SCREENING, CLONING AND EXPRESSION OF A XYLANASE FROM PALM OIL MILL EFFLUENT (POME) BY FUNCTIONAL METAGENOMICS APPROACH

BY

ADIBAH BINTI PARMAN

A thesis submitted in fulfilment of the requirement for the degree of Master of Science (Biotechnology Engineering)

Kulliyyah of Engineering
International Islamic University Malaysia

SEPTEMBER 2020

ABSTRACT

Metagenomics approach is an alternative method to study the novel enzyme. Therefore, a metagenomic fosmid library of approximately 50,000 clones was screened to identify the novel xylanase enzymes. The metagenomics deoxyribonucleic acid (DNA) used in this study is from the samples of palm oil mill effluent (POME) from Felda Palm Industry Sdn. Bhd. in Mempaga, Pahang. Clones of the metagenomics were screened using fluorescence substrate of chlorocoumarin xylobioside. The high score reads obtained from screening were sent for next-generation sequencing (NGS) of Illumina HiSeq2000. The sequences were further analysed to identify the predicted xylanase genes using automated and manual bioinformatic tools. A total of 34 predicted xylanases were identified, and five predicted xylanase genes #11, #15, #16, #17, and #18 of various microbial origins were chosen to be further analysed. The translated sequences of these five genes later were analysed to determine the primary, secondary, and tertiary protein structures in predicting feature and function of predicted xylanases. Next, based on the integrity checking in agarose gel electrophoresis (GE), Gene #15 with approximately 1.2 kb was chosen to be cloned into the pBAD-TOPO vector. Gene #15 has the percentage identity of 99.3% to the *Ochrobactrum intermedium* with glycoside hydrolase (GH) 10 as the conserved domain. After cloning, pBAD-TOPO-Xyl was used to transform *E. coli* cells and expressed using the inducer, L-arabinose. Protein #15 (P15) was later purified using immobilised metal affinity chromatography (IMAC), and the molecular mass of SDS-PAGE of approximately 46 kDa was confirmed. The P15 also fluoresced when checked with chlorocoumarin xylobioside substrate, which suggests the protein is a xylanase enzyme.

خلاصة البحث

يعتبر نهج الميتاجينوميات (دراسة المادة الوراثية) طريقة بديلة ومبتكرة لدراسة الإنزيمات. لذلك تم فحص مكتبة فوزميد للمادة الجينية (ميتاجينومية) لحوالي 50.000 استنساخ لتحديد إنزيمات الزيلانيز الجديدة. وتم استخلاص الحمض النووي الريبوزي منزوع الأكسجين من عينات من المرتجع السائل لمعاصر زيت النخيل (POME) من شركة فِلدا للتنمية المستدامة في صناعة النخيل بمدينة ممفاجا في ولاية باهانج بماليزيا. وتم فحص استنساخ الجينات الوراثية (الميتاجينومية) باستخدام مادة الكلوروكيومارين المشعة (Chloro-coumarin xylo-bioside). وتم إرسال القراءات العالية التي تم الحصول عليها من الفحص لدراسة التسلسل الجيني للجيل التالي (NGS) من Illumina HiSeq2000. وتم تحليل المتواليات بشكل أكبر لتحديد جينات انزيم الزيلانيز المتوقعة باستخدام أدوات المعلوماتية الحيوية الآلية واليدوية. وتم تحديد ما مجموعه 34 جين متوقع لإنزيم الزيلانيز، وكذلك تم اختيار خمسة جينات زيلانيز متوقعة . وأيضاً تم اختيار #11 و #15 و #16 و #17 و #18 جينات من الأصول الميكروبية المختلفة لمزيد من الفحص والتحليل. وتم تحليل المتواليات المترجمة لهذه الجينات الخمسة في وقت لاحق لتحديد تراكيب البروتين الأولية والثانوية والثالثية في التنبؤ بسمات ووظائف إنزيمات الزيلانيز المتوقعة. وبعد ذلك، واستنادًا إلى فحص السلامة في الرحلان الكهربائي للهلام (GE) ، تم اختيار الجين رقم #15 مع حوالي 1.2 كيلوبايت لاستنساخه في ناقل pBAD-TOPO. ويمتلك الجين رقم #15 نسبة مئوية 99.3٪ (*Ochrobactrum intermedium*) مع جليكوسيد الهيدروليز 10 (GH) كمجال محفوظ. وبعد الاستنساخ، تم استخدام pBAD-TOPO-Xyl لتحويل خلايا *E. coli* وتم التعبير عنه باستخدام المحرض، L-arabinose. وتم تنقية البروتين رقم #15 (P15) لاحقًا باستخدام كروماتوغرافيا تقارب المعدن المثبت (IMAC) ، وتم تأكيد الكتلة الجزيئية لـ SDS-PAGE التي تبلغ 46 كيلو دالتون تقريبًا. واتضح أن البروتين P15 أيضًا مشع عند فحصه مع مادة الكلوروكيومارين (Chloro-coumarin xylo-bioside) كركيزة للتفاعل (substrate) ، مما يشير ويؤكد أن البروتين هو إنزيم زيلانيز.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology Engineering)

……
Hamzah Mohd. Salleh
Supervisor

……
Ibrahim Ali Noorbatcha
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology Engineering)

……
Raha Ahmad Raus
Internal Examiner

……
Farah Diba Abu Bakar
External Examiner

This thesis was submitted to the Department of Biotechnology Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology Engineering)

……
Nor Fadhillah Mohamed Azmin
Head, Department of
Biotechnology Engineering

This thesis was submitted to the Kulliyyah of Engineering. and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology Engineering)

……
Sany Izan Ihsan
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Adibah Parman

Signature ..... Date .....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

SCREENING, CLONING AND EXPRESSION OF A XYLANASE FROM PALM OIL MILL EFFLUENT (POME) BY FUNCTIONAL METAGENOMICS APPROACH

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.

1. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

2. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Adibah Parman

……..…….. ……..
Signature Date

ACKNOWLEDGEMENTS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيم

In the name of Allah, the Most Beneficent, the Most Merciful. All praises belong to Allah, the supreme power, who is the right guider of humankind. Firstly, a special thanks to both of my supervisors, Prof. Dr. Hamzah Mohd. Salleh and Prof. Dr. Ibrahim Ali Noorbatcha for their continuous guide, teaching, support, encouragement, and advice. I will be forever grateful. *Jazakumullahu khairan kathiran*. May Allah always repay both of you with goodness. I also would like to thank Prof. S.G. Withers (University of British Columbia, Canada) for kindly providing chlorocoumarin xylobioside needed for this project.

I wish to express my appreciation and thanks to those who provided their time, effort, and support for this project, especially my colleague, sister Farah. She's my partner in research who has a lot of experience in molecular technique and the one that always shared and understood the hardship in finishing this vast research. Both of us always sit together, discussing and wondering when we will be able to finish this research. To the members under the same supervisory committee, brother Oualid and brother Aziz, thank you very much for guiding me in the skills of laboratory work. I am very grateful to have another three Ph.D. students who are in the supervisory committee that can always guide me in the upcoming problem of this research.

Finally, it is my highest pleasure to dedicate this work to my dear parents, Rohenah Binti Hasan and Parman Bin Sirad. My utmost appreciation to both, especially when I always in need of motivational advice and support in continuing my research. They are indeed the backbone and supporter of my work either in terms of financial or emotional advice. Not forget to my siblings, who always have a firm belief in my ability to accomplish this goal, thank you for all of your support and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4MU	4-methylumbelliferone
APS	Ammonium Persulfate
BAC	Bacterial Artificial Chromosomes
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like Alignment Tool
BOD	Biochemical Oxygen Demand
CAZy	Carbohydrate-Active enZYmes
CCX	Chlorocoumarin xylobioside
CDS	Coding Sequence
CMC	Carboxymethyl cellulose
COGs	Clusters of Orthologous Groups
CUI	Command User Interface
DH_2O	Distilled water
DNA	Deoxyribonucleic Acid
DOE-JGI	Department of Energy Joint Genome Institute
DTT	Dithiothreitol
EBI	European Bioinformatics Institute
EC	Enzyme Commission Number
EFB	Empty Fruit Brunches
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archi
FFB	Fresh Fruit Bunches
FPLC	Fast Protein Liquid Chromatography
FR	Forward
GE	Gel Electrophoresis
GH	Glycoside Hydrolase
GMQE	Global Model Quality Estimation
GO	Gene Ontology
GOLD	Genomes Online Database
GRAVY	Grand Average of Hydropathicity
GUI	Graphical User Interface
HTS	High-Throughput Screening
IMAC	Immobilized Metal Affinity Chromatography
IMG/M	Integrated Microbial Genomes with Microbiomes
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
LB	Luria-Bertani medium
M	Marker
MAD	Median Absolute Deviation
MAP	Metagenome Annotation Pipeline
Med	Median
MGI	Malaysia Genome Institute
MG-RAST	Metagenome Rapid Annotation using Subsystem Technology
NGS	Next-Generation Sequencing
OD	Optical Density

OPF	Oil Palm Fronds
OPT	Oil Palm Trunks
PCR	Polymerase Chain Reaction
Pfam	Protein Family Database
PHA	Polyhydroxyalkanoates
PKC	Palm Kernel Cake
PKS	Palm Kernel Shell
POME	Palm Oil Mill Effluent
PPF	Palm Pressed Fibres
PRODIGAL	PROkaryotic DynamIc programming Gene finding Algorithm
QMEAN Z-Score	Qualitative Model Energy Analysis of Z-Score
RBB	Remazol Brilliant Blue
RFU	Relative Fluorescence Unit
RNA	Ribonucleic Acid
RV	Reverse
Rz	Robust z-score
S.O.B	Super Optimal Broth
S.O.C	Super Optimal Broth with Catabolite repression
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SOPMA	Self-Optimized Prediction Method with Alignment
SSPACE	SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension
TE	Tris-EDTA buffer
TEMED	Tetramethylethylenediamine
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XOs	Xylooligosaccharides

LIST OF SYMBOLS

kb	kilo basepairs
ºC	celcius
µg	microgram
ml	millilitre
µl	microlitre
rpm	revolutions per minute
g	gram
g (RCF)	relative centrifugal force
bp	base pairs
s	second
mM	millimolar
kDa	kilo dalton

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

The advancement in the genomics and metagenomics field has widened the view on microbial diversity and benefits future potential biotechnological applications, including the novel enzyme mining process. Metagenomics is an advanced alternative approach to genomics in understanding the microorganism's diversity in a given sample without necessarily culturing any microbes. Cultivable microorganisms constitute only a tiny fraction of microbial diversity, which limits the mining process of novel enzymes (Ferrer et al., 2005). A lot of microorganisms cannot be cultured normally in the laboratory set-up. The metagenomics approach is the key to getting access to the uncultivable microbes that potentially possess unique enzymes.

The metagenomics approach will be used in this research to study potential xylan-degrading enzymes from the entire microorganisms living in palm oil mill effluent (POME) samples. POME is a by-product of processed fresh fruit bunches (FFB) from the palm oil industry, which contains high nutrient concentrations like carbohydrate, nitrogen, protein, phosphorus, potassium, magnesium, and calcium (Madaki & Seng, 2013).

Xylanase is one of the xylan degrading enzymes contained in hemicellulose rich POME. Xylanase is currently accessible in different industrial applications because of its catalytic ability in bio-bleaching of paper pulp, improvement of animal feed, bread making, application in solid waste treatment, preparation of juice from fruits or

vegetables, improve retting of flax fibers, production of biofuels, and others (Asish, 2015). Throughout this research, genes encoding xylanase enzymes are expected to be found from the POME sample through the metagenomics method.

The methodologies to be employed will be through the construction of the metagenomics library, fluorescence high throughput screening (HTS), next-generation sequencing (NGS), homology modeling, cloning, expression, and purification of the recombinant xylanase enzyme.

1.2 PROBLEM STATEMENT

To date, most of the available enzymes used in industrial applications originated from microbes. Enzymes are recognized as a favorable catalyst which can accelerate the rate of reaction of a process. The traditional method used in getting microbial enzymes of interest is by cultivating a microorganism through standard laboratory techniques. However, research has proven that only less than 1% of environmental microorganisms can be cultivated (Uchiyama & Miyazaki, 2009). This situation indicates that more than 99% of bacteria from the environment cannot be cultured using conventional approaches.

As a counter method, the metagenomics approach appeared to be an alternative to conventional microbial screening in studying the diversity of another 99% of non-cultured microorganisms (Kennedy et al., 2011). Therefore, this research will focus on the functional metagenomics approach to search for xylanases from the entire microorganisms in palm oil mill effluent (POME) without culturing the microorganisms from the POME sample. There are already microorganisms reported producing xylanases in POME, like *Bacillus*, *Micrococcus*, and *Staphylococcus* (Soleimaninanadegani & Manshad, 2014). However, by investigating through the

metagenomics approach, xylanase enzymes will be screened from the total microorganisms in POME, which involve 99% of still undiscovered microorganisms and including from 1% of cultivable microorganisms. The bacterial species of *Bacillus, Micrococcus,* and *Staphylococcus* can be considered as microorganisms from the 1% that can be investigated through standard laboratory methods. On the contrary, the expected xylanase to be found in this research will be more diverse because of direct mining from the POME sample.

1.3 RESEARCH OBJECTIVES

1. To screen for xylanase from a metagenomics library of palm oil mill effluent (POME) sample.
2. To identify unique DNA sequences encoding xylanases and model the structure of xylanase from POME metagenome.
3. To clone, express, and purify a xylanase obtained from POME metagenomic library.

1.4 SIGNIFICANCE OF STUDY

Xylanase is a major xylan degrading enzyme for lignocellulosic materials. It plays a huge role in the paper and pulp bleaching industry during the past several years besides having potential applications in bioconversion of lignocellulosic biomass and agro-wastes into a fermentative product, the digestibility of animal feedstocks, and the clarification of juice (Motta et al., 2013).

1.5 RESEARCH METHODOLOGY

This functional metagenomics approach begins with metagenomic DNA extraction from palm oil mill effluent (POME) samples. Next, the metagenomic DNA was cloned into fosmid with a size of approximately 40 kb to construct a metagenomic library. In mining the xylanases enzyme, high throughput screening using chlorocoumarin xylobioside substrate was used in this study. Next, the gene of interest was further identified through Illumina HiSeq2000 next-generation sequencing (NGS) method and was analysed using metagenomics bioinformatics analysis. Several potential xylanases that have been found were further cloned and expressed using the pBAD-TOPO expression system. The recombinant enzyme was then finally purified using immobilised metal affinity chromatography (IMAC)

1.6 SCOPE OF RESEARCH

This study was conducted to study the possible xylanase enzymes in the microorganisms of palm oil mill effluent using the metagenomics approach. It was only limited to the lab-scale production of recombinant xylanases using molecular cloning method and a specific substrate. The library was constructed by inserting the fragmented metagenomic DNA of a specific size into the fosmid vector. Besides that, the fluorescence high-throughput screening (HTS), next-generation sequencing (NGS), and bioinformatics analysis method are also used to find the genes of interest. Later, the molecular method of cloning, expression, and purification were also done to further investigate on this enzyme.

1.7 DISSERTATION ORGANISATION

The thesis is systematically organised in 5 chapters with chapter 1 discussing an overview, background, significance, and objective of the study. Next, Chapter 2 is focusing on literature review and information related to the functional metagenomics approach, xylanases, and palm oil mill effluent, especially in the five recent years. Following this chapter, chapter 3 elaborates on the materials and methodology being used in this study. Accordingly, chapter 4 details the results obtained in this study with a clear and concise discussion. Finally, chapter 5 is the concluding part of the thesis.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

Screening for microbial biocatalysts directly from environmental samples is more applicable and convenient compared to the conventional plate cultivation approach (Hu et al., 2008). The metagenomic technique has an advantage over conventional plate cultivation approach and provides an alternative approach in understanding 99% missing biodiversity of unculturable or difficult to culture microbes (Ferrés et al., 2015). In this study, the xylanase gene obtained from palm oil mill effluent (POME) through a functional metagenomic approach was screened, identified, analysed, cloned, expressed, and finally purified.

Xylanase is one of the xylan degrading enzymes for hemicellulosic materials metabolism which is currently popular in different industries because of its catalytic ability in bio-bleaching of paper pulp, improvement of animal feed, bread making, application in solid waste treatment, preparation of juice from fruits or vegetables, improve retting of flax fibres, production of biofuels, and others (Asish, 2015). However, this research will be limited to screening, cloning, and expression of a xylanase enzyme found from the POME sample only.

2.2 AGRO-RESIDUE BIOMASS

Malaysia is endowed with massive biomass supply from the agricultural and plantation residues. One of the major plantation residue in Malaysia besides rubber, cocoa, and pepper is the palm oil (Shafie et al., 2012). Malaysia was the largest producer and exporter of palm oil in the world from until 2007, when Indonesia replaced Malaysia as the largest palm oil producer. Nevertheless, Malaysia is still among the top palm oil producers and is only second to Indonesia (Otieno et al., 2016).

Agro-industrial wastes in the form of lignocellulosic biomass are accumulated every year in huge quantities (Mussatto & Teixeira, 2010). The wastes of palm oil industry include empty fruit bunches (EFB), palm pressed fibres (PPF), palm kernel cake (PKC) and palm kernel shell (PKS). Other wastes from the palm oil industry which contain lignocellulosic materials are oil palm trunks (OPT), oil palm fronds (OPF) and palm oil mill effluent (POME) (Abdullah & Sulaiman, 2013).

2.2.1 Palm Oil Mill Effluent (POME)

POME is the final stage effluent of palm oil mill production. Fresh POME is a hot (80-90 °C) and thick brownish colloidal mixture of oil, water, and fine suspended solids. In this condition, POME is acidic at a pH of 4.5 with very high non-toxic biochemical oxygen demand (BOD) (Madaki & Seng, 2013). Raw POME contains high concentrations of carbohydrate (lignocellulosic materials), protein, nitrogenous compound, lipids and minerals which are summarised in Table 2.1 (Habib et al., 1997). The nutrient rich POME makes it as an ideal habitat for microorganisms.

Table 2.1: The approximate compositions (%) of major constituents in POME (Habib et al., 1997)

Major Constituents	Composition (%)
Moisture	6.99
Crude Protein	12.75
Crude lipid	10.21
Ash	14.88
Carbohydrate	29.55
Nitrogen-free extract	26.39
Total Carotene	0.019
Total	100.789

2.2.2 POME as Microbial Fermentation Medium

POME has been used as microbial fermentation medium to produce several value added products like antibiotics, bio insecticides, solvents, polyhydroxyalkanoates (PHA), organic acids as well as enzymes (Wu et al., 2009). Microorganisms are known as a good source of useful enzymes as they could produce in extremely high rates and able to synthesise active biological product (Motta et al., 2013). Some of the bacteria found in POME is highlighted in Table 2.2. The indigenous microorganisms in POME or any habitat can be isolated and identified in laboratory conditions but the great majority (~99%) of the microbial population cannot be cultured, isolated and identified in laboratory conditions (Ferrés et al., 2015). As a counter method, metagenomics approach (also known as non-cultivable method) is a solution in making accessible of 99% biodiversity of microorganisms.

Table 2.2: Microorganisms found in POME by previous research

References	Microorganisms	Genera
(Soleimaninanadegani & Manshad, 2014)	Bacteria	*Bacillus* sp.
		Micrococcus sp.
		Pseudomonas sp.
		Staphylococcus sp.
	Fungi	*Aspergillus* sp.
		Candida sp.
		Fusarium sp.
		Mucor sp.
		Penicillium sp.
(Nwuche et al., 2014)	Bacteria	*Flavobacterium* sp.
	Fungi	*Trichoderma* sp.
(Bala et al., 2018)	Bacteria	*Stenotrophomonas* sp.
		Providencia sp.
		Klebsiella sp.
	Fungi	*Meyerozyma* sp.

2.3 XYLAN FROM HEMICELLULOSE

The agricultural biomass residue rich in lignocellulosic biomass materials mentioned previously in section 2.1 basically consists of 10 to 25% of lignin, 35 to 50% of cellulose and 20 to 35% of hemicellulose (Saha, 2003).

Hemicellulose consists of linear and branched polymer which basically made up of five and six different sugars of xylose, mannose, glucose, galactose and arabinose with other components of acetic, ferulic and glucuronic acid. It is being classified into xylans, glucans, mannans, glucuronoxylans, glucomannans, arabinoxylans,