BIOPROSPECTING BACTERIAL DIVERSITY FOR THERMOSTABLE HYDROLYTIC ENZYMES AND CHARACTERIZATION OF A THERMOSTABLE GH9 CELLULASE FROM SABAH HOT SPRING

BAK ZAIBAH BINTI FAZAL

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITY MALAYSIA SABAH
2023

BIOPROSPECTING BACTERIAL DIVERSITY FOR THERMOSTABLE HYDROLYTIC ENZYMES AND CHARACTERIZATION OF A THERMOSTABLE GH9 CELLULASE FROM SABAH HOT SPRING

BAK ZAIBAH BINTI FAZAL

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

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITY MALAYSIA SABAH
2023

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL : **BIOPROSPECTING BACTERIAL DIVERSITY FOR THERMOSTABLE HYDROLYTIC ENZYMES AND CHARACTERIZATION OF A THERMOSTABLE GH9 CELLULASE FROM SABAH HOT SPRING**

IJAZAH : **SARJANA SAINS**

BIDANG : **BIOTEKNOLOGI**

Saya **BAK ZAIBAH BINTI FAZAL**, Sesi **2019-2023**, mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

☐	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
☐	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
/	TIDAK TERHAD	

Disahkan Oleh,

ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

BAK ZAIBAH BINTI FAZAL
MZ1911006T

(Tandatangan Pustakawan)

Tarikh : Oktober 2023

(Prof. Madya Dr. Cahyo Budiman)
Penyelia Utama

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged.

8 August 2023

Bak Zaibah binti Fazal
MZ1911006T

CERTIFICATION

NAME	:	**BAK ZAIBAH BINTI FAZAL**
MATRIC NUM.	:	**MZ191106T**
TITTLE	:	**BIOPROSPECTING BACTERIAL DIVERSITY FOR THERMOSTABLE HYDROLYTIC ENZYMES AND CHARACTERIZATION OF A THERMOSTABLE GH9 CELLULASE FROM SABAH HOT SPRING**
DEGREE	:	**MASTER OF SCIENCE**
FIELD	:	**BIOTECHNOLOGY**
VIVA DATE	:	**8 AUGUST 2023**

CERTIFIED BY;

Signature

1. MAIN SUPERVISOR
Prof. Madya Dr. Cahyo Budiman

2. CO-SUPERVISOR
Prof. Madya Ts. Dr. Zarina Amin

3. CO-SUPERVISOR
Prof. Dr. Clemente Michael Vui Ling Wong

ACKNOWLEDGMENT

Alhamdulillah, all praise and thanks are due to Allah (SWT) for His infinite mercy and blessings that have enabled me to successfully complete this research study. I am grateful to Him for granting me the strength, perseverance, and determination to overcome the challenges and obstacles that I encountered along the way.

I would like to take this opportunity to express my deepest gratitude to my supervisors Assoc. Prof. Dr. Cahyo Budiman, Assoc. Prof. Ts Dr. Zarina Amin and Prof. Dr. Clemente Michael Wong Vui Ling for their invaluable guidance, patience, and encouragement throughout my academic journey. They have been an exceptional mentor, providing me with unwavering support, constructive feedback and expertise in shaping my research and academic study.

I would also like to thank the people and organizations that contributed for the completion of this study. This includes my project member Nurshafrina Aida Yahya, for her invaluable assistance and dedication for this research project, the Minister of Higher Education (MOHE) for the research grant (FRGS0489-2019) funding this project and Sabah Biodiversity Centre (SaBC) for access to Sabah resources.

Finally, I would like to express my deepest appreciation to all the individuals who have played a crucial role in my academic journey, including my protein folding group members, lab mates, pasca mates and family. Their tireless support has been the pillars of my success, and I am forever grateful for their presence in my life.

Bak Zaibah binti Fazal
8 August 2023

ABSTRACT

Hydrolytic enzymes are often required to catalyse reactions under harsh conditions, including high-temperature environments. Consequently, thermostable hydrolytic enzymes hold significant industrial importance due to their stability in high temperature working conditions. The Poring hot spring (PHS) in Sabah shows promise as a source for prospecting thermophilic microorganisms and their hydrolytic enzymes. However, the bacterial diversity and availability of thermostable hydrolytic enzymes at this site have not been explored to date. Therefore, this study aims to determine the bacterial diversity of Sabah hot springs and their hydrolytic enzyme production. To achieve this goal, water and sediment samples were collected from the site for physicochemical analysis, genome extraction and hydrolytic enzyme screening. During sampling, the recorded temperature ranged from 42 °C to 59 °C and the pH ranged from 6.5 to 7.0. The V3-V4 region of the 16S rRNA gene was amplified using the total genomic DNA (gDNA) from both water and sediment samples as templates. Subsequently, the amplicons were sequenced using the Illumina paired-end platform and subjected to analysis. The results showed that 99 % of the 1,127 OTUs obtained from water samples were bacteria, followed by 0.3 % archaea and 0.01 % unclassified microorganisms. In sediment samples, the majority of 1,559 OTUs were classified as 93 % bacteria, 0.03 % archaea and 7 % unclassified microorganisms. The most abundant phylum in water samples was Proteobacteria (84 %), while Cyanobacteria (49 %) dominated in sediment samples. Furthermore, hydrolytic enzyme screening, targeting viable bacteria yielded positive colonies producing proteases (11 colonies), amylases (9), lipases (8) and cellulases (2). As cellulases were considered the most relevant hydrolytic enzymes for Malaysia due to advanced oil palm plantations, the best colony exhibiting thermostable cellulases, was identified as *Thermoflavifilum aggregans* SP1 and selected for further studies. This strain was found to possess a gene encoding cellulase under glycoside hydrolase family 9 (GH9), which is less studied compared to other GH families. Accordingly, the GH9 cellulase of the SP1 strain (SP1-GH9) was chosen for further studies. For this purpose, an expression system for this gene was constructed in pET28a(+) for overexpression in *Escherichia coli* BL21(DE3). The protein was successfully expressed with an apparent size of 66 kDa and could be purified using Ni^{2+}-NTA affinity chromatography coupled with size exclusion chromatography (SEC). Interestingly, under SEC, this protein eluted at an apparent size of 154.2 kDa, indicating that it forms a dimeric structure. Far-UV circular dichroism (CD) spectroscopy confirmed that the pure protein was correctly folded and exhibited a dominant helical structure, consistent with the structural model of this protein, characterized by an $(\alpha\text{-}\alpha)_6$-barrel structural motif. The purified protein exhibited a specific activity of 0.53 x 10^{-2} U/mg against a *p*-Nitrophenyl β-D-Cellobioside (*p*NPC) substrate, with optimum activity observed at 60 °C and pH 4.0. Interestingly, the activity of SP1-GH9 was found to be enhanced in the presence of various metal ions (Zn, Mn, Ca, Li, Mg, K, Na), EDTA, Tween 20 and Triton X-100. This study highlights the potential development of biocatalysts with multifunctional thermostable glycoside hydrolases from local hot springs for various industrial applications.

ABSTRAK

BIOPROSPEK KEPELBAGAIAN BAKTERIA UNTUK ENZIM HIDROLITIK TERMOSTABIL DAN PENCIRIAN SELULASE GH9 TERMOSTABIL DARI MATA AIR PANAS SABAH

Enzim hidrolitik sering diperlukan untuk memangkin tindak balas dalam keadaan yang mencabar, termasuk persekitaran suhu tinggi. Oleh itu, enzim hidrolitik thermostabil mempunyai kepentingan industri yang besar kerana kestabilan mereka dalam keadaan kerja suhu tinggi. Sumber air panas Poring (PHS) di Sabah menunjukkan potensi sebagai sumber untuk prospek mikroorganisma termofilik dan enzim hidrolitik mereka. Walau bagaimanapun, kepelbagaian bakteria dan ketersediaan enzim hidrolitik thermostabil di tapak ini belum pernah dikaji hingga kini. Oleh itu, kajian ini bertujuan untuk menentukan kepelbagaian bakteria di mata air panas Sabah serta penghasilan enzim hidrolitik mereka. Bagi mencapai matlamat ini, sampel air dan sedimen telah dikumpulkan dari tapak tersebut untuk analisis fizikokimia, pengekstrakan genom dan pemeriksaan enzim hidrolitik. Semasa mengambil sampel, suhu dicatatkan dalam kadar 42 °C hingga 59 °C dan pH dicatat pada kadar 6.5 hingga 7.0. Kawasan V3-V4 pada gen 16S rRNA diamplifikasi daripada semua jumlah genomik DNA (gDNA) daripada kedua-dua sampel air dan sedimen sebagai templat. Kemudian, amplikon tersebut dijujuk menggunakan platform berpasangan Illumina dan dianalisis. Hasil kajian menunjukkan bahawa 99 % daripada 1,127 OTUs yang diperoleh daripada sampel air adalah bakteria, diikuti 0.3 % arkea dan 0.01 % mikroorganisma yang tidak dikelaskan. Dalam sampel sedimen, majoriti daripada 1,559 OTUs dikelaskan sebagai 93 % bakteria, 0.03 % arkea dan 7 % adalah tidak dikelaskan. Filum yang paling banyak terdapat dalam sampel air adalah Proteobacteria (84 %), sementara *Cyanobacteria* (49 %) *mendominasi dalam sampel sedimen. Selanjutnya, saringan enzim hidrolitik dengan mensasarkan bakteria memberi beberapa koloni positif menghasilkan protease (11 koloni), amilase (9), lipase (8) dan selulase (2). Oleh kerana selulase dianggap sebagai enzim hidrolitik yang paling relevan untuk Malaysia kerana pengusahaan ladang kelapa sawit yang maju, koloni terbaik mempamerkan selulase termostable, yang dikenal pasti sebagai Thermoflavifilum aggregans SP1, telah dipilih untuk kajian lanjut. Strain ini didapati mempunyai pengekodan gen selulase di bawah keluarga hidrolase glukosida 9 (GH9), yang kurang dikaji berbanding keluarga GH yang lain. Oleh itu, selulase GH9 daripada strain SP1 (SP1-GH9) telah dipilih untuk kajian lanjut. Bagi tujuan ini, sistem ekspresi gen ini telah dibina dalam pET28a(+) untuk diexpresi dalam Escherichia coli BL21(DE3). Protein ini berjaya diekspresikan dengan saiz 66 kDa dan dapat ditulenkan menggunakan kromatografi afiniti Ni^{2+}-NTA ditambah dengan kromatografi pengecualian saiz (SEC). Menariknya, di bawah SEC, protein ini diperhatikan pada saiz 154.2 kDa menunjukkan ia membentuk struktur dimerik. Spektroskopi dichroism pekeliling UV jauh (CD) mengesahkan bahawa protein dilipat dengan betul dan mempunyai struktur utama berbentuk helikal, selari dengan model struktur protein ini, yang dicirikan oleh motif struktur $(\alpha\text{-}\alpha)_6$-'barrel'. Protein yang ditulenkan menunjukkan aktiviti khas sebanyak 0.53 x 10^{-2} U/mg terhadap substrat p-Nitrophenyl β-D-Cellobioside (pNPC), dengan aktiviti optimum diperhatikan pada 60 °C dan pH 4.0. Menariknya, aktiviti SP1-GH9 didapati meningkat dengan kehadiran pelbagai ion logam (Zn, Mn, Ca, Li, Mg, K, Na), EDTA, Tween 20 dan Triton X-100. Kajian ini menunjukkan potensi pembangunan*

biopemangkin dengan enzim hidrolase glikosil termostabil pelbagai fungsi daripada mata air panas tempatan untuk pelbagai aplikasi industri.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Å	–	Angstrom
α	–	Alpha
β	–	Beta
λ	–	Lambda
ºC	–	Degree Celsius
%	–	Percentage
µL	–	Microliter
µmol	–	Micromole
µg	–	Microgram
bp	–	Base pair
K_M	–	Michaelis constant
k_{cat}	–	Turnover number
kb	–	Kilobase pair
kDa	–	Kilo Dalton
Mbp	–	Mega base pair
mM	–	Millimolar
mL	–	Millilitre
U	–	Unit
V_{MAX}	–	Maximal velocity

LIST OF ABBREVIATIONS

bp	–	Base pair
CBM	–	Carbohydrate Binding Module
CD	–	Circular Dichroism
CMC	–	Carboxymethyl Cellulose
CI	–	Cellulolytic Index
DSC	–	Differential Scanning Calorimetry
E. coli	–	*Escherichia coli*
EDTA	–	Ethylenediaminetetraacetic Acid
gDNA	–	Genomic Deoxyribonucleic Acid
GF	–	Gil Filtration
GMQE	–	Global Model Quality Estimation
GH	–	Glycoside Hydrolase
IMAC	–	Immobilized Metal Affinity Chromatography
IPTG	–	Isopropyl β-D-1-thiogalactopyranoside
LB	–	Luria-Bertani
Ni-NTA	–	Nickle-Nitrilotriacetic Acid
OD	–	Optical density
PDB	–	Protein Data Bank
PHS	–	Poring Hot Spring
***p*NPC**	–	*p*-Nitrophenyl β-D-Cellobioside
R.M.S.D	–	Root-mean-square deviation
rpm	–	Revolution per minute
rRNA	–	Ribosome Ribonucleic Acid
SEC	–	Size Exclusion Chromatography
SDS-PAGE	–	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
TEMED	–	Tetramethyl ethylenediamine
UV	–	Ultraviolet ray

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Hydrolytic enzymes are essential for various biological processes, including digestion, food processing and waste treatment, as well as for industrial applications like biofuel production, detergent manufacturing, paper and pulp processing (Chapman et al., 2018). The demand for these enzymes is on the rise due to the increasing need for sustainable and eco-friendly methods across industries. A valuable source of hydrolytic enzymes is hot springs, which are known for their high temperatures and unique microbial communities (Choure et al., 2021; Sahay et al., 2017). These extreme conditions like high temperature, while challenging for most organisms, support the growth of microbes capable of producing enzymes that break down complex compounds. Moreover, hot springs foster the evolution of novel enzymes with exceptional properties, making them ideal for high-temperature industrial applications. Hot spring-derived enzymes have proven to be thermally stable and active, reducing the reliance on chemical additives and promoting sustainability (Sahay et al., 2017).

Apart from bacterial isolation, previous studies have focused on characterizing the enzymes produced by thermostable bacteria (Knapik et al., 2019; Alrumman et al., 2018). This research extends to exploring enzyme structure and function, as well as optimizing production conditions, resulting in various commercial applications, including biofuel production and other industrial processes. The bioprospecting of hot springs for thermostable bacteria producing hydrolytic enzymes remains an active research area aimed at discovering improved enzymes for diverse industrial applications.

Poring hot spring in Ranau, Sabah, Malaysia, is one such hot spring with the potential to be a promising source for isolating thermostable bacteria producing hydrolytic enzymes. This hot spring has a temperature range of approximately 40-70 °C makes it an ideal environment for thermophilic bacteria (Hua, 2016) Additionally, the rich and diverse environment created by the high level of dissolved minerals and gases in hot spring waters provides fertile ground for bacterial growth (Yu et al., 2022; Gu et al., 2017). Although previous studies have identified thermophilic bacteria in Malaysian hot springs, including amylolytic, lipolytic, cellulolytic and proteolytic bacteria, none have specifically explored Poring hot spring (Abidin et al., 2021; Msarah et al., 2020; Msarah et al., 2018; Chan et al., 2015). Consequently, further investigation of Poring hot spring may reveal novel thermostable bacteria with hydrolytic enzyme-producing capabilities.

Bioprospecting of environmental sources often coupled with the study of its bacterial diversity (Vuong et al., 2022; Adegboye et al., 2021). Bacterial diversity and bioprospecting are intertwined, as the former refers to the variation in the number and types of bacteria in a given environment, and the latter aims to discover biologically active compounds and organisms for industrial and commercial purposes (Mutalib et al., 2020). Bacteria are one of the most diverse groups of organisms and are found in a wide range of environments, including hot springs, soil and water. Exploring these environments opens doors to discovering new bacteria and their associated metabolic processes, leading to the identification of biologically active compounds, including enzymes and secondary metabolites (Adegboye et al., 2021). Furthermore, the diversity of bacterial communities can influence the potential for bioprospecting success, as different bacterial communities may produce different metabolic processes and biologically active compounds. Mutalib et al. (2020) suggested that a higher diversity often correlates with a greater likelihood of finding beneficial compounds for mankind. For example, hot springs, with their high diversity and unique microorganisms, have been fruitful sources of thermophilic bacteria producing hydrolytic enzymes. Therefore, bacterial diversity plays a pivotal role in successful bioprospecting by providing a pool of potential bacteria and associated metabolic processes for industrial and commercial purposes.

Studies across various countries, including Malaysia, have demonstrated that hot springs are rich reservoirs of bacterial diversity, housing unique and rare bacterial strains. Studies of hot springs in peninsular Malaysia, have revealed a significant presence of thermophilic bacteria, including genera such as *Geobacillus, Thermosipho* and *Meiothermus* (Abidin et al., 2021; Lee et al., 2018) known for their biotechnological applications due to their enzyme-producing abilities. Similarly, studies from Indian hot springs, have unveiled diverse bacterial populations, including thermophilic and halophilic bacteria, with potential biotechnological uses (Chikkanna et al., 2018; Ghati & Paul, 2015). These findings highlight the importance of hot springs as sources of novel and valuable bacterial strains.

Bharti and Grimm (2021) emphasized the effectiveness of using the 16S metagenome for investigating bacterial diversity, providing insights into the dominant bacterial taxa within a community. This metagenomic approach involves sequencing the 16S ribosomal RNA (rRNA) gene, which is a commonly used target in metagenomic studies. It has been applied in studies of hot springs conducted in various countries, such as Malaysia, India, Africa, and the United States (Podar et al., 2020; Chaudhuri et al., 2017; Chan et al., 2015; Tekere et al., 2011). These studies have revealed commonly found phyla in hot springs, including Firmicutes, Proteobacteria, Actinobacteria, Thermotogae and Cyanobacteria (Chen et al., 2022; Sahoo et al., 2017). Additionally, bacterial genera commonly identified in hot springs include *Thermus, Bacillus, Geobacillus, Legionella, Sulfurihydrogenibium* and *Deinococcus*. Importantly, bacterial profiles can vary based on the specific conditions and location of the hot spring. Despite the promise of Poring hot spring for bioprospecting hydrolytic enzymes, research on the bacterial diversity at this site and the presence of thermostable hydrolytic enzymes remain limited.

In a previous study, Yahya (2023) successfully isolated a thermostable bacterium, *Thermoflavifilum aggregans* SP1, indicating the potential of the location as a source of thermostable bacteria. Further whole genome analysis revealed that this SP1 strain harbors genes encoding cellulose-degrading enzymes (cellulases) belonging to glycoside hydrolase (GH) families 9 (GH9) and 5 (GH5). Ravachol et al. (2014) classified cellulases into 17 families, with family 9 and family 48 displaying greater specificity for cellulases (EC 3.2.1.4, EC 3.2.1.91, and EC

3.2.1.176). Out of 1,525 GH9 sequences that have been analyzed and identified, 132 were designated as cellulases, showing greater activity on cellodextrin, except for one enzyme from *Photobacterium profundum*, which exhibited higher activity on chito-oligosaccharides than on cellodextrins (Honda et al., 2011). Many GH9 cellulases also display side activities on related polysaccharides such as 1,3–1,4-glycans (Arai et al., 2003), xylan (Eckert et al., 2002), or xyloglucan (Hirano et al., 2013). However, their preferred substrate is either soluble (carboxymethyl cellulose and cellodextrins) or insoluble (amorphous) cellulose. Structurally, GH9 cellulases often feature additional domains aside from the catalytic domain. Many known GH9 enzymes contain ancillary modules/domains, such as carbohydrate-binding modules (CBM) (Ravachol et al., 2014; Najmudin et al., 2006) or modules with yet unknown functions. Therefore, the study of GH9 cellulases of the SP1 strain is of interest. Notably, the gene encoding GH9 of cellulase in this strain was only detected from the genome sequence and has never been tested for functionality. Further investigation into the functionality of this enzyme is essential to confirm its significance and potential applications. In summary, the study on bacterial diversity and insight into GH9 cellulase from SP1 strain of Poring hot spring aims to unveil the uniqueness of this cellulase enzyme and highlight the potential of hot springs as a site for bioprospecting hydrolytic enzymes, particularly cellulases.

1.2 Problem Statements

In Malaysia, local production of hydrolytic enzymes is limited, leading to a reliance on imported enzymes. This situation highlights the need to explore local resources for identifying potential sources of hydrolytic enzymes. Additionally, the demand for thermostable enzymes across various industries remains high, but there is a lack of availability of such enzymes from local sources. This underscores the necessity for further exploration of local resources to identify potential sources of thermostable enzymes. There are various hot springs in Sabah, Malaysia, which could potentially serve as sources of thermostable bacteria and enzymes. However, limited knowledge exists regarding the richness of these sources in terms of the types of bacteria and hydrolytic enzymes they can produce. Therefore, there is a pressing

need for further exploration of the hot springs in Sabah to determine their potential for producing thermostable enzymes.

Among the many hydrolytic enzymes needed, there is a high demand for cellulose-degrading enzymes due to the growth of the oil palm-related industry. Oil palm production generates a significant amount of waste that can be converted into valuable products through the use of cellulose-degrading enzymes. While a bacterial strain known as *Thermoflavifilum aggregans* SP1 has been successfully isolated from Poring hot spring in Sabah and encodes a gene for cellulase enzyme under the GH9 family (GH9-SP1), this gene has never been functionally confirmed. Consequently, studying this enzyme is crucial to confirm its activity and assess its potential for use in the production of valuable products from oil palm waste. This study holds particular significance as it can serve as a proof-of-concept for Sabah's hot spring bioprospecting potential and provide insights into the types of enzymes that could potentially be produced from hot springs in Sabah and their industrial applications.

1.3 Research Questions

The study aims to address the following questions:

a) What is the taxonomic composition and diversity of the bacterial community in the Poring hot spring, and how does it compare to other hot springs?
b) What hydrolytic enzymes are produced by the cultivable thermostable bacteria found in Poring hot spring?
c) What are the properties of the cellulase gene isolated from the thermophilic bacterial strain found in Poring hot spring?