biofilm

communities. Cyanobacteria and Chloroflexi are chlorophyll-based phototrophs bacteria and are highly affected by the temperature, pH, sulphide concentration, sunlight and/or chemical gradients (Klatt et al., 2013). High sulphide concentration (12.3 mg L^{-1}) with different chemical compositions (Table 4.1) potentially led to the increases in the abundance of chlorophototrophic compared to our previously studied SK hot spring (sulphide, 0.2 mg L^{-1}) which detected low abundance of chlorophototrophs (Chan et al., 2015). Moreover, the stagnant spring water of SK-Y hot spring has flavoured the green biofilm formation compared to the fast-flowing SK hot spring.

This study focused on identification of thermophiles involved in foliage degradation at the deeper part of the spring. The dissolved oxygen level at the base of SK-Y hot spring is expected to be low as spring head temperature could reach 90°C. The present study shows that Firmicutes and Proteobacteria represented the dominated phyla in the microbiota of degraded foliage. This might be due to higher cellulose content of 7.0%-11.3% among *Vitex*, *Ficus*, and *Stenochlaena* foliage samples compared to hemicellulose content of 2.9%–4.3%. This is supported by Eichorst et al. (2013) whereby Firmicutes are the primary degraders of cellulose in laboratory enrichment experiment. Members of Firmicutes included *Geobacillus*, *Thermoanaerobacter*, *Thermoanaerobacterium*, *Candidatus Desulforudis*, and *Caldicellulosiruptor*, which could participate in different stages of halocellulose degradation as reported elsewhere (Bhalla et al., 2013; Cobucci-Ponzano et al., 2015; De Maayer et al., 2014; Eichorst et al., 2013; Vishnivetskaya et al., 2015). In the decomposition of sugarcane bagasse waste at 50°C, the predominance phylum is Proteobacteria (Mhuantong et al., 2015). Thus, the reports elucidated that Firmicutes and Proteobacteria are the important phyla for biomass degradation at high temperature. Another important bacterial component in SK-Y hot spring was Acidobacteria, particularly the major genus of *Thermoanaerobaculum* (7.3%),

a chemo-organotrophic that thrived in anaerobic habitats (Losey et al., 2013). Acidobacteria exclusively or preferentially use organic substrates as an energy.

Some of the thermophiles in SK-Y hot spring have been noticed too in natural geothermal or heated laboratory setups related to biomass degradation (Table 4.5). For instance, the microbial community in heated sediments surrounding vegetated (*Juncus tweedyi*) wetland (68°C) in Obsidian Pool (site OBP10) in YNP was mainly constituted of phyla Firmicutes, Proteobacteria, Aquificae, Deinococcus-Thermus, Spirochaetes, Verrucomicrobia, and a huge proportion of unclassified bacteria. Majority of Firmicutes members including lignocellulolytic degraders *Clostridium*, *Anaerobacter*, *Caloramator*, *Caldicellulosiruptor*, and *Thermoanaerobacter* (Vishnivetskaya et al., 2015). In the same report, when OBP10 samples were inoculated with various lignocellulolytic materials (Avicel, switchgrass, *Populus*, and xylan) and incubated at 55-85°C in anaerobic laboratory setup. The main bacteria after three passages of culturing were *Thermoanaerobacter*, *Caloramator*, *Caldicellulosiruptor*, *Clostridium*, *Dictyoglomus*, and *Fervidobacterium*; and their distributions in these experiments varied with experimental parameters such as temperature and substrates (Vishnivetskaya et al., 2015).

In another report, Anaerolineales, Clostridiales, Bacteroidales, Thermotogales, Methanobacteriales, and Methanosarcinales were the major orders in an anaerobic digestion sludge enriched with microcrystalline cellulose at 55°C (Xia et al., 2014). In a separate study, Park et al. (2012) suggested that *Thermus thermophilus*, *Paenibacillus* spp., *Rhodothermus marinus*, *Thermobispora bispora*, and *Thermomicrobia* sp. are important thermophiles for microcrystalline cellulose degradation at 60°C. This result is similar to Eichorst et al. (2013) which reported that Firmicutes, Bacteroidetes, and Thermus are important for cellulose degradation. A different microbiota community was

observed in degradation (55°C, aerobic) of switchgrass under high-solid condition (Yu et al., 2015), where the dominant phyla included Chloroflexi, Proteobacteria, Bacteroidetes, and Actinobacteria. Great Boiling Spring (GBS) is a hot spring located at Nevada (77-85°C) (Peacock et al., 2013). This site generally lacks of lignocellulosic plant materials. A microbial diversity analysis was conducted to compare the microbial diversity in GBS water-sediments as well as man-made *in situ* enrichment using ammonia fiber explosion-treated corn stover and aspen shavings. The biomass enriched samples consisted of potential biomass degraders, sugar fermenters, and hydrogenotrophs such as *Thermotoga*, *Dictyoglomus*, Desulfurococcales, and Archaeoglobales. Interestingly, the microbial flora in biomass enriched samples and GBS indigenous samples were different, therefore Peacock et al. (2013) suggested that the additional lignocellulosic biomass stimulated the growth of the potent biomass degraders in a natural environment. The current data in this work agreed with aforementioned observation since microbiota in SK-Y hot spring water sample alone is different to foliage samples.

The main purpose of current work summarises a list of important thermophiles for biomass degradation at high temperature and circumneutral pH. The major potential biomass degraders found in this study are slightly different from the other results (Table 4.5), but my list is more comprehensive. During the time course of biomass decomposition process, compositions of phyla could be altered (Eichorst et al., 2013; Yu et al., 2015). The length of time needed for new fallen foliage to reach degradation stage in SK-Y hot spring was uncertained. Yet, microorganisms attached on top layer of the submerged plant litter bed (i.e., submerged foliage with no apparent biofilm and green biofilm attached on submerged foliage) could mimic taxa that are involved in the early stage of degradation process. Slowly over a certain period of time, these foliage will be covered and compressed by new layers of fallen plant litters, and the initial foliage

eventually would be occupied by a slightly different microbiota community to complete the degradation (i.e. the decay sample in this study).

5.2 Biomass-degrading metabolic potential

Hydrolysis of cellulolytic materials involved various groups of GHs (Singh et al., 2014). A complete list of important GHs and lignin acting enzymes (or known as auxiliary enzymes, AA) for complete biomass decomposition is available in several earlier publications (Banerjee et al., 2010b; Mhuantong et al., 2015; van der Lelie et al., 2012). Synergistic actions of the different types of core halocellulolytic enzymes, such as the exoglucanase, endoglucanase, cellobiase, endo-β-1,4-xylanase, xylan-1,4-β-xylosidase, α-L-arabinofuranosidase, α-glucuronidase, acetylxylan esterase, feruloyl esterase, mannan endo-1,4-β-mannanase, β-1,4-mannosidase, and arabinan endo-1,5-α-L-arabinosidase, are crucial for efficient hydrolysis of halocellulose to xylose or glucose (Van Dyk & Pletschke, 2012).

Several publications have elucidated some examples of thermophiles for decomposing biomass at elevated temperatures (Eichorst et al., 2013; Mhuantong et al., 2015; Park et al., 2012; Vishnivetskaya et al., 2015; Yu et al., 2015). The present study is an attempt to investigate the functional profiles using PICRUSt and CAZy database to determine the composition of thermophilic microbial communities (Table 4.4). Genera with incomplete genome data, or those with lower OGT (<40°C), or OTUs which are unable to be classified confidently to genus level, or OTUs lower than 0.85% of total population were excluded from the analysis. At this point, we assume that species or strains within the same genus will have similar genomic content and thus exhibit homologous genes for cellulolytic enzymes. Despite many of the OTUs for SK-Y hot spring are able to assign at species level, yet the affiliation may be ambiguous, and

therefore we are not able to pinpoint the actual species present in SK-Y hot spring. Due to this limitation, a general overview of foliage degradation at high temperature was provided using complete genomes information available in databases or literatures. Shotgun metagenome sequencing and culture collections of SK-Y hot spring is yet to be performed to validate the information provided in Table 4.4.

The abundance of sequences associated with specific carbohydrates metabolism from KEGG annotation is illustrated in Figure 4.7. A total of twenty-four carbohydrates and energy pathways were identified in SK-Y hot spring, with a substantial overlap between the bacterial and archaeal data sets. *In silico* PICRUSt analysis shows the increase of the prevalence of functional genes encoding carbohydrate and energy metabolic pathways to convert cellulose, hemicellulose, and lignin into various types of monosaccharides that are essential aerobic energy sources via the citrate cycle (TCA) (Mhuantong et al., 2015). Carbohydrate catabolic pathways, including glycolysis, oxidative pentose phosphate pathway, and TCA, are active mainly during the dark phase of the light-dark cycle (Quintana et al., 2011). These pathways are responsible for producing NAD(P)H and other biosynthetic metabolites that are involved in the normal cellular functions (Quintana et al., 2011). The pentoses, xylose, and arabinose are metabolized by the non-oxidative pentose phosphate pathway in thermophilic bacteria, and are later converted into intermediates within the glycolysis pathway (Chang & Yao, 2011; Zeikus et al., 1980). The different carbohydrates metabolic pathways in SK-Y hot spring may be explained by the high diversity of thermophilic biomass degraders, which account for at least 18 genera, based on CAZy analysis.

Among many GH families, GH 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 61, 74, and 124 which comprised of endoglucanases, exoglucanases, and beta-glucosidases are important

biocatalyst to degrade cellulose (Mhuantong et al., 2015; van der Lelie et al., 2012). The intermediate products of cellulose (i.e., cello-oligosaccharides and cellobiose) are further processed by the β-glucosidase from GH 1, 3, 9, 30, and 116 involving the hydrolysis of beta-linked dimers of oligosaccharides (Mhuantong et al., 2015; van der Lelie et al., 2012). Some of the candidates listed in Table 4.4 have been well characterised, for instances, *Melioribacter* (Podosokorskaya et al., 2013) and *Thermoanaerobacterium* (Currie et al., 2014). Majority of the dominated bacterial genera in SK-Y hot spring are abundance of GH enzymes, thus again supported the findings that the genera listed in Table 4.4 are generally important for biomass degradation at circumneutral pH and high temperature. Apparently, our data also suggested that bacteria consortium instead of archaea plays a more important role in biomass degradation as genomes of archaeal genera harbour lower numbers of GH enzymes. Generally, the genome size of archaea is relatively smaller than bacteria, and probably due to genome streaming process, many genes for GH enzymes have been omitted (Urbieta et al., 2015).

According to Mhuantong et al. (2015), the AA 1 (multicopper oxidase), AA 3 (choline dehydrogenase), AA 4 (vanillyl-alcohol oxidase), AA 5 (copper radical oxidase), AA 6 (1,4-benzoquinone reductase), AA 7 (glucooligosaccharide oxidases), and AA 9 (copper-dependent monooxygenases) are crucial in lignin degradation. Besides that, AA 2 (dye-decolorising peroxidases, homologue to fungi lignolytic peroxidases) is an enzyme that has the ability to degrade lignin (de Gonzalo et al., 2016). So far, most well studied AA enzymes are originated from fungi (Karnaouri et al., 2014; Levasseur et al., 2013). Based on the complete genomes information shown in Table 4.4, only three AA families (i.e., AA 1 from *Melioribacter roseus* P3M, AA3 from *Meiothermus ruber* DSM 1279, and AA 6 from *Geobacillus stearothermophilus* X1 and *Geobacillus thermoglucosidasius* DSM 2542) were identified in SK-Y hot spring. Based on CAZy

classification for auxiliary enzymes, genes for AA are lacking or missing for most of the thermophiles listed in Table 4.4. Several *Thermus* OTUs were identified in SK-Y hot spring as minority taxa. A thermostable and chloride-tolerant laccase (AA 1, E.C 1.10.3.2) from *Thermus thermophilus* SG0.5JP17-16 was lately cloned and overexpressed (Liu et al., 2015). Interestingly, based on information provided in CAZy, the genome sequence of *T. thermophilus* SG0.5JP17-16 was lacked of laccase. This example of contradictory could suggest to us that AA database may require further validation by curators and researchers. In fact, the total numbers of identified AA enzymes in genomic sequencing projects are indeed lower than well-established GHs that have been categorised in according to CAZy and PeroxiBase Database (Fawal et al., 2013). By further examining the minority genera (<0.85%) detected in SK-Y hot spring, we managed to identify the presence of some AA genes in some sequenced genomes. Examples of these include *Nocardioides* (AA 3), *Amycolatopsis* (AA 3, AA 6, AA 7, and AA 10), *Azoarcus* (AA 3, AA 4, and AA 6), *Pseudoxanthomonas* (AA 3 and AA 6), and other AA enzymes present in *Anoxybacillus*, *Thermodesulfobacteriaceae*, *Clostridium*, *Geobacter*, *Bacillus*, and some other minority members. Despite *Bacillus* spp. are generally mesophilic, proteins from *Bacillus* could be thermostable, for instance laccase from *Bacillus licheniformis* ATCC 9945a had an optimum temperature of 90°C and a half-life of 50 minutes at 70°C (Lončar et al., 2016). We are not able to rule out the presence of fungi or enzymes from fungi in SK-Y hot spring that may help in partial lignin removal. Nevertheless, as the average temperature of SK-Y hot spring is relatively high, the presence of fungi and the stability of its enzymes are doubtful.

CHAPTER 6: CONCLUSION

The work presented here describes a 'natural biomass degrading bioreactor' as one of the hot springs with the highest temperature in Malaysia which remains largely under-exploited, but it possesses a great potential for the biofuel, fermentation, and cellulosic biorefinery industries. It also provides a unique ecological niche, as the microorganisms in this lignocellulose-rich environment are able to thrive in relatively high temperature with varying microenvironments of aerobic surface to anoxic interior. The microbial community in this SK-Y hot spring thus presents a good resource for the discovery and characterization of new biomass-degrading enzymes. Detailed 16S rRNA gene amplicon sequencing using high-throughput Illumina MiSeq platform revealed a diverse and distinct landscape of bacterial and archaeal fauna in both water sample and foliage samples with prevalence of the bacterial phyla Proteobacteria, Chloroflexi, Firmicutes, and Cyanobacteria; and archaeal phyla Thaumarchaeota, Euryarchaeota, Crenarchaeota, and *Candidatus* Korarchaeota. Collectively, biomass degradation at high temperature may involve various community structure and factors such as temperature, dissolved oxygen level, type of biomass, and stage of degradation (or complexity of substrate) which will greatly influence the thermophiles population. The current study also discusses the analyses of carbohydrate active enzymes from thermophiles involved in plant biomass degradation by a microbial community speculated from the SK-Y hot spring. The metagenomic data from 16S rRNA sequences demonstrated that the microbial population contained many taxa known for their ability to degrade lignocellulosic biomass. A total of 18 genera from the microbial community are considered promising candidates for the mining of novel enzymatic cocktails for the conversion of lignocellulosic biomass into fermentable sugars. Analysis of the CAZy helped in identification of more than 60 GH enzymes, which needs to be further explored regarding their collaboration and interaction for plant biomass degradation in this microbial

community. Despite the relatively easy access to their genes, biochemical characterizations of all enzymes for biomass decomposition at high temperature is still under explored. We showed here the potential biomass degraders through bioinformatics analysis and demonstrated how these potent biomass degraders are important for the mining of broad range of lignocellulose-degrading enzymes for the production of enzyme cocktails.

To improve hydrolysis ability, enzymes with better desirable properties and enzyme systems with optimized compositions can be obtained using high-throughput screening approaches. Furthermore, deep understanding of the biological process for lignocellulolytic enzyme production will provide critical targets for rational strain engineering for high productivities. Nevertheless, low enzymatic digestibility of lignocelluloses (a scientific problem) and the immature platform for systematic engineering of bacterial strains (a technical problem) are still the major barriers for the present work.

REFERENCES

Antoine, E., Cilia, V., Meunier, J., Guezennec, J., Lesongeur, F., & Barbier, G. (1997). *Thermosipho melanesiensis* sp. nov., a new thermophilic anaerobic bacterium belonging to the order Thermotogales, isolated from deep-sea hydrothermal vents in the southwestern Pacific Ocean. *International Journal of Systematic Bacteriology, 47*(4), 1118-1123.

Asgher, M., Bashir, F., & Iqbal, H. M. N. (2014). A comprehensive ligninolytic pre-treatment approach from lignocellulose green biotechnology to produce bio-ethanol. *Chemical Engineering Research and Design, 92*(8), 1571-1578.

Badhai, J., Ghosh, T. S., & Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Frontiers in Microbiology, 6*, 1166.

Balat, M. (2011). Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review. *Energy Conversion and Management, 52*(2), 858-875.

Baldrian, P. (2006). Fungal laccases–occurrence and properties. *FEMS Microbiology Reviews, 30*(2), 215-242.

Banerjee, G., Car, S., Scott-Craig, J. S., Borrusch, M. S., Aslam, N., & Walton, J. D. (2010b). Synthetic enzyme mixtures for biomass deconstruction: Production and optimization of a core set. *Biotechnology and Bioengineering, 106*(5), 707-720.

Banerjee, G., Scott-Craig, J. S., & Walton, J. D. (2010a). Improving enzymes for biomass conversion: A basic research perspective. *Bioenergy Research, 3*(1), 82-92.

Barnard, D., Casanueva, A., Tuffin, M., & Cowan, D. (2010). Extremophiles in biofuel synthesis. *Environmental Technology, 31*(8-9), 871-888.

Barott, K. L., Rodriguez-Mueller, B., Youle, M., Marhaver, K. L., Vermeij, M. J., Smith, J. E., & Rohwer, F. L. (2012). Microbial to reef scale interactions between the reef-building coral *Montastraea annularis* and benthic algae. *Proceedings of the Royal Society B, 279*(1733), 1655-1664.

Bayer, E. A., Chanzy, H., Lamed, R., & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology, 8*(5), 548-557.

Beeson, W. T., Phillips, C. M., Cate, J. H., & Marletta, M. A. (2011). Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *Journal of the American Chemical Society, 134*(2), 890-892.

Bey, M., Zhou, S., Poidevin, L., Henrissat, B., Coutinho, P. M., Berrin, J. G., & Sigoillot, J. C. (2013). Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from *Podospora anserina*. *Applied and Environmental Microbiology, 79*(2), 488-496.

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., & Sani, R. K. (2013). Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology, 128*, 751-759.

Brock, B. J., Rieble, S., & Gold, M. H. (1995). Purification and characterization of a 1, 4-benzoquinone reductase from the basidiomycete *Phanerochaete chrysosporium*. *Applied and Environmental Microbiology, 61*(8), 3076-3081.

Brock, T. D. (1967). Micro-organisms adapted to high temperatures. *Nature, 214*(5091), 882-885.

Brock, T. D. (2012). *Thermophilic microorganisms and life at high temperatures*. Berlin: Springer Science & Business Media.

Brumm, P., Land, M. L., Hauser, L. J., Jeffries, C. D., Chang, Y.-J., & Mead, D. A. (2015b). Complete genome sequences of *Geobacillus* sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences, 10*, 81.

Brumm, P. J., Land, M. L., & Mead, D. A. (2015a). Complete genome sequence of *Geobacillus thermoglucosidasius* C56-YS93, a novel biomass degrader isolated from Obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences, 10*, 73.

Brumm, P. J., Land, M. L., & Mead, D. A. (2016). Complete genome sequences of *Geobacillus* sp. WCH70, a thermophilic strain isolated from wood compost. *Standards in Genomic Sciences, 11*, 33.

Cai, L., Ye, L., Tong, A. H. Y., Lok, S., & Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE, 8*(1), e53649.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335-336.

Chaffron, S., Rehrauer, H., Pernthaler, J., & von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research, 20*(7), 947-959.

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods, 69*(2), 330-339.

Chan, C. S., Chan, K. G., Tay, Y. L., Chua, Y. H., & Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Frontiers in Microbiology, 6*, 177.

Chan, C. S., Sin, L. L., Chan, K. G., Shamsir, M. S., Manan, F. A., Sani, R. K., & Goh, K. M. (2016). Characterization of a glucose-tolerant β-glucosidase from *Anoxybacillus* sp. DT3-1. *Biotechnology for Biofuels, 9*, 174.

Chang, T., & Yao, S. (2011). Thermophilic, lignocellulolytic bacteria for ethanol production: Current state and perspectives. *Applied Microbiology and Biotechnology, 92*(1), 13-27.

Cho, I., & Blaser, M. J. (2012). The human microbiome: At the interface of health and disease. *Nature Reviews Genetics, 13*(4), 260-270.

Claesson, M. J., Wang, Q., O'sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., & O'toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research, 38*(22), e200.

Clum, A., Nolan, M., Lang, E., Rio, T. G., Tice, H., Copeland, A., . . . Bruce, D. (2009). Complete genome sequence of *Acidimicrobium ferrooxidans* type strain (ICP^T). *Standards in Genomic Sciences, 1*(1), 38-45.

Cobucci-Ponzano, B., Strazzulli, A., Iacono, R., Masturzo, G., Giglio, R., Rossi, M., & Moracci, M. (2015). Novel thermophilic hemicellulases for the conversion of lignocellulose for second generation biorefineries. *Enzyme and Microbial Technology, 78*, 63-73.

Codron, D., Lee-Thorp, J. A., Sponheimer, M., & Codron, J. (2007). Nutritional content of savanna plant foods: Implications for browser/grazer models of ungulate diversification. *European Journal of Wildlife Research, 53*, 100-111.

Collins, R. E., & Higgs, P. G. (2012). Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Molecular Biology and Evolution, 29*(11), 3413-3425.

Costas, G., Amaya, M., Liu, Z., Tomsho, L. P., Schuster, S. C., Ward, D. M., & Bryant, D. A. (2012). Complete genome of *Candidatus Chloracidobacterium thermophilum*, a chlorophyll-based photoheterotroph belonging to the phylum Acidobacteria. *Environmental Microbiology, 14*(1), 177-190.

Currie, D., Guss, A. M., Herring, C., Giannone, R. J., Johnson, C. M., Lankford, P. K., . . . Lynd, L. R. (2014). Profile of secreted hydrolases, associated proteins, and SlpA in *Thermoanaerobacterium saccharolyticum* during the degradation of hemicellulose. *Applied and Environmental Microbiology, 80*(16), 5001-5011.

de Gonzalo, G., Colpa, D. I., Habib, M. H., & Fraaije, M. W. (2016). Bacterial enzymes involved in lignin degradation. *Journal of Biotechnology, 236*, 110-119.

de Jong, E., van Berkel, W. J., van der Zwan, R. P., & de Bont, J. A. (1992). Purification and characterization of vanillyl-alcohol oxidase from *Penicillium simplicissimum*. A novel aromatic alcohol oxidase containing covalently bound FAD. *European Journal of Biochemistry, 208*(3), 651-657.

De Maayer, P., Brumm, P. J., Mead, D. A., & Cowan, D. A. (2014). Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics, 15*, 836.

DeCastro, M.-E., Rodríguez-Belmonte, E., & González-Siso, M.-I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Frontiers in Microbiology, 7*, 1521.

Delver, E., Belogurova, N., Tupikova, E., Varfolomeyev, S., & Belogurov, A. (1996). Characterization, sequence and mode of replication of plasmid pNB2 from the thermophilic bacterium *Clostridium thermosaccharolyticum*. *Molecular and General Genetics, 253*(1), 166-172.

Demain, A. L., Newcomb, M., & Wu, J. D. (2005). Cellulase, Clostridia, and ethanol. *Microbiology and Molecular Biology Reviews, 69*(1), 124-154.

Eastwood, D. C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., . . . Bendiksby, M. (2011). The plant cell wall–decomposing machinery underlies the functional diversity of forest fungi. *Science, 333*(6043), 762-765.

Eichorst, S. A., Varanasi, P., Stavila, V., Zemla, M., Auer, M., Singh, S., . . . Singer, S. W. (2013). Community dynamics of cellulose-adapted thermophilic bacterial consortia. *Environmental Microbiology, 15*(9), 2573-2587.

Fan, X., Liu, X., Wang, K., Wang, S., Huang, R., & Liu, Y. (2011). Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. *Journal of Molecular Catalysis B: Enzymatic, 72*(3-4), 319-326.

Fan, Z., Oguntimein, G. B., & Reilly, P. J. (2000). Characterization of kinetics and thermostability of *Acremonium strictum* glucooligosaccharide oxidase. *Biotechnology and Bioengineering, 68*(2), 231-237.

Fawal, N., Li, Q., Savelli, B., Brette, M., Passaia, G., Fabre, M., . . . Dunand, C. (2013). PeroxiBase: A database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Research, 41*(Database issue), D441-D444.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., . . . Peng, X. (2007). Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America, 104*(13), 5602-5607.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., . . . Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America, 109*(52), 21390-21395.

Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunæs, A. C., Stenstrøm, Y., MacKenzie, A., . . . Eijsink, V. G. (2011). Cleavage of cellulose by a CBM33 protein. *Protein Science, 20*(9), 1479-1483.

Fuciños, P., Atanes, E., López-López, O., Solaroli, M., Cerdán, M. E., González-Siso, M. I., . . . Rúa, M. L. (2014). Cloning, expression, purification and characterization of an oligomeric His-tagged thermophilic esterase from *Thermus thermophilus* HB27. *Process Biochemistry, 49*(6), 927-935.

Georgieva, T. I., Mikkelsen, M. J., & Ahring, B. K. (2008). Ethanol production from wet-exploded wheat straw hydrolysate by thermophilic anaerobic bacterium *Thermoanaerobacter* BG1L1 in a continuous immobilized reactor. *Applied Biochemistry and Biotechnology, 145*(1-3), 99-110.

Giffhorn, F. (2000). Fungal pyranose oxidases: Occurrence, properties and biotechnical applications in carbohydrate chemistry. *Applied Microbiology and Biotechnology, 54*(6), 727-740.

Gladden, J. M., Allgaier, M., Miller, C. S., Hazen, T. C., VanderGheynst, J. S., Hugenholtz, P., . . . Singer, S. W. (2011). Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. *Applied and Environmental Microbiology, 77*(16), 5804-5812.

Gladden, J. M., Eichorst, S. A., Hazen, T. C., Simmons, B. A., & Singer, S. W. (2012). Substrate perturbation alters the glycoside hydrolase activities and community composition of switchgrass-adapted bacterial consortia. *Biotechnology and Bioengineering, 109*(5), 1140-1145.

Goh, K. M., Chan, K. G., Lim, S. W., Liew, K. J., Chan, C. S., Shamsir, M. S., . . . Adrian, T. G. S. (2016). Genome analysis of a new Rhodothermaceae strain isolated from a hot spring. *Frontiers in Microbiology, 7*, 1109.

Göker, M., Daligault, H., Mwirichia, R., Lapidus, A., Lucas, S., Deshpande, S., . . . Goodwin, L. (2011). Complete genome sequence of the thermophilic sulfur-reducer *Desulfurobacterium thermolithotrophum* type strain (BSA^T) from a deep-sea hydrothermal vent. *Standards in Genomic Sciences, 5*(3), 407-415.

Hamilton-Brehm, S. D., Mosher, J. J., Vishnivetskaya, T., Podar, M., Carroll, S., Allman, S., . . . Elkins, J. G. (2010). *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Applied and Environmental Microbiology, 76*(4), 1014-1020.

Hemsworth, G. R., Henrissat, B., Davies, G. J., & Walton, P. H. (2014). Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature Chemical Biology, 10*(2), 122-126.

Henriksson, G., Johansson, G., & Pettersson, G. (1993). Is cellobiose oxidase from *Phanerochaete chrysosporium* a one-electron reductase? *Biochimica et Biophysica Acta, 1144*(2), 184-190.

Henriksson, G., Johansson, G., & Pettersson, G. (2000). A critical review of cellobiose dehydrogenases. *Journal of Biotechnology, 78*(2), 93-113.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., . . . Zhang, T. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science, 331*(6016), 463-467.

Hess, V., Poehlein, A., Weghoff, M. C., Daniel, R., & Müller, V. (2014). A genome-guided analysis of energy conservation in the thermophilic, cytochrome-free acetogenic bacterium *Thermoanaerobacter kivui*. *BMC Genomics, 15*, 1139.

Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgard, M., & Rusch, D. B. (2013). The YNP metagenome project: Environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Frontiers in Microbiology, 4*, 67.

Jensen Jr, K. A., Ryan, Z. C., Vanden Wymelenberg, A., Cullen, D., & Hammel, K. E. (2002). An NADH: Quinone oxidoreductase active during biodegradation by the brown-rot basidiomycete *Gloeophyllum trabeum*. *Applied and Environmental Microbiology, 68*(6), 2699-2703.

Jiang, B., Liang, X., Chen, Y., Ma, T., Liu, L., Li, J., . . . Li, S. (2012). Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. *Scientific Reports, 2*, 936.

Kadnikov, V. V., Mardanov, A. V., Podosokorskaya, O. A., Gavrilov, S. N., Kublanov, I. V., Beletsky, A. V., . . . Ravin, N. V. (2013). Genomic analysis of *Melioribacter roseus*, facultatively anaerobic organotrophic bacterium representing a novel deep lineage within Bacteriodetes/Chlorobi group. *PLoS ONE, 8*(1), e53047.

Karnaouri, A., Topakas, E., Antonopoulou, I., & Christakopoulos, P. (2014). Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila*. *Frontiers in Microbiology, 5*, 281.

Kawaguchi, H., Hasunuma, T., Ogino, C., & Kondo, A. (2016). Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks. *Current Opinion in Biotechnology, 42*, 30-39.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., . . . Ankai, A. (1999). Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes, 6*(2), 83-101.

Keshk, S. M. (2016). Cellulase application in enzymatic hydrolysis of biomass. *New and Future Developments in Microbial Biotechnology and Bioengineering*, 185-191.

Kiss, H., Cleland, D., Lapidus, A., Lucas, S., Del Rio, T. G., Nolan, M., . . . Pitluck, S. (2010). Complete genome sequence of '*Thermobaculum terrenum*' type strain ($YNP1^T$). *Standards in Genomic Sciences, 3*(2), 153-162.

Klatt, C. G., Inskeep, W. P., Herrgard, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., . . . Bryant, D. A. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Frontiers in Microbiology, 4*, 106.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1), e1.

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America, 102*(7), 2567-2572.

Kuhad, R. C., Singh, A., & Eriksson, K. E. (1997). Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Advances in Biochemical Engineering/Biotechnology, 57*, 45-125.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., . . . Knight, R. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology, 31*, 814-821.

Lee, J., Lee, C., Hugunin, K., Maute, C., & Dysko, R. (2010). Bacteria from drinking water supply and their fate in gastrointestinal tracts of germ-free mice: A phylogenetic comparison study. *Water Research, 44*(17), 5050-5058.

Lee, L. S., Goh, K. M., Chan, C. S., Annie Tan, G. Y., Yin, W. F., Chong, C. S., & Chan, K. G. (2018). Microbial diversity of thermophiles with biomass deconstruction potential in a foliage-rich hot spring. *MicrobiologyOpen*, e00615.

Lee, S., Moon, D., Choi, H. T., & Song, H. (2007). Purification and characterization of an intracellular NADH: Quinone reductase from *Trametes versicolor*. *The Journal of Microbiology, 45*(4), 333-338.

Lee, Y.-J., Jeong, H., Park, G.-S., Kwak, Y., Lee, S.-J., Lee, S. J., . . . Lee, D.-W. (2015). Genome sequence of a native-feather degrading extremely thermophilic *Eubacterium*, *Fervidobacterium islandicum* AW-1. *Standards in Genomic Sciences, 10*, 71.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels, 6*(1), 41.

Liu, H., Cheng, Y., Du, B., Tong, C., Liang, S., Han, S., . . . Lin, Y. (2015). Overexpression of a novel thermostable and chloride-tolerant laccase from *Thermus thermophilus* SG0. 5JP17-16 in *Pichia pastoris* and its application in synthetic dye decolorization. *PLoS ONE, 10*(3), e0119833.

Lončar, N., Božić, N., & Vujčić, Z. (2016). Expression and characterization of a thermostable organic solvent-tolerant laccase from *Bacillus licheniformis* ATCC 9945a. *Journal of Molecular Catalysis B: Enzymatic, 134*(Part B), 390–395.

Losey, N. A., Stevenson, B. S., Busse, H. J., Sinninghe Damsté, J. S., Rijpstra, W. I. C., Rudd, S., & Lawson, P. A. (2013). *Thermoanaerobaculum aquaticum* gen. nov., sp. nov., the first cultivated member of Acidobacteria subdivision 23, isolated from a hot spring. *International Journal of Systematic and Evolutionary Microbiology, 63*(Pt 11), 4149-4157.

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology, 71*(12), 8228-8235.

Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology, 73*(5), 1576-1585.

Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., . . . Sheehan, J. (2008). How biotech can transform biofuels. *Nature Biotechnology, 26*(2), 169-172.

Maitan-Alfenas, G. P., Visser, E. M., & Guimarães, V. M. (2015). Enzymatic hydrolysis of lignocellulosic biomass: Converting food waste in valuable products. *Current Opinion in Food Science, 1*, 44-49.

Manjula, A., Sathyavathi, S., Gunasekaran, P., & Rajendhran, J. (2011). Comparison of seven methods of DNA extraction from termitarium for functional metagenomic DNA library construction. *Journal of Scientific and Industrial Research, 70*, 945-951.

Martiny, A. C., Treseder, K., & Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal, 7*(4), 830-838.

Matsumura, H., Umezawa, K., Takeda, K., Sugimoto, N., Ishida, T., Samejima, M., . . . Nakamura, N. (2014). Discovery of a eukaryotic pyrroloquinoline quinone-dependent oxidoreductase belonging to a new auxiliary activity family in the database of carbohydrate-active enzymes. *PLoS ONE, 9*(8), e104851.

Mhuantong, W., Charoensawan, V., Kanokratana, P., Tangphatsornruang, S., & Champreda, V. (2015). Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnology for Biofuels, 8*, 16.

Mirete, S., Morgante, V., & González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Current Opinion in Biotechnology, 38*, 143-149.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., . . . Snapper, S. B. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology, 13*(9), R79.

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., . . . Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science, 332*(6032), 970-974.

Murray, M., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8*(19), 4321-4325.

Mussatto, S. I., Dragone, G., Guimarães, P. M., Silva, J. P. A., Carneiro, L. M., Roberto, I. C., . . . Teixeira, J. A. (2010). Technological trends, global market, and challenges of bio-ethanol production. *Biotechnology Advances, 28*(6), 817-830.

Nesbø, C. L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., . . . Doolittle, W. F. (2009). The genome of *Thermosipho africanus* TCF52B: Lateral genetic connections to the Firmicutes and archaea. *Journal of Bacteriology, 191*(6), 1974-1978.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial ecology and evolution: A ribosomal RNA approach. *Annual Review of Microbiology, 40*, 337-365.

Parisutham, V., Kim, T. H., & Lee, S. K. (2014). Feasibilities of consolidated bioprocessing microbes: From pretreatment to biofuel production. *Bioresource Technology, 161*, 431-440.

Park, J. I., Steen, E. J., Burd, H., Evans, S. S., Redding-Johnson, A. M., Batth, T., . . . Sale, K. L. (2012). A thermophilic ionic liquid-tolerant cellulase cocktail for the production of cellulosic biofuels. *PLoS ONE, 7*(5), e37010.

Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics, 30*(21), 3123-3124.

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., & Hedlund, B. P. (2013). Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after *in situ* lignocellulose enrichment. *PLoS ONE, 8*(3), e59927.

Peng, X., Qiao, W., Mi, S., Jia, X., Su, H., & Han, Y. (2015). Characterization of hemicellulase and cellulase from the extremely thermophilic bacterium *Caldicellulosiruptor owensensis* and their potential application for bioconversion of lignocellulosic biomass without pretreatment. *Biotechnology for Biofuels, 8*, 131.

Pessela, B. C., Torres, R., Fuentes, M., Mateo, C., Carrascosa, A. V., Vian, A., . . . Fernandez-Lafuente, R. (2004). A simple strategy for the purification of large thermophilic proteins overexpressed in mesophilic microorganisms: Application to multimeric enzymes from *Thermus* sp. strain T2 expressed in *Escherichia coli*. *Biotechnology Progress, 20*(5), 1507-1511.

Phillips, C. M., Beeson, W. T., Cate, J. H., & Marletta, M. A. (2011). Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chemical Biology, 6*(12), 1399-1406.

Podosokorskaya, O. A., Kadnikov, V. V., Gavrilov, S. N., Mardanov, A. V., Merkel, A. Y., Karnachuk, O. V., . . . Kublanov, I. V. (2013). Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environmental Microbiology, 15*(6), 1759-1771.

Qing, Q., & Wyman, C. E. (2011). Hydrolysis of different chain length xylooliogmers by cellulase and hemicellulase. *Bioresource Technology, 102*(2), 1359-1366.

Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P., & Verpoorte, R. (2011). Renewable energy from Cyanobacteria: Energy production optimization by metabolic pathway engineering. *Applied Microbiology and Biotechnology, 91*(3), 471-490.

Ragauskas, A. J., Beckham, G. T., Biddy, M. J., Chandra, R., Chen, F., Davis, M. F., . . . Keller, M. (2014). Lignin valorization: Improving lignin processing in the biorefinery. *Science, 344*(6185), 1246843.

Rakitin, A. L., Ermakova, A. Y., & Ravin, N. V. (2015). Novel endoxylanases of the moderately thermophilic polysaccharide-degrading bacterium *Melioribacter roseus*. *Journal of Microbiology and Biotechnology, 25*(9), 1476-1484.

Roh, C., & Schmid, R. D. (2013). Isolation of an organic solvent-tolerant lipolytic enzyme from uncultivated microorganism. *Applied Biochemistry and Biotechnology, 171*(7), 1750-1758.

Saini, J. K., Saini, R., & Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. *3 Biotech, 5*(4), 337-353.

Sarmiento, F., Peralta, R., & Blamey, J. M. (2015). Cold and hot extremozymes: Industrial relevance and current trends. *Frontiers in Bioengineering and Biotechnology, 3*, 148.

Satyanarayana, T., Littlechild, J., & Kawarabayasi, Y. (2013). *Thermophilic microbes in environmental and industrial biotechnology*. London: Springer Netherlands.

Schoenfeld, T. W., Murugapiran, S. K., Dodsworth, J. A., Floyd, S., Lodes, M., Mead, D. A., & Hedlund, B. P. (2013). Lateral gene transfer of family A DNA polymerases between thermophilic viruses, Aquificae, and Apicomplexa. *Molecular Biology and Evolution, 30*(7), 1653-1664.

Schwarz, W. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology, 56*(5-6), 634-649.

Segata, N., & Huttenhower, C. (2011). Toward an efficient method of identifying core genes for evolutionary and functional microbial phylogenies. *PLoS ONE, 6*(9), e24704.

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., & Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *The ISME Journal, 8*(6), 1166-1174.

Shi, H., Zhang, Y., Li, X., Huang, Y., Wang, L., Wang, Y., . . . Wang, F. (2013). A novel highly thermostable xylanase stimulated by Ca^{2+} from *Thermotoga thermarum*: Cloning, expression and characterization. *Biotechnology for Biofuels, 6*(1), 26.

Shintani, M., Ohtsubo, Y., Fukuda, K., Hosoyama, A., Ohji, S., Yamazoe, A., . . . Hatta, T. (2014). Complete genome sequence of the thermophilic polychlorinated biphenyl degrader *Geobacillus* sp. strain JF8 (NBRC 109937). *Genome Announcements, 2*(1), e01213-01213.

Sikorski, J., Tindall, B. J., Lowry, S., Lucas, S., Nolan, M., Copeland, A., . . . Han, C. (2010). Complete genome sequence of *Meiothermus silvanus* type strain (VI-$R2^T$). *Standards in Genomic Sciences, 3*(1), 37-46.

Singh, K., Reddy, B., Patel, D., Patel, A., Parmar, N., Patel, A., . . . Joshi, C. (2014). High potential source for biomass degradation enzyme discovery and environmental aspects revealed through metagenomics of Indian buffalo rumen. *BioMed Research International, 2014*.

Skinner, K. A., & Leathers, T. D. (2004). Bacterial contaminants of fuel ethanol production. *Journal of Industrial Microbiology and Biotechnology, 31*(9), 401-408.

Snel, B., Bork, P., & Huynen, M. A. (1999). Genome phylogeny based on gene content. *Nature Genetics, 21*(1), 108-110.

Somboonna, N., Assawamakin, A., Wilantho, A., Tangphatsornruang, S., & Tongsima, S. (2012). Metagenomic profiles of free-living archaea, bacteria and small eukaryotes in coastal areas of Sichang island, Thailand. *BMC Genomics, 13*(7), S29.

Suen, G., Scott, J. J., Aylward, F. O., Adams, S. M., Tringe, S. G., Pinto-Tomás, A. A., . . . Barry, K. W. (2010). An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics, 6*(9), e1001129.

Taherzadeh, M. J., & Karimi, K. (2007). Acid-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *BioResources, 2*(3), 472-499.

Takami, H., Nishi, S., Lu, J., Shimamura, S., & Takaki, Y. (2004a). Genomic characterization of thermophilic *Geobacillus* species isolated from the deepest sea mud of the Mariana Trench. *Extremophiles, 8*(5), 351-356.

Takami, H., Takaki, Y., Chee, G.-J., Nishi, S., Shimamura, S., Suzuki, H., . . . Uchiyama, I. (2004b). Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Research, 32*(21), 6292-6303.

Taylor, M. P., Eley, K. L., Martin, S., Tuffin, M. I., Burton, S. G., & Cowan, D. A. (2009). Thermophilic ethanologenesis: Future prospects for second-generation bioethanol production. *Trends in Biotechnology, 27*(7), 398-405.

Thiel, V., Tomsho, L. P., Burhans, R., Gay, S. E., Schuster, S. C., Ward, D. M., & Bryant, D. A. (2015). Draft genome sequence of the Deinococcus-Thermus bacterium *Meiothermus ruber* strain A. *Genome Announcements, 3*(2), e00202-00215.

Thiel, V., Wood, J. M., Olsen, M. T., Tank, M., Klatt, C. G., Ward, D. M., & Bryant, D. A. (2016). The dark side of the Mushroom Spring microbial mat: Life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA amplicons and metagenomic sequencing. *Frontiers in Microbiology, 7*, 919.

Turbe-Doan, A., Arfi, Y., Record, E., Estrada-Alvarado, I., & Levasseur, A. (2013). Heterologous production of cellobiose dehydrogenases from the basidiomycete *Coprinopsis cinerea* and the ascomycete *Podospora anserina* and their effect on saccharification of wheat straw. *Applied Microbiology and Biotechnology, 97*(11), 4873-4885.

Turner, P., Mamo, G., & Karlsson, E. N. (2007). Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories, 6*, 9.

Urbieta, M. S., Donati, E. R., Chan, K. G., Shahar, S., Sin, L. L., & Goh, K. M. (2015). Thermophiles in the genomic era: Biodiversity, science, and applications. *Biotechnology Advances, 33*(6), 633-647.

Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B., & Eijsink, V. G. (2005). The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *The Journal of Biological Chemistry, 280*(31), 28492-28497.

van den Brink, J., & de Vries, R. P. (2011). Fungal enzyme sets for plant polysaccharide degradation. *Applied Microbiology and Biotechnology, 91*(6), 1477-1492.

van der Lelie, D., Taghavi, S., McCorkle, S. M., Li, L.-L., Malfatti, S. A., Monteleone, D., . . . Himmel, M. E. (2012). The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS ONE, 7*(5), e36740.

Van Dyk, J., & Pletschke, B. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-factors affecting enzymes, conversion and synergy. *Biotechnology Advances, 30*(6), 1458-1480.

van Hellemond, E. W., Leferink, N. G., Heuts, D. P., Fraaije, M. W., & van Berkel, W. J. (2006). Occurrence and biocatalytic potential of carbohydrate oxidases. *Advances in Applied Microbiology, 60*, 17-54.

Van Soest, P. J., & Wine, R. H. (1967). Use of detergents in the analysis of fibrous feeds. IV. Determination of plant cell-wall constituents. *Journal of the Association of Official Analytical Chemists, 50*(1), 50-55.

Vishnivetskaya, T. A., Hamilton-Brehm, S. D., Podar, M., Mosher, J. J., Palumbo, A. V., Phelps, T. J., . . . Elkins, J. G. (2015). Community analysis of plant biomass-degrading microorganisms from Obsidian Pool, Yellowstone National Park. *Microbial Ecology, 69*(2), 333-345.

Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R., & Marletta, M. A. (2014). A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America, 111*(38), 13822-13827.

Wagner, I. D., & Wiegel, J. (2008). Diversity of thermophilic anaerobes. *Annals of the New York Academy of Sciences, 1125*, 1-43.

Walton, J., Banerjee, G., & Car, S. (2011). GENPLAT: An automated platform for biomass enzyme discovery and cocktail optimization. *Journal of Visualized Experiments,* (56), e3314.

Wang, S. D., Guo, G. S., Li, L., Cao, L. C., Tong, L., Ren, G. H., & Liu, Y. H. (2014). Identification and characterization of an unusual glycosyltransferase-like enzyme with β-galactosidase activity from a soil metagenomic library. *Enzyme and Microbial Technology, 57*, 26-35.

Whittaker, J. W. (2003). Free radical catalysis by galactose oxidase. *Chemical Reviews, 103*(6), 2347-2363.

Whittaker, M. M., Kersten, P. J., Nakamura, N., Sanders-Loehr, J., Schweizer, E. S., & Whittaker, J. W. (1996). Glyoxal oxidase from *Phanerochaete chrysosporium* is a new radical-copper oxidase. *The Journal of Biological Chemistry, 271*(2), 681-687.

Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W. R., & Liesegang, H. (2013). Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announcements, 1*(2), e00092-00013.

Wilson, D. B. (2009). Cellulases and biofuels. *Current Opinion in Biotechnology, 20*(3), 295-299.

Wilson, D. B. (2011). Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology, 14*(3), 259-263.

Wissuwa, J., Stokke, R., Fedøy, A.-E., Lian, K., Smalås, A. O., & Steen, I. H. (2016). Isolation and complete genome sequence of the thermophilic *Geobacillus* sp. 12AMOR1 from an Arctic deep-sea hydrothermal vent site. *Standards in Genomic Sciences, 11*, 16.

Wu, Y. W., Joshua, C., Eichorst, S. A., Gladden, J. M., Simmons, B. A., & Singer, S. W. (2015). Genomic analysis of xylose metabolism in members of the Deinoccocus-Thermus phylum from thermophilic biomass-deconstructing bacterial consortia. *Bioenergy Research, 8*(3), 1031-1038.

Xia, Y., Cai, L., Zhang, T., & Fang, H. H. (2012). Effects of substrate loading and co-substrates on thermophilic anaerobic conversion of microcrystalline cellulose and microbial communities revealed using high-throughput sequencing. *International Journal of Hydrogen Energy, 37*(18), 13652-13659.

Xia, Y., Ju, F., Fang, H. H., & Zhang, T. (2013). Mining of novel thermo-stable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics. *PLoS ONE, 8*(1), e53779.

Xia, Y., Wang, Y., Fang, H. H., Jin, T., Zhong, H., & Zhang, T. (2014). Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Scientific Reports, 4*, 6708.

Xing, M. N., Zhang, X. Z., & Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnology Advances, 30*(4), 920-929.

Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., . . . Latreille, P. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biology, 5*(7), e156.

Yang, B., & Wyman, C. E. (2008). Pretreatment: The key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining, 2*(1), 26-40.

Yao, N., Ren, Y., & Wang, W. (2013). Genome sequence of a thermophilic *Bacillus*, *Geobacillus thermodenitrificans* DSM465. *Genome Announcements, 1*(6), e01046-01013.

Yu, C., Reddy, A. P., Simmons, C. W., Simmons, B. A., Singer, S. W., & VanderGheynst, J. S. (2015). Preservation of microbial communities enriched on lignocellulose under thermophilic and high-solid conditions. *Biotechnology for Biofuels, 8*, 206.

Yu, T., Anbarasan, S., Wang, Y., Telli, K., Aslan, A. S., Su, Z., . . . Havukainen, S. (2016). Hyperthermostable *Thermotoga maritima* xylanase XYN10B shows high activity at high temperatures in the presence of biomass-dissolving hydrophilic ionic liquids. *Extremophiles, 20*(4), 515-524.

Zaneveld, J. R., Lozupone, C., Gordon, J. I., & Knight, R. (2010). Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Research, 38*(12), 3869-3879.

Zeikus, J. G., Ben-Bassat Arie, & Hegge, P. W. (1980). Microbiology of methanogenesis in thermal, volcanic environments. *Journal of Bacteriology, 143*(1), 432-440.

Zhang, Z., Donaldson, A. A., & Ma, X. (2012). Advancements and future directions in enzyme technology for biomass conversion. *Biotechnology Advances, 30*(4), 913-919.

Zhao, C., Chu, Y., Li, Y., Yang, C., Chen, Y., Wang, X., & Liu, B. (2017). High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnology Letters, 39*(1), 123-131.

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology, 62*(2), 316-322.

LIST OF PUBLICATIONS AND PAPER PRESENTED

A LIST OF PUBLICATIONS

Lee, L. S., Goh, K. M., Chan, C. S., Annie Tan, G. Y., Yin, W. F., Chong, C. S., & Chan, K. G. (2018). Microbial diversity of thermophiles with biomass deconstruction potential in a foliage-rich hot spring. *MicrobiologyOpen*, e00615.

Chan, C.S., Sin, L.L., Chan, K.-G., Shamsir, M.S., Manan, F.A., Sani, R.K., Goh, K.M. 2016. Characterization of a glucose-tolerant β-glucosidase from *Anoxybacillus* sp. DT3-1. *Biotechnology for Biofuels*, 9, 174.

Urbieta, M.S., Donati, E.R., Chan, K.-G., Shahar, S., Sin, L.L., Goh, K.M. 2015. Thermophiles in the genomic era: Biodiversity, science, and applications. *Biotechnology Advances*, *33*(6), 633-647.

B LIST OF PRESENTATION

Characterization of a glucose-tolerant β-glucosidase from *Anoxybacillus* sp. DT3-1, Monash Science Symposium 2016, 21 November 2016 to 23 November 2016, Monash University (International).