Biochemical characterization of β-xylosidase and β-glucosidase isolated from a thermophilic horse manure metagenomic library

By

Kanyisa Ndata

A thesis submitted in fulfillment of the requirements for the degree of

MAGISTER SCIENTIAE (M. Sc)

Institute for Microbial Biotechnology and Metagenomics

Department of Biotechnology, University of the Western Cape,

Bellville, Cape Town

South Africa

Supervisor: Prof. Marla Trindade

Co-supervisor: Dr. Lonnie van Zyl

Declaration

I, Kanyisa Ndata declare that the thesis titled "***Biochemical characterization of β-xylosidase and β-glucosidase isolated from a thermophilic horse manure metagenomic library***" is my work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Signed…….

Date……..

Acknowledgments

First and foremost, I would like to thank God for with Him all things are possible.

I would like to thank my supervisor Prof. Marla Trindade for granting me an opportunity to be part of IMBM, funding, and for patience with me and my writing skills. I appreciate all the suggestions and the reading of my thesis.

I would also like to express a great deal of appreciation to my co-supervisor Dr. Lonnie van Zyl for always keeping his office door open, valued contributions, and help in my laboratory experiments and writing. I will forever be in debt for all the help.

Special thanks to Dr. Anita Burger for all the support, encouragement, and help throughout my time in IMBM.

IMBMers for the chats, the check-ups, and keeping me sane throughout the project.

Last but not least my family, (Phumela, Mkhuseli, Nikiwe, and Sinovuyo) for being home keeping me laughing in hard times, and all my friends for the random calls and check-ups and moral support and prayers.

Dedication

To my late grandmother

Abstract

The complete degradation of recalcitrant lignocellulose biomass into value-added products requires the efficient and synergistic action of lignocellulose degrading enzymes. This has resulted in a need for the discovery of new hydrolytic enzymes which are more effective than commonly used ones. β-xylosidases and β-glucosidases are key glycoside hydrolases (GHs) that catalyse the final hydrolytic steps of xylan and cellulose degradation, essential for the complete degradation of lignocellulose. Functional-based metagenomics has been employed successfully for the identification and discovery of novel GH genes from a metagenome library. Therefore, this approach was used in this study to increase the chances of discovering novel glycoside hydrolase genes from a horse manure metagenomic DNA library constructed in a previous study. Three fosmid clones P55E4, P81G1, and P89A4 exhibiting β-xylosidase activity were found to encode putative glycosyl hydrolases designated XylP55, XylP81, and BglP89. Amino acid sequence analysis revealed that XylP55, XylP81, and BglP89 are members of the GH43, GH39, and GH3 glycoside hydrolase families, respectively. Phylogenetic analysis of XylP81 and BglP89 indicated that these showed relatively low sequence similarities to other homologues in the respective GH families. The enzymes were expressed and purified, and only XylP81 and BglP89 were biochemically characterized. XylP81 (~58 kDa) and BglP89 (~84 kDa) both showed optimum activity at pH 6 and 50°C and retained 100% residual activity at 55°C after 1-hour indicating that they are moderately thermostable. XylP81 had high specific activity against 4-nitrophenyl-β-D-xylopyranoside (*p*NPX; 122 U/mg) with a K_M value of 5.3 mM, k_{cat}/K_M of 20.3 $s^{-1}mM^{-1}$, and it showed enzyme activity against α-L-arabinofuranosidase, β-galactosidase, and β-glucosidase activity. BglP89 had a high specific activity for 4-nitrophenyl-β-D-glucopyranoside (*p*NPG; 133.5 U/mg) with a K_M value of 8.4 mM, k_{cat}/K_M of 22 $s^{-1}mM^{-1}$ and also showed α-L-arabinofuranosidase, β-galactosidase, β-glucosidase, and low β-xylosidase activity. BglP89 also showed low hydrolytic activity on cellobiose, β-glucan, and lichenan indicating that it is a broad specificity β-glucosidase. XylP81 retained ~40% activity in the presence of 3 M xylose whilst BglP89 showed considerable glucose tolerance at 150 mM glucose and retained ~46% residual activity. This study reveals two metagenomic derived enzymes (β-xylosidase and β-glucosidase)

showing characteristics that could make them potential candidates for lignocellulose biomass degradation in biotechnological and industrial applications.

Keywords: lignocellulose, β-xylosidase, β-glucosidase, metagenome, function-based screening, compost.

Table of contents

List of Figures

List of Table

Abbreviations

℃	Degrees Celsius
Amp^R	Ampicillin Resistance
APS	Ammonium Persulfate
BACs	Bacterial artificial chromosomes
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
$CaCl_2$	Calcium Chloride
CAZy	Carbohydrate Active Enzymes
CMC	Carboxymethyl cellulose
DNS	Dinitrosalicylic
EC	Enzyme Commission number
EDTA	Ethylenediamine tetraacetic acid
g	Gram
x *g*	Gravitational force
GH	Glycoside hydrolase
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilo bases
kDa	Kilo daltons
LB	Lysogeny Broth
MCS	Multiple Cloning Site
$MgCl_2$	Magnesium chloride
mg	Milligram
mL	Milliliters
mM	Millimolar
M	Molar
min	Minute
NaCl	Sodium chloride
NCBI	National Centre of Biotechnological Information
NGS	Next-generation sequencing
nm	Nanometer

OD	Optical Density
ORF	Open reading frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
*p*NP	para-Nitrophenol
*p*NPG	4-nitrophenyl-β-D-glucopyranoside
*p*NPX	4-nitrophenyl-β-D-xylopyranoside
SDS	Sodium Dodecyl Sulphate
TAE	Tris-base Acetic acid EDTA
TEMED	Tetramethylethylenediamine
TLC	Thin Layer Chromatography
UV	Ultraviolet
v/v	Volume/volume percentage
Vmax	Maximum velocity
w/v	Weight/volume percentage
µg	Micrograms
µL	Microlitres
µM	Micromolar

Chapter 1: Literature Review

1.1 Introduction

Lignocellulose is an abundant, inexpensive, and renewable biomass and is typically found in large amounts as waste material from forestry, agricultural, municipal, and agro-industrial processes (Pérez *et al.*, 2012; Isikgor and Becer, 2015). The continuous accumulation of lignocellulosic biomass on Earth is creating a great opportunity for it to be explored for biotechnological and industrial applications which include, biofuels, pharmaceuticals, pulp and paper, as well as food and cosmetics (Kucharska *et al*., 2018; Isikgor and Becer, 2015). The lignocellulose biomass is mainly made up of cellulose, hemicellulose, lignin, and small amounts of pectin and proteins (**Figure 1**). The lignocellulose polymers are strongly linked to one another by covalent, non-covalent, and hydrogen bonds making it rigid and resistant to enzymatic degradation (Usui *et al*., 1999).

The recalcitrant nature of the plant cell wall has evolved to protect the plant against the microbial attack which hinders its degradation into simple fermentable sugars for industrial and biotechnological applications (Usui *et al*., 1999; Ji *et al*., 2012). Enzymatic degradation (biocatalysis) is an environmentally friendly, sustainable, and viable approach used for the degradation of lignocellulose biomass (Wang *et al*., 2019; Li *et al*., 2014). Therefore, discovering glycoside hydrolase (GH) enzymes with high catalytic efficiency, high substrate specificity, and affinity, which can also withstand harsh industrial conditions and are resistant to product feedback inhibition is required. There are GH enzymes currently reported and commercially available but there is also a growing need for more novel enzymes with improved catalytic performance or functional properties to meet industrial requirements and lower the cost of enzyme production (Colombo *et al*., 2016).

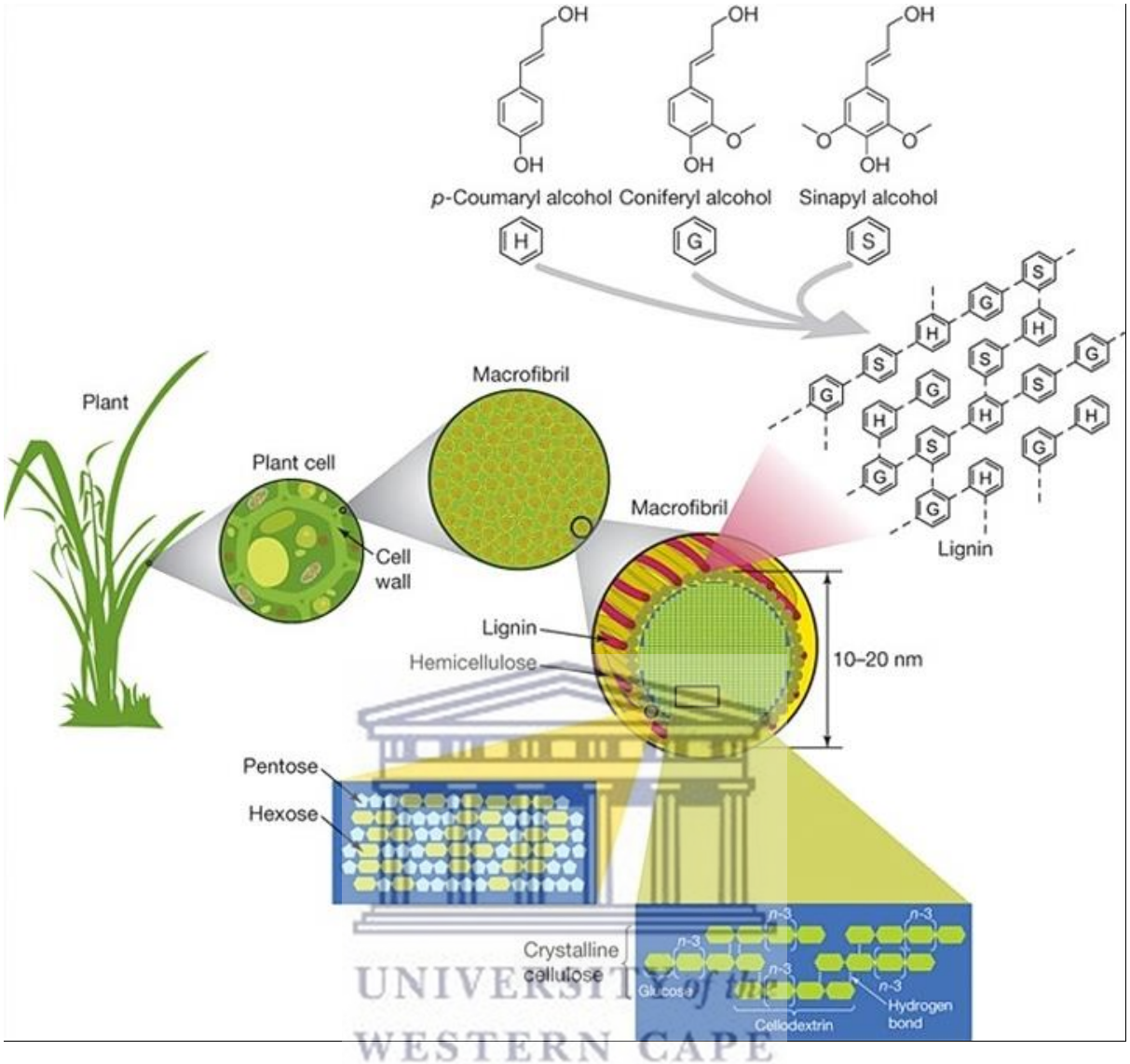

Figure 1: Schematic representation of lignocellulosic biomass showing the three polymers (cellulose, hemicellulose, and lignin) in the plant cell wall (Adapted from Rubin 2008).

In nature, complex lignocellulosic biomass is synergistically degraded by environmental microorganisms (archaea, bacteria, and fungi) which then produce a diversity of lignocellulose degrading GH enzymes (Yang *et al*., 2011). These enzymes could improve the efficiency of lignocellulose degradation in various industrial processes. However, the majority of these microorganisms cannot be cultured using traditional culturing methods in laboratories (Linares-Pastén *et al.*, 2014), however researchers are now tapping into unculturable isolates using a metagenomic approach. The metagenomic approach is a bioprospecting tool enabling the discovery of novel lignocellulose-degrading enzymes directly from microbial

environmental samples. Several industrially relevant enzymes have been successfully discovered through metagenomic analysis of various environmental samples which include soil, compost, rumen content, hot-springs, etc (Datta *et al*., 2020). The discovery and screening of novel GH enzymes with new and improved activities from environmental DNA are achieved by using the function-based and sequence-based screening approaches.

1.2 Lignocellulose biomass composition

Lignocellulose is the major component of the plant cell wall and mainly consists of polymers which include cellulose (approximately 40-50%), hemicellulose (25-35%), and lignin (10-25%) depending on the plant source (**Table 1**) (Zoghlami and Paës, 2019). The fermentable sugars that make up the cellulose and hemicellulose have many industrial and biotechnological applications hence enzymatic degradation is essential. The cellulose and hemicellulose polymers are embedded in the hydrophobic lignin matrix and this interaction strongly limits the access to these polymers to enzymatic degradation.

Table 1: The percentage composition of cellulose, hemicellulose, and lignin from different plant sources.

Biomass source	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Barley straw	33.8	21.9	13.8
Corn cobs	45.0	39.0	15
Cotton residues	58.5	14.4	21.5
Rice residues	36.2	19.0	9.99
Sugar cane	40.0	27.0	10
Wheat straw	32.9	24.0	8.9
Switch grass	31.0	20.4	17.6

1.2.1 Cellulose

Cellulose is the most abundant component of plant cell walls and accounts for about 40-50% of lignocellulose biomass (Liu *et al*., 2012; Kousar *et al*., 2013). It is a linear polymer made up of thousands of D-glucose subunits that are linked together by β-1,4 glycosidic linkages forming cellulose chains. The linear cellulose chains are then held together by hydrogen bonds, hydrophobic interactions, and van der Waals forces to form microfibrils (Tandon 2015; Strakowska *et al*., 2018). Microfibrils can be less structured (amorphous cellulose) and easily degraded or highly structured (crystalline cellulose) which is rigid and less susceptible to degradation (Ahmed 2013). These cellulose microfibrils are connected to the hemicellulose and lignin making the lignocellulose plant cell wall structure recalcitrant and thus providing it with its rigid nature. This complex and packed structure of cellulose chains make the backbone-less accessible to hydrolysis and limits or slows down degradation of cellulose (Sørensen *et al*., 2013)

1.2.2 Hemicellulose

Hemicellulose is the second most abundant polymer making up about 35% of plant biomass and is tightly bound to the surface of cellulose microfibrils (Maitan-Alfenas *et al*., 2016). Unlike cellulose, hemicellulose is a heteropolymer composed of five different sugar monomers which may include D-xylose, L-arabinose, D-galactose, D-mannose, and D-glucose, (Kucharska *et al*., 2018). There are different types of hemicellulose polymers which include xylans, glucomannans, xyloglucans, arabinoxylans, galactomannans, arabinogalactans, and mannans depending on the sugar in the backbone and plant source. Xylan is the most abundant hemicellulose and the second most abundant polymer after cellulose (Adelsberger *et al*., 2004; Maitan-Alfenas *et al*., 2016).

1.2.2.1 Xylan

Xylan is a highly branched heteropolysaccharide made up of β-1,4- linked xylose sugars forming the main chain with substitute side chains such as L-arabinofuranosyl, glucuronyl, 4-*O*-methylglucuronyl, and acetyl groups (Collins *et al*., 2005). These xylan side chains can either be acetylated or esterified by phenolic residues such as ferulic acid or *p*-coumaric acid (Adelsberger *et al*., 2004; Wongratpanya *et al*., 2016). The type and nature of side-chains also depend on the source and the type of xylan (Ratnadewi *et al*., 2013). Xylans are classified into different types which include homoxylans, arabinoxylans, glucuronoxylans, and arabinoglucuroxylans (**Table 2**). This arrangement makes the structure of hemicellulose complex, but it is more susceptible to degradation compared to cellulose (Tandon 2015).

Table 2: Composition of different types of xylans from different xylan sources (Valenzuela *et al*., 2016; Dodd and Cann, 2009).

Type of xylan	Backbone	Side group(s)	Source
Homoxylans	β-1,4 or β-1,3-xylose		Seaweeds
Arabinoxylans	β-1,4-xylose	α-1,2- or α-1,3-arabinofuranosyl	Cereals, wheat, barley
Glucuronoxylans	β-1,4-xylose	α-1,2 or α-1,3-acetyl with α-1,2-*O*-methyl-α-D-glucuronic acid (MeGA)	Hardwood
Arabinoglucuronoxylans	β-1,4-xylose	α-1,2 or α-1,3-Ara, with α-1,2-*O*-methyl-α-D-glucuronic acid (MeGA)	Softwood (grasses)

1.2.3 Lignin

Lignin is a highly cross-linked aromatic polymer and is made up of three poly-phenolic propane units namely coniferyl, coumarily, and sinapyl alcohol, and they are linked together by ether linkages and carbon-carbon interactions. The composition and content of lignin varies depending on the plant origin. Lignin is the most recalcitrant component of plant cell walls and it links cellulose and hemicellulose polymers together (Thapa *et al.*, 2020). The main function of lignin is to bring rigidity and protection to the plant cell wall whilst hindering the chemical and enzymatic degradation of plant biomass as it limits the accessibility to cellulose and hemicellulose (Thornbury *et al.*, 2018). The removal of lignin is crucial to allow efficient degradation of cellulose and hemicellulose to produce fermentable sugars (Kumar and Chandra, 2020). Enzymes which include lignin peroxidase (EC 1.11.1.4), manganese peroxidase (EC 1.11.1.13), and laccase (EC 1.10.3.2) functionally work together to hydrolyse lignin (Fisher and Fong 2014).

1.3 Enzymatic degradation of cellulose and hemicellulose

The enzymatic degradation of lignocellulosic biomass has gained major interest in research and various industrial applications which include pulp and paper, agriculture, textile, biofuels, feeds, food and beverage, brewing, detergent, pharmaceuticals, and bioconversion to value-added products (Valášková and Baldrian 2006; Kuhad *et al.*, 2011; Ahmed 2013; Park *et al.*, 2018). The degradation of cellulose and hemicellulose polymers is carried out by cellulases and hemicellulases classified into different GH families in the Carbohydrate Active Enzymes database (CAZy, http://www.cazy.org). Glycoside hydrolases (GHs; EC 3.2.1.x) are responsible for the hydrolysis of glycosidic bonds between carbohydrates or oligosaccharides (Walker *et al*., 2017; DeCastro *et al*., 2016). The GH enzymes are widely distributed amongst prokaryotes, eukaryotes, and archaea with a variety of functions (Sathya and Khan 2014). They include glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate-binding molecules (CBMs), carbohydrate esterases (CEs), and auxiliary activity enzymes (AAs) which are also classified in the CAZy database based on their catalytic activities.

In the continuously updated CAZy database, glycoside hydrolase enzymes are grouped into different GH families based on their amino acid sequence similarities (http://www.cazy.org/GlycosideHydolase; Ahmed 2013; Cheng *et al*., 2017). This classification was first developed by (Henrissat 1991) and to date, there are 167 GH families described. Enzymes in the same family can have diverse or broad substrate specificities due to small changes in the amino acid sequences. GH enzymes are further grouped into 18 clans (GHA-GHR; **Table 3**) with GH-A being the largest clan. This clan currently contains 26 GH families which display a conserved $(\beta/\ \alpha)_8$-barrel structural catalytic fold, they also share a retaining mechanism for the hydrolysis of glycosidic bonds (Bhalla *et al*., 2014). Enzymes in a clan have a conserved protein fold or tertiary structure, catalytic amino acid residues, and catalytic mechanism suggesting that they evolved from the same ancestor (Park *et al*., 2018).

Table 3: GH enzymes classification into different clans in the CAZy database (http://www.cazy.org/Glycoside-Hydrolases.html).

Clan	Catalytic domain fold/ structure	Catalytic mechanism	GH families
GH-A	$(\beta/\alpha)_8$	Retaining	1, 2, 5, 10 ,17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128, 147, 148, 157, 158, 167
GH-B	β-jelly roll	Retaining	7, 16
GH-C	β-jelly roll	Retaining	11, 12
GH-D	$(\beta/\alpha)_8$	Retaining	27, 31, 36
GH-E	6-fold β-propeller	Retaining	33, 34, 83, 93
GH-F	5-fold β-propeller	Inverting	43, 62
GH-G	$(\alpha/\alpha)_6$	Inverting	37, 63, 100, 125
GH-H	$(\beta/\alpha)_8$	Inverting	13, 70, 77
GH-I	α+β	Inverting	24, 80
GH-J	5-fold β-propeller	Retaining	32, 68
GH-K	$(\beta/\alpha)_8$	Retaining	18, 20, 85
GH-L	$(\alpha/\alpha)_6$	Inverting	15, 65
GH-M	$(\alpha/\alpha)_6$	Inverting	8, 48
GH-N	β-helix	Inverting	28, 49
GH-O	$(\alpha/\alpha)_6$	Retaining	52, 116
GH-P	$(\alpha/\alpha)_6$	Retaining	127, 146
GH-Q	$(\alpha/\alpha)_6$	Inverting	94, 149, 161
GH-R	$(\beta/\alpha)_8$	Retaining	29, 107

1.3.1 GHs mode of action and mechanism

Glycoside hydrolases have endo- or exo-acting hydrolysis abilities as they can hydrolyse a substrate by cleaving the internal glycosidic bonds (endo-acting enzymes) or at the ends of the substrate chain (exo-acting enzymes). They work synergistically to efficiently accelerate the degradation of lignocellulosic substrates. GHs perform hydrolysis of the glycosidic bonds via one of the two mechanisms; the inverting or retaining mechanism and the catalytic mechanism is conserved for all enzymes of a particular GH family (Vuong and Wilson 2010). The inverting mechanism uses a single displacement action and occurs in a single reaction which results in a product with an inverted anomeric configuration. The retaining mechanism uses a double displacement action and occurs via two reactions; glycosylation and deglycosylation, resulting in a product with the same structural anomeric configuration (Park *et al*., 2018). Both these mechanisms occur via the general acid catalysis with the assistance of carboxylate groups of two amino acid residues, usually aspartic and glutamic acid, which act as general acid (proton donor) and a nucleophile or base (Davies and Henrissat 1995; Vuong and Wilson 2010).

1.3.2 Cellulose degradation

The degradation of β-glycosidic bonds of cellulose to fermentable glucose requires the synergistic action of a group of three cellulases which include β-1,4-endoglucanase (EC 3.2.1.4), β-1,4 exoglucanases (cellobiohydrolase, EC 3.2.1.91), and β-1,4-glucosidase (EC 3.2.1.21). The internal β-1, 4-glycosidic bonds of the cellulose linear polymer are cleaved by endoglucanases resulting in different lengths of cello-oligosaccharides (Kaur *et al*., 2007; Valášková and Baldrian, 2006). Exoglucanases are responsible for the hydrolysis of both the reducing and non-reducing ends of cellulose to release cellobiose (Horn *et al*., 2012). B-glucosidase subsequently hydrolyses the resultant cellobiose into fermentable glucose (Valášková and Baldrian, 2006; Ko *et al*., 2013).

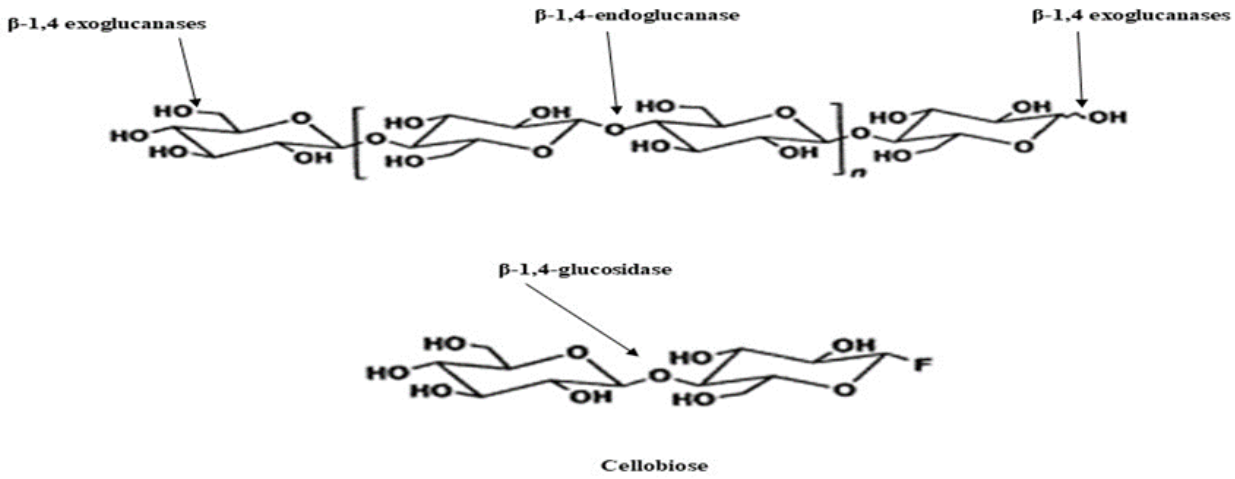

Figure 2: A schematic representation of the enzymatic degradation of a cellulose backbone and the resultant cellobiose by cellulases (Modified from Rånby 2001)

1.3.2.1 β-1,4-glucosidases

β-glucosidases are exo-acting GHs that hydrolyse the final step in cellulose degradation (Gumerov *et al.*, 2015; Li *et al*., 2014; Kaur *et al*., 2007). This step is considered rate-limiting as it prevents the accumulation of cellobiose which has end product inhibition on the activity of both endo- and exoglucanases slowing down the rate of cellulose degradation (Mohsin *et al*., 2019; Hwang *et al.*, 2018). Therefore β-glucosidases play a crucial role in ensuring efficient cellulose degradation and they enhance the production of glucose monomers. β-glucosidases catalyse the β-glycosidic bonds of the various substrates such as alkyl β-glycosides, aryl-β-glycosides, disaccharides, and short-chain oligosaccharides releasing glucose monomers. β-glucosidases have been reported and identified from fungi, bacteria, plants, animals, and metagenomes.

1.3.2.2 Classification of β-glucosidases

β-glucosidases are classified according to their substrate specificity and their amino acid sequence similarities. Based on substrate specificity they are grouped into three groups namely: aryl-β-glucosidases (specific activity against *p*-nitrophenyl-β-glucopyranoside (*p*NPG) chromogenic substrate), cellobiases (specific activity against cellobiose), and broad-spectrum β-glucosidases (activity against a wide range of substrates with different glycosidic linkages) which are the most common group of β-glucosidases (Hwang *et al*., 2018). The substrate specificity classification is not very useful or informative as a single enzyme can have activities on various substrates (Srivastava *et al*.,2019). They are therefore further classified into different GH families according to their amino acid sequence similarities. To date, β-glucosidases have been classified into nine different (GH) families which include; GH1, GH2, GH3, GH5, GH9, GH16, GH30, and GH116 on the CAZy database (http://www.cazy.org; Mohsin *et al*., 2019) with a newly discovered GH39 β-glucosidase isolated from a deep-sea bacterium (Shen *et al*., 2019). Most characterized β-glucosidases from the CAZy database are placed in the GH1 and GH3 families making them the largest β-glucosidases families (Zhang *et al*., 2017). They either perform hydrolysis in a retaining or inverting catalytic mechanism, with retaining β-glucosidases found in GH1, GH2 GH3, GH5, GH30, and GH39 families whilst the GH116 and GH9 β-glucosidases perform hydrolysis using the inverting mechanism.

1.3.2.3 Applications of β-glucosidases

β-glucosidases have a wide range of applications in the biofuel, beverage, paper and pulp, and other industries. They are used in the production of glucose sugar monomers from agricultural waste which is then fermented into biofuels (Srivastava *et al*., 2019). Most characterized β-glucosidases are inhibited by the accumulation of their end-product, glucose (Gumerov *et al*., 2015). This inhibition hinders the efficient production of biofuels hence the need to discover β-glucosidases that show high glucose-tolerance for efficient cellulose degradation and high production of glucose (Liu *et al*., 2012). The beverage industry uses β-glucosidases for clarification and removal of the bitter taste during juice extraction; while in the wine industry β-glucosidases are useful in flavour enhancement and liberation of aroma

(Escuder-Rodríguez *et al*., 2018; Gomes-Pepe *et al*., 2016). In the paper and pulp industries, β-glucosidases are employed in paper recycling as they are involved in the enzymatic de-inking of paper reducing the use of wood (Sahoo *et al*.,2020; Ahmed *et al*., 2013).

1.3.3 Xylan degradation

The complete degradation of the complex heterologous structure of xylan requires the synergistic action of several GH enzymes, collectively known as xylanases, to produce monomeric sugars. β-1,4-endoxylanases (EC 3.2.1.8) and β-1,4-xylosidases (EC 3.2.1.37) are crucial enzymes for degradation of the xylan backbone. Debranching enzymes enhance xylan degradation as they make the main chain more accessible for degradation (Bhalla *et al*., 2014; Huang *et al*., 2019. The β-1,4-endoxylanases catalyse the hydrolysis of the β-1,4-glycosidic bonds in the xylan backbone generating short xylooligosaccharides as the final product (Valenzuela *et al*., 2016). These xylooligosaccharides are then hydrolysed by the β-xylosidase to produce xylose. The side-chain degrading enzymes include the α-L-arabinofuranosidases (EC 3.2.1.55) which are responsible for the removal of α-1,2 or α-1,3 and α-1,5 arabinofuranosyl residues from the xylan main chain. The α-1, 2 linkage between glucuronic acid or 4-*O*-methyl glucuronic acid are hydrolysed by α-D-glucuronidases (EC 3.2.1.139). Acetyl xylan esterases (EC 3.1.1.6); and ferulic or coumaric acid esterases (EC 3.1.1.73 or EC 3.1.1.72) are responsible for removing the acetylated ferulic or coumaric acid groups that are usually esterified to the arabinofuranosyl group (Maitan-Alfenas *et al*., 2016; Usui 1999).

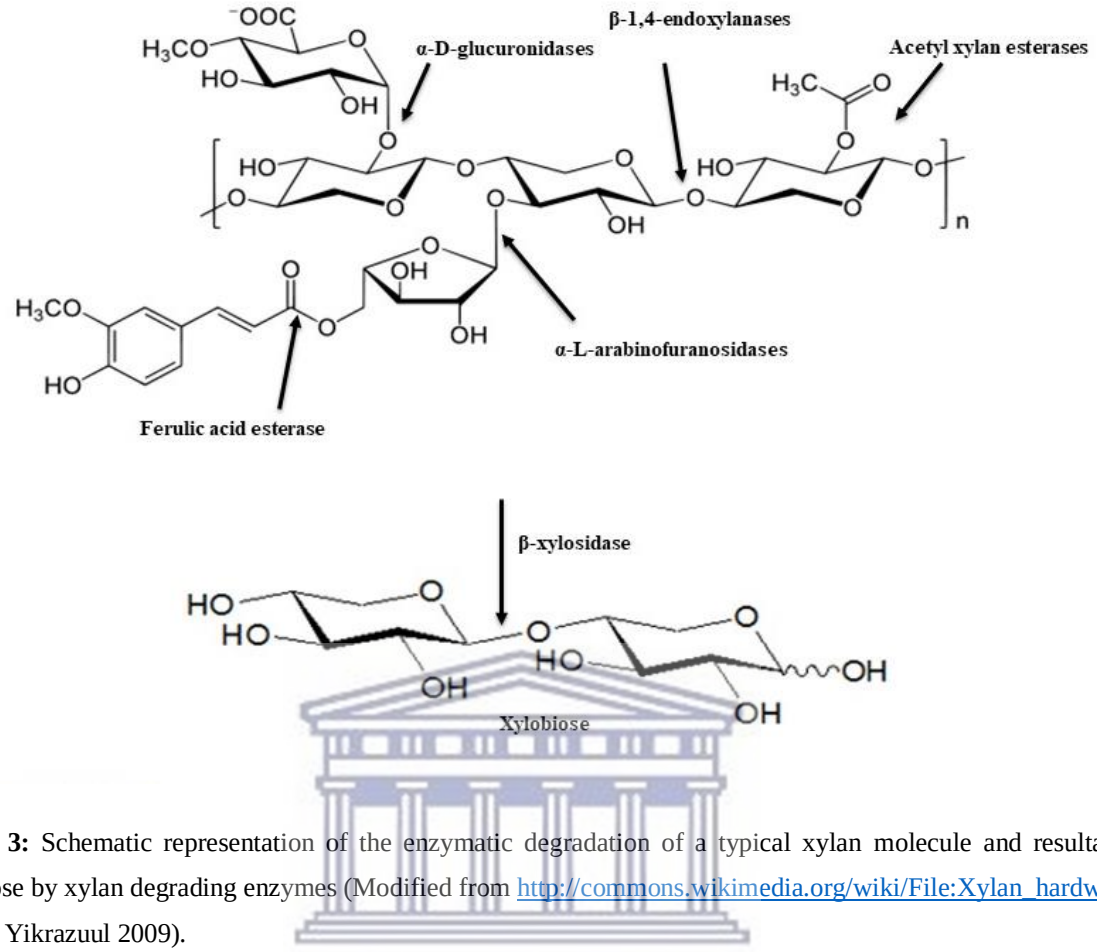

Figure 3: Schematic representation of the enzymatic degradation of a typical xylan molecule and resultant xylobiose by xylan degrading enzymes (Modified from http://commons.wikimedia.org/wiki/File:Xylan_hardwood.svg, Yikrazuul 2009).

1.3.3.1 β-1, 4-xylosidases

β-xylosidases are exo-acting GHs that hydrolyse the β-1,4-glycosidic bonds of short xylooligosaccharides from the non-reducing end releasing xylose sugar monomers (Bhalla *et al.*, 2014). This final step of xylan degradation is considered as the rate-limiting step as the enzymes act on xylooligosaccharides which are potent inhibitors of endoxylanase activity and thus alleviate inhibition and enhance xylan degradation. β-xylosidases also cleave artificial substrates like *p*-nitrophenyl-β-xylosidase (*p*NPX) and have been identified and isolated from bacteria, fungi, archaea, plants, and metagenomes (Li *et al*., 2018). β-xylosidases may either be extracellular or intracellular depending on the microorganisms and environmental conditions (Knob *et al.*, 2009).

1.3.3.2 Classification of β-xylosidases

To date, β-xylosidases are grouped into 13 different GH families according to their amino acid sequence similarities; GH1, GH3, GH5, GH30, GH39, GH43, GH51, GH52, GH54, GH116, and GH120. These perform hydrolysis using the retaining mechanisms except for GH43 β-xylosidases which use the inverting hydrolysis mechanism. GH43 also represents the family with the most characterized β-xylosidases (Rohman *et al*., 2019; Huang *et al*., 2019; Dodd and Cann 2009). β-xylosidases in family GH3 and GH43 have also been reported to have bifunctional β-xylosidase / α-L-arabinofuranosidase activities, this may be because D-xylopyranose and L-arabinofuranose sugars are spatially similar (Wagschal *et al*., 2009).

1.3.3.3 Applications of β-xylosidases

β-Xylosidases have gained great interest due to their potential application in several industrial processes and applications which include biofuel production, food industries, as well as bleaching of paper and pulp (Li *et al*., 2018). The hydrolysis of xylooligosaccharides into xylose is important in the production of biofuels as xylose is one of the sugars used for the production of bioethanol through fermentation. In the food industry, β-xylosidases are used as additives during baking for improvement of nutrition and quality of bread dough (Kumar *et al*., 2008). β-Xylosidase also hydrolyses xylosylated compounds from fruit removing the bitterness during fruit juice extraction (Jordan and Wagschal, 2010). The complete degradation of xylan is important for the improvement of the biobleaching process which is an environmentally friendly alternative to chlorine in the paper and pulp industry (Bravman *et al.*, 2001). β-xylosidases can be added to animal feed with other lignocellulosic enzymes to enhance digestibility whilst improving nutritional properties and thus speeding up weight gain in animals (Nieto-Domínguez *et al*., 2015).

1.4 Compost as the source of thermostable enzymes

Compositing is a microbial-mediated biodegradation of organic material (food waste, manure, plant biomass, crop waste, etc) into carbon dioxide, water, and compost humus material under controlled conditions. Compost contains a diversity of microorganisms that mainly include bacteria, fungi, and actinomycetes which produce enzymes that are responsible for the hydrolysis of the plant biomass (Mladenov 2018; Zang *et al*., 2018). Several enzymes including lipases, esterases, cellulases, amylases, xylanases, and proteases, amongst others, have been reported to be discovered from bioprospecting of compost. The composting process has different phases or stages controlled by a specific microbial community due to constant changes in the physicochemical parameters like temperature, oxygen, moisture content, pH, and nutrient availably (Li *et al.*, 2014).

The organic material during composting is initially degraded by mesophilic microorganisms during the mesophilic phase (15-45°C) and contains microorganisms such as *Lactobacillus* and *Acetobacter* spp (Meng *et al*., 2019; Partanen *et al*., 2010). This microbial activity causes an increase in temperature (50-70°C) leading to the thermophilic phase which is responsible for the majority of organic material decomposition, dominated by thermophilic bacteria and actinomycetes (Wang *et al*., 2019; Hubbe *et al*., 2010). Several thermophilic microorganisms including *Geobacillus stearothermophilus*, *Geobacillus toebii*, *Paenibacillus humicus*, *Rhodothermus marinus*, *Thermus thermophilus*, *Clostridium thermocellum*, *Caldicellulosiruptor saccharolyticus*, *Dictoglomus thermophilum*, *Thermoanaerobacterium* sp. and *Pseudomonas* sp. have been isolated from different types of compost (Sung *et al*., 2002; Bhalla *et al*., 2014; Sato *et al*., 2017). This makes the compost a rich source of thermostable enzymes which are in great demand and have the potential to improve different industrial and biotechnological processes.

1.4.1 Thermophiles

Thermophiles are microorganisms that strive and are found at higher temperatures. Thermophiles are classified into three groups: moderate thermophiles with stability at (50-60°C), extreme thermophiles with stability at (60-80°C), and hyperthermophiles that have stability at (80-110°C). The majority of thermophiles are found in archaea and bacteria. Thermophiles are naturally found in heated regions like the hot springs, deep sea, and hydrothermal vents, and decaying plant matter such as composts (Mohammad *et al*., 2017; Panda *et al*., 2013). Thermophiles possess high metabolism and due to their environmental conditions have adapted to high, fluctuating temperatures and harsh conditions (DeCastro *et al*., 2016). Therefore, thermophiles have attracted biotechnological interest as they are a source of efficient and thermostable enzymes that can catalyse reactions at elevated temperatures (Mehta *et al*., 2016).

1.4.2 Thermostable β-xylosidases and β-glucosidases

Thermostable enzymes are therefore able to withstand harsh conditions and show increased resistance to proteolysis and chemical denaturation agents (Sahoo *et al.*, 2020). Thermostable enzymes have added advantages over mesophilic enzymes as they have accelerated activity, increased solubility of substrates making it more accessible, high thermal stability and performing industrial processes at high temperatures reduces the risk of microbial contamination (Park *et al*., 2018; Murphy and Walsh, 2019). The thermostability of these enzymes is due to amino acid substitution, hydrophobic interactions, polar core, and interactions between subunits (Mohammad *et al*., 2017; Wang *et al*., 2016). The application of thermostable GHs in biotechnological industrial applications has increased the need for discovering more thermostable GHs with novel and improved catalytic properties.

Thermostable β-xylosidases and β-glucosidases have been isolated from microorganisms such as bacteria, archaea, and fungi in nature. The composting environments in particular have proven to be an important source of several thermostable lignocellulose degrading enzymes through culture-independent approaches (Lemos *et al*., 2017; Ellilä *et al*., 2019; Verma *et al*., 2013). A few thermostable β-xylosidases and β-glucosidases have been reported

in literature thus far and are discussed in a later chapter (Table 11). The examples include enzymes from compost metagenome libraries with optimum activities ranging between 50℃ and 75℃, and where activities are retained at temperatures as high as 90℃ incubation (Sato *et al*., 2017). The robust thermostabilities of these enzymes is also an added advantage as they are highly suitable for the harsh conditions required for efficient hydrolysis of lignocellulose to fermentable sugars. The discovery of novel β-xylosidases and β-glucosidases with enhanced and novel functionalities could be of great benefit for industrial applications.

1.5 Metagenomics in the discovery of new GH enzymes

Microbial diversity, representing the major life form on the surface of Earth, is enormous and largely unexplored (Wilkens *et al*., 2017). The majority of these microorganisms are unfortunately unculturable using traditional culturing techniques due to difficulties in mimicking their growth conditions in the laboratory, leaving a diversity of microorganisms unknown and unexplored (Dougherty *et al*., 2012). The use of metagenomic tools has opened a platform to overcome this limitation and thus study and explore unculturable microorganisms within different microbial communities (Cheng *et al*., 2017). Metagenomics is defined as the study of genomic DNA isolated directly from environmental samples omitting the culturing steps hence the culture-independent approach (Dougherty *et al*., 2012; Wierzbicka-Woś *et al*. , 2019 Montella *et al*., 2017). Metagenomic tools have been successfully used for over two decades as a powerful technology for studying the diversity of different environments and the functioning of their microbial communities (Datta *et al*., 2020). This has also led to the discovery of novel enzymes, drugs, antimicrobials, therapeutic compounds, and other natural products from uncultured and unknown species that are highly beneficial in many industrial and biotechnological applications (Liu *et al*., 2012; Colombo *et al*., 2016; Datta *et al*., 2020).

The metagenomic approach involves sampling and sample processing, followed by the extraction of DNA from environmental samples. The environmental DNA is then cloned into a suitable vector, such as cosmids, fosmids, bacterial artificial chromosomes (BACs), or plasmids depending on the DNA size to create metagenomic libraries which are subsequently transformed into a suitable host, usually *E. coli* (Armstrong *et al*., 2019; Ahmad *et al*., 2019).

The metagenomic library clones are then screened by either sequence-based or function-based methods **(Figure 4)**. Both methods have been successfully used in discovering novel lignocellulose degrading GH enzymes from various environmental samples including; cow/bovine rumen, soil, composts, hot springs, and mammalian guts (**Table 4**; Ellilä *et al*., 2019)**.** Environments where lignocellulose is naturally degraded represent an abundant and promising source of novel GH (Ahmad *et al*., 2019).

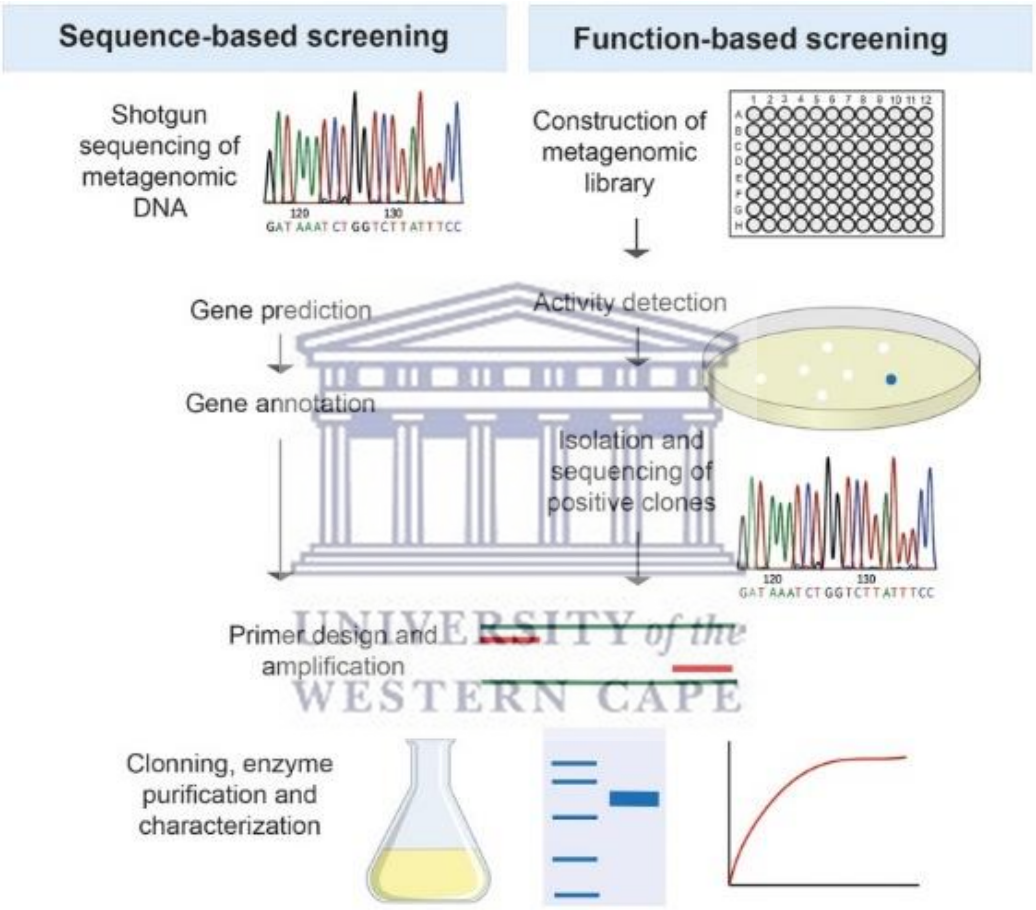

Figure 4: A representation of the screening of metagenomic DNA using the sequence-based and function-based screening approaches (DeCastro *et al*., 2016).

Table 4: Novel glycoside hydrolase enzymes discovered from various environments using metagenomic tools.

Metagenome DNA source	GH enzyme activities	Metagenomic screening approach	References
Brazilian forest soil	A novel β-glucosidase	Functional-based	Alves *et al*., 2018
Hot spring soil	A highly thermostable β-xylosidase	Sequence and function-based	Sato *et al*., 2017
Cow rumen	A multifunction β-glucosidase/ β-xylosidase/ α-arabinofuranosidase	Functional-based	Gruninger *et al*., 2014
Compost	A GH43 β-xylosidase/ α-arabinofuranosidase	Functional-based	Matsuzawa *et al*., 2015
Compost	Several GHs	Functional-based	Dougherty *et al*., 2012
Yak rumen	Two β-glucosidase/β-xylosidase	Functional-based	Bao *et al*., 2012
Yak rumen	Bifunctional xylanase/endoglucanase	Functional-based	Chang *et al*., 2011
Amazon soil	Two novel β-glucosidase	Functional-based	Bergmann *et al*., 2014
Soil	Endo-β-1,4-Glucananse	Sequence-based	Hua *et al*., 2015
Rhizosphere	Endo-β-1,4-Glucananse	Functional-based	Wierzbicka-Woś *et al*., 2019
Lignocellulose-enriched compost	Novel Xylanase	Functional-based	Ellilä *et al*., 2019
Soil	β-glucosidase	Sequence-based	Gomes-Pepe *et al*., 2016
Switchgrass compost	Lignocellulose degrading enzymes	Sequence-based	Allgaier *et al*., 2010
Cow rumen	Biomass degrading genes	Sequence-based	Hess *et al*., 2011

1.5.1 Sequence-based screening

Through sequence-based screening, the discovery of novel genes is achieved by using DNA probes or primers designed based on conserved regions of homology. Target genes from metagenomic DNA are amplified using polymerase chain reaction (PCR) and the resultant amplicons are sequenced (Kodzius and Gojobom, 2015). This strategy is highly dependent on known or similar genes for designing the primers and DNA probes limiting the discovery of unknown and novel genes. The decrease in sequencing cost and advances in next-generation DNA sequencing (NGS) technologies has allowed the use of a shotgun metagenomic sequencing approach for the discovery of novel microbial genes present in an environmental sample (Jünemann *et al*., 2017). In shotgun metagenomic sequencing the environmental DNA is extracted and independently sequenced. It has the potential to provide a complete understanding of the diversity of the entire microbial community in a certain environment (Boers *et al*., 2019; Allgaier *et al*., 2010; Escobar-Zepeda *et al*., 2015).

The resultant metagenomic sequenced data is then subjected to computational analysis using bioinformatics tools to compare and find homologous sequences from known sequence data for functional annotation (Cheng *et al*., 2017; Knapik *et al*., 2019). The sequence-based approaches depend on the availability of closely related gene sequences for annotation of genes. The target genes from metagenomic sequence data can be synthesized for functional analysis (Culligan *et al*., 2014). In 2018 Thornbury identified and synthesized novel lignocellulose degrading enzymes from the porcupine microbiome using synthetic metagenomics. The sequence-based screening approach only detects a potential gene sequence without revealing function or activity of the gene. This would require an activity-based screening approach (DeCastro *et al*., 2016).

1.5.2 Function-based screening

Function-based screening, also known as activity-based screening, involves the expression of the metagenomic libraries in a suitable heterologous host followed by activity screening for enzymatic activities (Knapik *et al*., 2019; Cheng *et al*., 2017). Screening requires the use of reliable, high-throughput, and reproducible assay methods that have been developed over years for the detection of enzyme activities. High-throughput screening of glycoside hydrolases is achieved by using chromogenic or fluorogenic substrates and insoluble substrates to identify novel GHs. The success of functional screening also highly depends on proper expression of the metagenomic genes by the heterologous host, usually *E. coli*, and subsequently the assays used for the detection of enzymatic functions (Lam *et al*., 2015). The low-level expression and protein insolubility are the main limitations of the function-based screening approach. This is due to differences in codon usages, protein-folding elements, transcription and translation problems, and toxicity of the environmental genes to the host (DeCastro *et al*., 2016; Alessi *et al*., 2020). Different heterologous hosts like *Streptomyces lividans* or *Pseudomonas putida* and others have been used as alternatives to *E. coli* (Yun and Ryu, 2005). The functional-based metagenomic screening approach has however proven to be successful and led to the identification and discovery of many genes that encode completely novel and efficient GH from various environmental samples (Knapik *et al*., 2019). The advantage of using the function-based screening lies in its ability to identify and discover complete novel genes and enzyme activities without requiring knowledge of their sequences (Maruthamuthu and Van Elsas, 2017; Wang *et al*., 2019). Functional base analysis complements sequence-based screening in the annotation of metagenome genes (Lam *et al*., 2015).

1.6 Aims and objectives of the study

In a previous Ph.D. study by W. Nevondo (2016), a fosmid metagenomic library was constructed from the thermophilic stage of horse manure compost in search of lignocellulose degrading enzymes. The functional screening revealed 26 clones with β-xylosidase activity. In this study, three fosmid clones (P55E4, P81G1, and P89A4) were selected for further analysis. The main aim of this study was to functionally characterize the enzymes encoded by these genes to identify whether they conferred any novel properties. The objectives of the study were as follows:

- Cloning and expression of the GH genes in the pET21a expression vector.
- Protein purification using nickel-ion affinity chromatography.
- Biochemical characterization of the purified enzymes.
- Evaluation of hydrolytic products using TLC analysis
- Bioinformatics analysis of GH genes.

Chapter 2: Material and methods

2.1 Materials

All the chemicals and reagents used in this study were purchased from Sigma-Aldrich or Merck unless stated otherwise. The chromogenic PNP substrates: 4-nitrophenyl-β-D-glucopyranoside (*p*NPG), 4-nitrophenyl-β-D-xylopyranoside (*p*NPX), p-nitrophenyl-α-L-arabinofuranoside (*p*NPA), and 4-nitrophenyl-β-D-galactoside (*p*NPGal) were from Sigma-Aldrich. Polysaccharides used included beechwood xylan, birchwood xylan, Avicel PH101, and carboxymethyl cellulose (CMC) were from Sigma-Aldrich. The β-glucan (from barley) and lichenan (from moss) were from Megazyme; xylose, glucose, and cellobiose (Sigma); xylobiose and xylotriose standards from Megazyme.

2.2 Bacterial strains, plasmids, and culture conditions

All *E. coli* strains and plasmids used in this study are listed in (**Table 5** and **Table 6**) respectively.

Table 5: Bacterial strains used in this study.

Name	Genotype	Supplier
E. coli JM109	endA1, recA1, gyrA96, thi, hsdR17 (rK-, mK+), relA1, supE44, Δ (lac-proAB), [F' traD36, proAB, laqI^{q}ZΔM15]	NEB
E. coli BL21	F$^-$ ompT hsdSb (rB$^-$, mB$^-$) gal dcm araB: T7RNAP-tetA	Novagen (USA)
E. coli Rosetta pLysS	F- ompT hsdSB (rB- mB-) gal dcm (DE3) pLysSRARE (CamR)	Novagen (USA)

Table 6: Plasmids used and constructed in this study.

Plasmid(s)	Description	Supplier
pJET1.2/blunt	Blunt DNA ends for ligation of inserts in MCS, Amp^R	Thermo scientific
pET21a (+)	C-terminal His-tag expression vector, Amp^R	Novagen (USA)
pJET-xylP55	pJET1.2/ blunt ligated with a 1.1 kb P55 DNA insert.	In this study
pJET-bglP89	pJET1.2/ blunt ligated with a 2.3 kb P89 DNA insert.	In this study
pET21-XylP55	pET21a (+) containing the 1.1 kb *xylP55* gene cloned in the NheI and XhoI sites.	In this study
pET21-XylP81	pET21a (+) containing the 1.3 kb *xylP81* gene cloned in the NdeI and XhoI sites.	Biomatik
pET21-BglP89	pET21a (+) containing the 2.3 kb *bglP89* gene cloned in the NheI and XhoI sites.	In this study

All *E. coli* strains were cultured in Lysogeny Broth (LB) containing 1% (w/v) tryptone powder, 1% (w/v) sodium chloride (NaCl) and 0.5% (w/v) or LB agar (LB broth supplemented with 1% agar). Cultures were supplemented with ampicillin (100 µg/mL) unless stated otherwise and incubated at 37℃. Glycerol stocks were prepared and stored at -80℃ freezer for long time storage.

2.3 DNA methods

2.3.1 Fosmid DNA extraction

A previously constructed metagenomic library cloned in pCCFOS and transformed in *E. coli* Epi300) (Ph.D. thesis, W Nevondo) was screened for xylosidase activity. Three fosmid clones, P55E4, P81G1, and P89A4, were selected for this study. Single colonies from the -80℃ fosmid glycerol stock were inoculated in 10 mL LB broth supplemented with 0.02% (w/v) L-arabinose and 12.5 µg/mL chloramphenicol and incubated with shaking at 37℃ overnight. A 4 mL overnight culture was centrifuged at 3 000 x *g* for 20 minutes at 4℃ and fosmid DNA was

extracted using the QIAprep® Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions.

2.3.2 Alkaline lysis plasmid DNA extraction

E. coli cells with the plasmid of interest were picked with sterile tips, inoculated in 5mL LB broth, and incubated at 37℃ with shaking at (150 rpm) overnight. The overnight culture was harvested by centrifugation at 6 000 x *g*, for 3 min at 4℃ and plasmid DNA was extracted using the alkaline lysis method. Briefly, the cell pellet was resuspended in 100 µL Solution 1 (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8 and with 100 µg/mL RNase) and 200 µL solution 2 (200 mM NaOH, 1% (w/v) SDS; freshly prepared) was added, mixed and incubated on ice for 5 min. Following that 150 µL of Solution 3 (3 M potassium acetate) was added, incubated on ice for 5 min and the solution was centrifuged at 13 000 x *g* for 10 min at 4℃. Then 400 µL of the supernatant was mixed with 600 µL ice-cold absolute ethanol and incubated at -20℃ for 20 min. The plasmid DNA was collected by centrifugation at 13 000 x *g* for 10 min at 4℃, the resultant pellet was air-dried at room temperature for 30 min and resuspended in 50 µL deionized water.

2.3.2 DNA quantification

The DNA concentration and purity were determined by measuring absorbance at 260 nm and 280 nm using the Nanodrop spectrophotometer (ND-1000; Nanodrop Technologies, USA). Purity was determined by calculating the A260/280nm ration. Plasmid DNA was stored at 4℃ for further analysis.

2.3.3 Agarose gel electrophoresis

The size of PCR amplicons and double restricted DNA fragments were analysed on 1% (w/v) and 0.8% (w/v) agarose gels, respectively. Agarose gel was made using 1X TAE (40 mM Tris base, 20 mM acetic acid, and 1 mM EDTA at pH 8.0) buffer containing ethidium bromide

(0.5 µg/ml). The nucleic acid samples were prepared by mixing with 10X loading dye and 15 µL sample solutions were then loaded alongside a 1 kb DNA ladder (NEB) in the agarose gel placed in an electrophoresis tank filled with 1X TAE buffer. The gel was electrophoresed at 90V for 1 hour and the bands were viewed under UV illumination and analysed using Alpha-imager software. The Spectroline transilluminator (360nm) was used to view the gels for excision.

2.3.4 Restriction enzyme digestion

Plasmid constructs were confirmed by performing double restriction enzyme digestion with appropriate enzyme(s) in 50 µL reactions. Reactions contained 1 µg of plasmid DNA, 5µL 10X CutSmart buffer, 1 µL of each restriction enzyme and adjusted to 50 µL with distilled H_2O. The reaction mixture was incubated at 37°C overnight, terminated by heat inactivation at 65°C for 20 min, and analysed by agarose gel electrophoresis as described in (**section 2.3.3**)**.** The required DNA fragments were excised from the gel and purified using Nucleospin® Gel and PCR Clean-up kit (Machery-Nagel) according to the manufacturer's instruction and quantified as described in (**section 2.3.2**)**.**

2.4 Preparation of chemically competent cells

The *E. coli* cells glycerol stock (JM109 for cloning, BL21, and Rosetta pLysS for protein expression) were streaked on LB agar plates and grown overnight at 37°C. A single colony was inoculated into a 10 mL LB broth and incubated at 37°C with shaking overnight. A 1 mL overnight starter culture was inoculated into 100 mL LB broth incubated at 37°C shaking until the OD600 of 0.35-0.4 was reached. The cultures were cooled on ice for 30 min as were the sterile buffers, centrifuge bottles, and Eppendorf tubes. The cells were harvested by centrifugation at 3 000 x *g* for 15 min at 4°C in an Eppendorf centrifuge 5810 R (Merck). The cell pellet was resuspended in 20 mL ice-cold 100 mM $MgCl_2$ followed by centrifugation at 2 000 x *g* for 15 min at 4°C. The supernatant was discarded, and the pellet was resuspended in 40 mL ice-cold 100 mM $CaCl_2$ and incubated on ice for 20 min. The suspension was centrifuged at 2 000 x *g* for 15 minutes at 4°C and the resultant pellet was resuspended in 10 mL

ice-cold 85 mM $CaCl_2$ and 15% (v/v) glycerol and centrifuged at 1000 x *g*, for 15 min at 4°C. The pellet was finally resuspended in 2 mL ice-cold 85 mM $CaCl_2$ and 15% (v/v) glycerol and 50 µL aliquots were prepared into the ice-cold 1. 5 mL Eppendorf tubes. The cells were stored at -80°C.

2.5 Transformation of *E. coli* cells

Ligation reactions were transformed into chemically competent *E. coli* JM109, BL21, and Rosetta pLysS cells by the heat shock method. Competent cells (section 2.4) were thawed on ice and 5 µL of the ligation reaction was added to the competent cells and incubated on ice for 30 min. Cells were then heat shocked at 42°C for 30-45 sec and immediately incubated on ice for 2 min. After that 950 µL of LB broth (no antibiotic) was added and incubated at 37°C shaking (150 rpm) for 1 hour. After incubation 50 µL and 100 µL of the transformation mix was spread on LB plates with 100 µg/mL ampicillin and incubated overnight at 37°C. Random colonies were then sterilely picked for screening and confirmation by restriction enzyme digestion and further analysis.

2.6 Amplification and cloning

2.6.1 PCR amplification

The ORF of the fosmid clones P55 and P89 were amplified by PCR using Phusion® High-Fidelity DNA polymerase (NEB). Primers were designed with incorporated restriction sites to allow cloning into the pET21a (+) expression vector (Inqaba Biotech). Amplification of *xylP55* gene was amplified using forward primer 5´-**GCTAGC**ATGATAAAAGGCTTTA AC-3´ and reverse primer 5´-**CTCGAG**GTTCAAATATGGGTC-3´ which contained the NheI and XhoI restriction site in bold. On the other hand, *bglP89* gene was with forward primer 5´-**CATATG**AAAAGAATAATTGCTCTT-3´ and reverse primer 5´-**CTCGAG**GTCGATC AACTCAAAA-3´ containing NdeI and XhoI restriction sites in bold. Each 50 µL reaction contained 1X Phusion GC buffer, 200 µM dNTPs, 10 µM forward and reverse primers, 1U Phusion DNA polymerase, 250 ng template DNA, and water. Amplification was carried out in a Bio-Rad T100™ Thermal cycler for 25 cycles using the following conditions: 98°C, 30 s;

98°C, 10 s; 72°C, 30 s for *xylP55* and 2 min for *bglP89* and 72°C, 10 min. The PCR products were visualised on a 1% (w/v) agarose gel and the amplified PCR bands of *xylP55 and bglP89* gene were excised from the agarose gel and purified.

2.6.2 Cloning into pJET1.2/blunt vector

The purified PCR products were ligated into the pJET1.2/blunt cloning vector (**Figure 5;** CloneJet™) PCR Cloning Kit system in a vector to insert ratio of 3:1 following the manufacture's instruction. The 20 µL ligation reaction contained 2X reaction buffer, 1 µL pJET1.2/blunt cloning vector (50 ng/µL), 1 µL T4 DNA ligase (5 U/µL), the appropriate concentration of the purified PCR product, and nuclease-free water to make up to 20 µL. The ligation reaction was mixed and incubated at room temperature for 30 min.

The ligation reaction was then transformed into *E. coli* JM109 chemically competent cells as described in (**section 2.5**). Five colonies were randomly picked, and putative recombinant plasmid pJET-P55 and pJET-P89 were extracted and screened using restriction enzyme digestion to confirm the presence of the respective genes. Both plasmids pJET-P55 and pJET-P89 were digested with NdeI and Xho1 and the inserts were subsequently used for subcloning into pET21a (+) for overexpression.

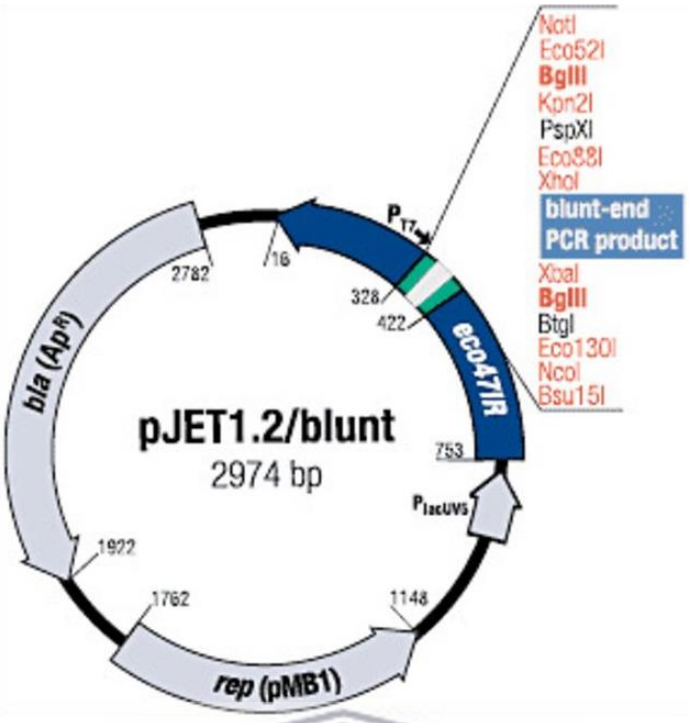

Figure 5: The pJET1.2/blunt cloning vector map (Thermo scientific).

2.6.3 Sub-cloning into pET21a (+) expression vector

The purified double digested *xylP55* and *bglP89* genes were ligated to a pET21a (+) expression vector (**Figure 6**) linearized with compatible restriction enzymes using the T4 DNA ligase (Thermo Fisher). The 20 µL ligation reaction contained 100 ng linear vector, insert DNA at a vector to insert molar ratio of 5:1, 2 µL 10X T4 DNA ligase, T4 DNA ligase (5 U), and nuclease-free water to 20 µL. The ligation reaction was incubated at room temperature for 1 hour.

Chemically competent BL21 and Rosetta pLysS *E. coli* strains (**section 2.5**) were transformed with the resulting plasmids pET21-P55, pET21-P81 (synthesized from Biomatik). Recombinant colonies were randomly picked, and putative recombinant plasmid constructs pET21-P55, pET21-P81 and pET21-P89 were extracted and screened using restriction enzyme digestion to confirm the presence of an inserted fragment (**section 2.3.4**) as follows: NheI and XhoI for pET21-P55; NdeI and XhoI for pET21-P81; and NdeI and Xho1 for pET21-P89.

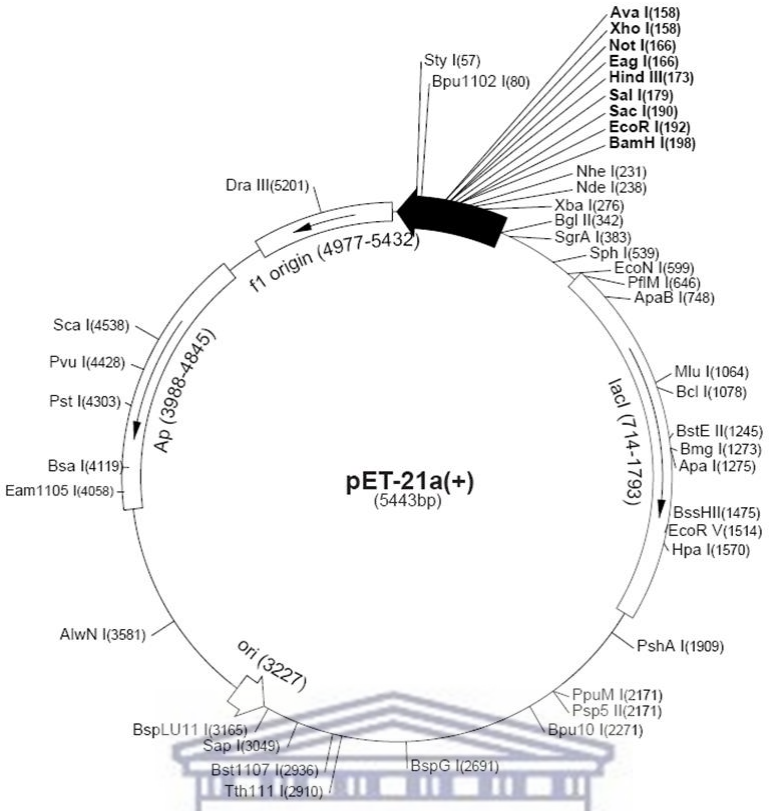

Figure 6: A map of the pET21a(+) expression vector used for expression of *xylP55*, *xylP81*, and *bglP89* genes with the ampicillin-resistance gene and C-terminal His-tag sequence for purification**.**

2.7 Protein expression and purification

2.7.1 Protein expression

E. coli BL21 and Rosetta pLysS cells transformed with plasmid constructs pET21-P55, pET21-P81, and pET21-P89 (**Table 6**) were inoculated in 10 mL LB with ampicillin (100µg/µl) and incubated at 37°C overnight. The overnight culture was inoculated into fresh 50 mL LB Amp and incubated at the same conditions until the $0D_{600nm}$ of 0.5-0.6 was reached. The optical density was monitored using a spectrophotometer at a wavelength of 600nm. Protein expression was thereafter induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the cultures were incubated at 37°C (XylP55 and XylP81) and 28°C (BglP89) with shaking (150 rpm) overnight. Thereafter, the cells were harvested by centrifugation at 3000 x *g* for 20 min at 4°C. The cell pellet was resuspended in 50 mM NaCl and 20 mM Tris-HCl at (pH 8) and cells were lysed by sonication on ice using

the Bandelin Sonopuls HD 2070 (58% power, 5 cycles of 30 sec and 30 sec pausing between cycles). Cell debris was centrifuged at 13 000 x *g*, for 20 min at 4℃ keeping both the supernatant (soluble fraction) and pellet (insoluble fraction). The soluble and insoluble fractions were analysed using SDS-PAGE.

2.7.2 Bradford Assay

The protein concentration of the resultant proteins was determined by the Bradford method (Braford, 1976) using bovine serum albumin (BSA) as a standard with a concentration range of 0-1.5 mg/mL. Briefly, 5 µL of the various standard and appropriately diluted protein samples were mixed with 200 µL Bradford's reagent solution in triplicate into a 96 well U-shaped microplate (Greiner biome) and incubated at room temperature for 10 min. The absorbance of all samples was measured at 595nm using a spectrophotometer plate reader (SPECTRO star Nano, BMG LabTech UK). A standard curve of absorbance vs BSA concentrations was used to determine the protein concentration of the purified protein samples in mg/mL.

2.7.3 SDS-PAGE Analysis

SDS-PAGE gels were made according to the Laemmli method (1970) to analyse protein expression levels in both soluble and insoluble fractions and determine purity. The gels consisted of a 12% resolving gel and a 4% stacking gel. Protein samples were prepared by mixing 40 µL of the uninduced, induced, purified fractions and 20 µL 5X sample loading buffer (0.05% (w/v) bromophenol blue, β-mercaptoethanol, 10% (w/v) SDS, 0.5 M Tris-HCl, (10% v/v) glycerol). The protein samples were heated at 95℃ for 10 min on a heating block. The first well on the gel was loaded with 10 µL broad range prestained protein molecular weight marker (NEB) alongside 15-20 µL protein samples. The gels were electrophoresed in a 1X running buffer (25 mM Tris-base, 186 mM glycine and 0.1% (w/v) SDS) at 200 V for 1 hour. After electrophoresis, the gel was stained with Coomassie brilliant blue staining solution (0.124% (w/v) Coomassie blue R250, 50% (v/v) methanol, and 10% (v/v) acetic acid) with gentle shaking at room temperature overnight. Following that the gels were destained

with a destaining solution (50% (v/v) methanol and 10% (v/v) acetic acid) until the bands were visible.

2.7.4 Affinity Chromatography

For purification of recombinant proteins, the cultures were scaled up to 250 mL and incubated at 37℃ for XylP55 and XylP81 and 28℃ for BglP89 for expression of genes. After expression, the cells were harvested by centrifugation at 3 000 x *g* for 20 min at 4℃ and the resultant pellet was resuspended in 50 mM NaCl and 20 mM Tris-HCl (pH 8). The cells were sonicated as described (section 2.7.1), centrifuged at 13 000 x *g*, for 20 min at 4℃ and the supernatant (soluble fraction) containing the His-tagged recombinant protein was filtered through a 0.45 µM pore syringe filter. The recombinant His-tag XylP55, XylP81, and BglP89 proteins were purified using nickel affinity chromatography. The column was packed under gravity flow with 4 mL His-bind resin, charged and equilibrated by washing the column with 6 mL of sterile deionized water, 10 mL 1X charge buffer (50 mM $NiSO_4$), and 6 mL 1X binding buffer (50 mM NaCl and 20 mM Tris-HCl (pH 8). The filtered protein extracts were then loaded onto the equilibrated column, washed with 20 mL 1X binding buffer to allow proteins to bind to the resin, and washed to remove any unbound proteins with 12 mL 1X wash buffer (500 Mm NaCl, 20 mM Tris-HCl, and 60 mM imidazole at pH 8). The bound recombinant protein of interest was eluted with 12 mL 1X elute buffer (0.5 M NaCl, 20 mM Tris-HCl, 1M imidazole at pH 8). All flow-through fractions were collected to be evaluated by SDS-PAGE analysis. The eluted and purified protein samples were dialyzed overnight to remove the imidazole and any other salts against the dialysis buffer (50 mM Tris-HCl at pH 7.5) using Slide-A-Lyzer® 10K dialysis cassettes (Thermo scientific). Dialysis was performed overnight with continuous stirring at 4℃. The dialyzed protein samples were concentrated using an Amicon® Ultra-15-centrifugal filter device with a 50 kDa molecular weight cut off according to the manufacturer's instructions. Proteins were stored at 4℃ for further analysis.

2.8 Enzyme activity assay

The recombinant proteins XylP81 and BglP89 were selected for further characterization. Enzyme activity of the recombinant XylP81 and BglP89 proteins was determined by measuring the hydrolysis of chromogenic substrates *p*NPX (4-nitrophenyl-β-D-xylopyranoside) and *p*NPG (4-nitrophenyl-β-D-glucopyranoside) respectively. The enzyme reaction mixture contained 2mM substrate dissolved in 50mM sodium phosphate buffer (pH 6) to a total volume of 240 µL and 0.5 µg of the purified enzyme was added to initiate the reaction. The reaction was incubated at 37℃ for 10 min and terminated by the addition of 1 mL 1 M Na_2CO_3. The amount of *p*-nitrophenol released was determined by measuring absorbance at 410 nm using a spectrophotometer plate reader (SPECTRO star Nano, BMG LabTech UK). A *p*-nitrophenol standard curve was constructed to determine the amount of *p*-nitrophenol released through enzyme activity. One unit (U) of enzyme activity was defined as the amount of enzyme that releases 1µmol of *p*-nitrophenol from the substrate under the assay conditions. All reactions were performed in triplicate with appropriate negative controls.

2.9 Biochemical characterization of XylP81 and BglP89

The optimum pH of the purified recombinant proteins (XylP81 and BglP89) was examined measuring enzyme activity in a pH range of 3-9 using three different buffer systems following the enzyme assay conditions above (section 2.8). The buffers used were 50 mM citrate phosphate buffer (pH 3-5), 50 mM sodium phosphate (pH 6 and 7), and 50 mM Tris-HCl (pH 8 and 9) respectively. The optimum temperature was determined by measuring the enzyme activity at different temperatures (20, 30, 40, 50, 60, 70, and 80℃) for 10 min at the optimum pH of 6. Relative activity (%) was calculated as the specific activity at each pH or temperature divided by the specific activity at the optimum.

The effect of metal ions was evaluated by measuring enzyme activity of XylP81 and BglP89 with the addition of 5 mM metal ion(s) (Ag^{2+}, Ca^{2+}, Cu^{2+}, Fe^{2+}, K^{+}, Mn^{+}, Mg^{2+}, Na^{+}, Ni^{2+} and Zn^{2+}) into the reaction mixture. Enzyme activity of the sample without the addition of metal ions was considered as 100% enzyme activity.

2.9.1 Thermal stability assay

The thermal stability of XylP81 and BglP89 was evaluated by incubating the purified enzymes in 50 mM sodium phosphate buffer at pH 6 for 1 hour at different temperatures range (50℃, 55℃, 60℃, and 70℃) for time intervals of 10, 20, 30, 40, and 60 min. 10 µL samples were withdrawn and cooled on ice for 5 minutes. Thereafter enzyme activity was determined for each time point at 50℃ and pH 6 as described in (**section 2.8)**. The control unheated sample enzyme activity was considered to be 100% enzyme activity.

2.9.2 Enzyme kinetics

Enzyme kinetic parameters: Michaelis-Menten constant (K_M), maximum reaction velocity (V_{max}), catalytic rate constant (k_{cat}), and catalytic efficiency (k_{cat}/ K_M) for XylP81 and BglP89 were determined by measuring the initial velocity of the enzyme reaction. The velocity was measured by assaying enzyme activity on *p*NPX for and *p*NPG at a substrate concentration of (1 mM – 20 mM). Enzyme assays were conducted as described (s**ection 2.8**) at 50℃ and pH 6 for 10 min. The kinetic parameters K_M and V_{max} were estimated by plotting the initial velocity data against substrate concentration using the non-linear regression curve fitting of the Michaelis-Menten equation using GraphPad Prism version 5.00 (GraphPad Software, Inc., San Diego, CA, USA).

2.9.3 Substrate specificity

Enzyme specificities of XylP81 and BglP89 were investigated using different chromogenic *p*NP-linked substrates which included: 4-nitrophenyl-β-D-glucopyranoside (*p*NPG), 4-nitrophenyl-β-D-xylopyranoside (*p*NPX), p-nitrophenyl-α-L-arabinofuranoside (*p*NPA), and 4-nitrophenyl-β-D-galactoside (*p*NPGal). Assays were performed at 50℃ in 50 mM sodium phosphate buffer pH 6 using the standard assay conditions above. The substrate specificity of XylP81 and BglP89 was also determined using different natural substrates (polysaccharides) and was determined by measuring the amount of reducing sugar released using the dinitrosalicylic (DNS) method (Miller, 1959). XylP81 was tested on 1% (w/v) arabinan (sugar

beet), arabinoxylan (rye), beechwood, and birchwood xylan to assay for β-1, 4-endoxylanse activity. BglP89 was assayed on various polysaccharides 1% (w/v): cellobiose, lichenan (moss), β-glucan (from barley), Avicel PH101, and carboxymethyl cellulose (CMC) to determine β-1, 4-endoglucanase activity. Briefly, 900 µL (1% w/v) substrate in 50 mM sodium phosphate buffer at pH 6 was mixed with 10 µg appropriately diluted enzyme. The reaction was incubated for 1 hour in a 50℃-water bath. The reaction was terminated by the addition of 1.5 mL DNS reagent (10 g of 3,4 dinitrosalicylic acid, 300 g potassium sodium tartrate tetrahydrate, and 16 g NaOH anhydrous), boiled for 10 minutes, and allowed to cool immediately on ice. The absorbance was measured at 540 nm and the amount of reducing sugars released and the specific activity was determined using xylose as the standard for XylP81 and glucose as the standard for BglP89. Each assay was performed in triplicate.

2.9.3.1 TLC analysis of the hydrolytic products

The hydrolytic products formed by the enzymes on natural polysaccharides were further analyzed. BglP89 activity was tested on a range of 1% substrates which included cellobiose, Avicel PH101, CMC, lichenan, and β-glucan. The resultant hydrolytic products were analyzed using thin-layer chromatography (TLC). All the enzymatic reactions were performed as described in (**section 2.9.3)** taking aliquots of 100 µL after 1, 6, 16, and 24 hours of incubation. Enzyme reactions were heated at 100°C for 10 min to terminate the reaction and centrifuged at 12 000 x *g* for 20 min to remove unhydrolyzed substrate from the reaction. The resultant supernatant was concentrated using the Refrigerated CentriVap Concentrator (Labconco) and resuspended in 10 µL distilled water. Appropriate controls which included enzymes without substrate were incubated for 1 hour. The samples (8 µL) as well as 2 µL mixture of standards (glucose, cellobiose, and cellotriose) solution (0.2 mM) were spotted on a TLC silica gel 60 F254 aluminium coated plates (Merck). The spotted TLC were separated in butanol: ethyl acetate: acetic acid: water (80:10:5:5 v/v) as mobile phase three to four times and then dried using a heating gun in a fume hood. TLC plates were visualized by dipping in 1:1 (0.2%) methanolic orcinol and (20%) sulphuric acid stain and heat dried until the spots are visible.

2.10 End product inhibition

End product inhibition on the activity of both XylP81 and BglP89 was evaluated by testing the effect of xylose and glucose on enzyme activity respectively. The effect of xylose on XylP81 activity was determining activity on 2 mM *p*NPX substrate in the presence of xylose (100 mM, 200 mM, 400 mM, 500 mM, 1000 mM, 2000 mM, and 3000 mM) over 20 min under the standard assay conditions. Also, the effect of glucose on BglP89 activity was determined by incubating it with 2 mM *p*NPG substrate in the presence of glucose (100 mM, 200 mM, 400 mM, 500 mM, 1000 mM, 2000 mM, and 3000 mM) over 20 min under the standard assay conditions. XylP81 and BglP89 reactions incubated without sugar were used as 100 % activity.

2.11 Bioinformatics and sequence analysis

The open reading frame (ORF) of the fosmid DNA clones P55E4, P81G1, and P89A4 was detected using the ORF finder tool from the National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/orffinder/). The DNA sequence was translated into a protein sequence using the translate tool in expasy (https://web.expasy.org/translate/). The deduced protein sequences were searched against the Pfam protein database (https://pfam.xfam.org/) to identify which glycoside hydrolase (GH) family each protein belongs to. Multiple sequence alignments were performed using ClustalW (https://www.ebi.ac.uk/Tools/clustalw2/) with closest related homologous protein sequences identified using the BLASTp tool (http://www.ncbi.nlm.nih.gov/BLAST) and characterized proteins from the same respective GH family from the CAZy database. The resultant multiple sequence alignments were used to generate a phylogenetic tree analysis for XylP55, BglP81, and XylP89 using MEGA 7 using the maximum like-hood analysis (1000 bootstrap replicates).

Chapter 3: Results and Discussion

3. Introduction

Metagenomics is a culture-independent approach for extraction and expression of genetic diversity from unculturable microorganisms found in various environmental samples. Metagenomic gene discovery usually involves one of two approaches, sequence-based or function-based screening. Environments like hot springs (Joshi *et al*., 2020, rumen (Cheng *et al*., 2012; Li *et al*., 2014), soil (Cheng *et al*., 2017; Alves *et al*., 2018), and compost (Ellilä *et al*., 2019) have become desirable environments to screen to identify novel GH enzymes for lignocellulose degradation. The composition of compost, in particular, makes it a promising sample site for the discovery of novel thermostable GH enzymes (Dougherty *et al*., 2012; Lemos *et al*., 2017). There are even a few examples of industrially relevant GH enzymes, discovered through metagenomics from compost environments, adding to the appeal of these environments for novel enzyme discovery (Matsuzawa *et al*., 2015; Uchiyama *et al*., 2013).

In a previous study, a metagenomic library was constructed using mDNA extracted from the thermophilic stage of horse manure compost, cloned into the fosmid vector pCCFOS, and transfected in *E. coli* Epi300 (Ph.D. thesis, W. Nevondo 2016). Analysis of the metagenomic microbial diversity indicated that the environment is mostly dominated by bacteria belonging to the Bacteroidetes phylum. Functional screening of the library for β-xylosidase activity gave 26 hits from 20000 clones screened, and the DNA sequence of 18 fosmid clones was determined. Based on preliminary bioinformatics analysis of the identifiable GHs, together with initial substrate range characterization, using cell lysates from transfectants harbouring complete fosmids, three fosmids (P55E4, P81G1, and P89A4) were identified as encoding putative novel β-xylosidases. Moreover, these were further subjected to secondary screening against three additional *p*NP chromogenic substrates (*p*NPA, *p*NPC (4-nitrophenyl-β-D-cellobioside), and *p*NPG) and interestingly showed multiple activities against the substrates in the secondary screening indicating potential novel multifunction glycoside hydrolase enzymes.

GHs with high catalytic efficiency, high substrate specificity and affinity, which can withstand harsh industrial conditions, and are resistant to product feedback inhibition, would be highly valuable in biotechnological and industrial processes (Colombo *et al*., 2016). Therefore, this study focuses on the functional characterization of the glycoside hydrolases identified from these metagenomic library clones to determine if their biochemical properties differ substantially from those that have been described previously, potentially signalling that these can act as novel catalysts.

3.1 Results

3.1.1 Annotation of fosmid clones

The P55E4 fosmid clone DNA sequence was analysed and the putative ORFs were identified (**Figure 7**). One ORF designated XylP55, coding a 366 amino acids protein with a predicted mass of ~42kDa was detected and showed a high sequence identity to a putative GH43 glycoside hydrolase from *Flammeovirgaceae* bacterium 311 (WP_061989802.1; 100% query coverage and 70.15% identity), belonging to the diverse Bacteroidetes phylum. This suggested that the protein encoded by this ORF might be responsible for the activity on *p*NPX detected from the initial screening of the fosmid library and was selected for further analysis. An ORF coding for an HTH-type transcriptional regulator upstream of the GH43 family protein ORF which might be responsible for regulation of the expression of the gene was observed.

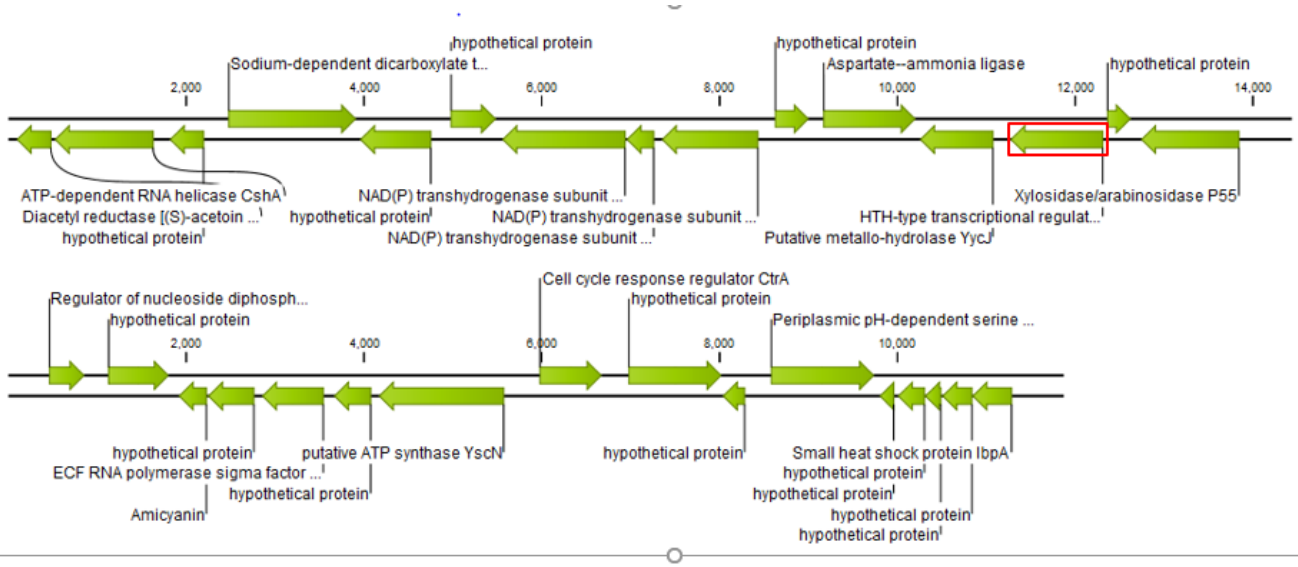

Figure 7: Gene organisation of ORFs identified in the fosmid clone P55E4 DNA sequence. The red box is the ORF coding for XylP55.

A similar analysis for the P81G1 fosmid clone DNA revealed one ORF (**Figure 8**) designated XylP81 in this study coding for a 456 amino acid protein (~53 kDa) which shared the highest sequence identity to a β-xylosidase (74%) from *Anaerolineaceae* bacterium (99% query coverage; MAU09869.1). This suggested that this ORF might be responsible for the activity on *p*NPX detected in the initial screening of the fosmid library and was selected for further analysis. The fosmid analysis also showed the presence of genes encoding three oligopeptide transport systems proteins in the vicinity of the β-xylosidase gene. These proteins showed high sequence identity 80.6%, 83.4% and 80% to an uncharacterised ABC transporter ATP-binding protein from *Chloroflexi* bacterium (NWG21180.1; 96% query coverage), a peptide ABC-transporter from *Anaerolineaceae* bacterium (MAU09876.1; 100% query coverage) and ABC-transporter from *Anaerolineaceae* bacterium (MAU09075.1; 100% query coverage), respectively. ABC transporters proteins are known to be responsible for the transportation of a variety of substrates including xylooligosaccharides (e.g. xylobiose or xylotriose) across the cellular membrane for hydrolysis to xylose by the intercellular enzymes (Wilkens 2015).

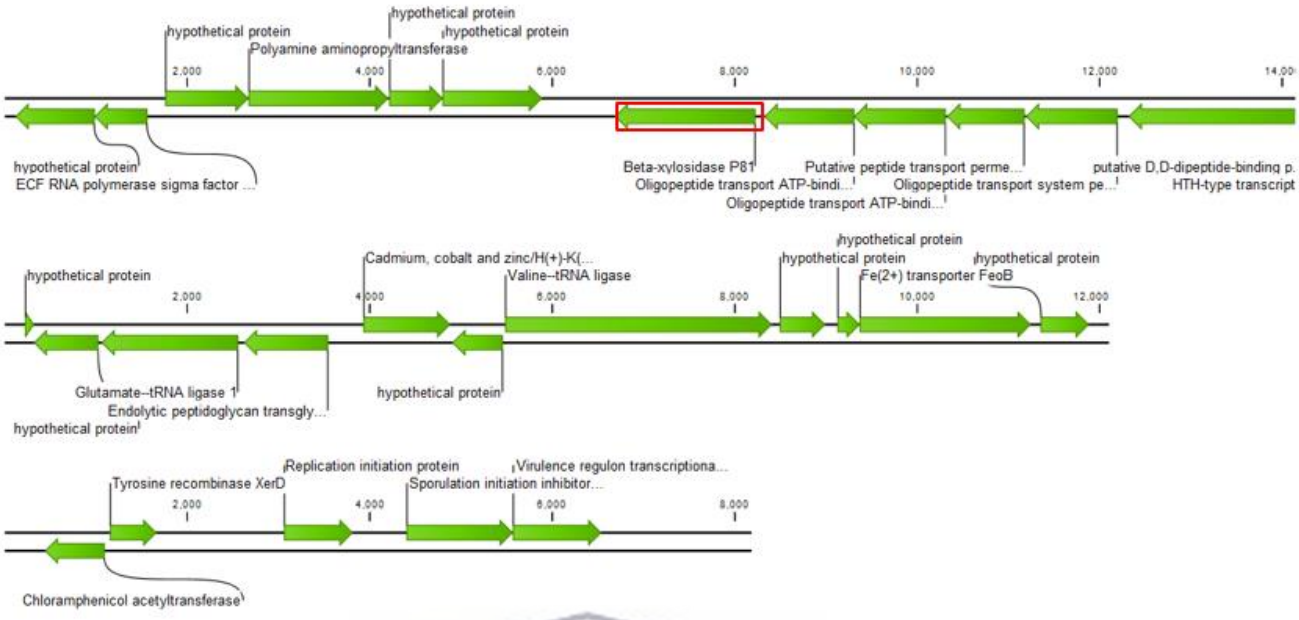

Figure 8: Gene organisation of ORFs identified in the fosmid clone P81G1 DNA sequence. The red box is the ORF coding for XylP81.

The DNA sequence of the P89A4 fosmid clone revealed three ORFs that are related to GHs (**Figure 9)**. The third ORF designated BglP89, coding for a putative β-glucosidase (766 amino acids, ~ 83kDa) showed the highest amino acid sequence identity to a putative β-glucosidase from *Pontibacter korlensis* (WP_148561686; 99% query coverage, the identity of 68%). Further analysis revealed that it could be a member of the GH3 family which is known to contain diverse enzyme activities with reports of bi- or multifunctional enzyme and could be responsible for the detectable β-xylosidase activity in the initial screening of the P89A4 fosmid clone. This ORF was then selected for further analysis and characterization in this study.

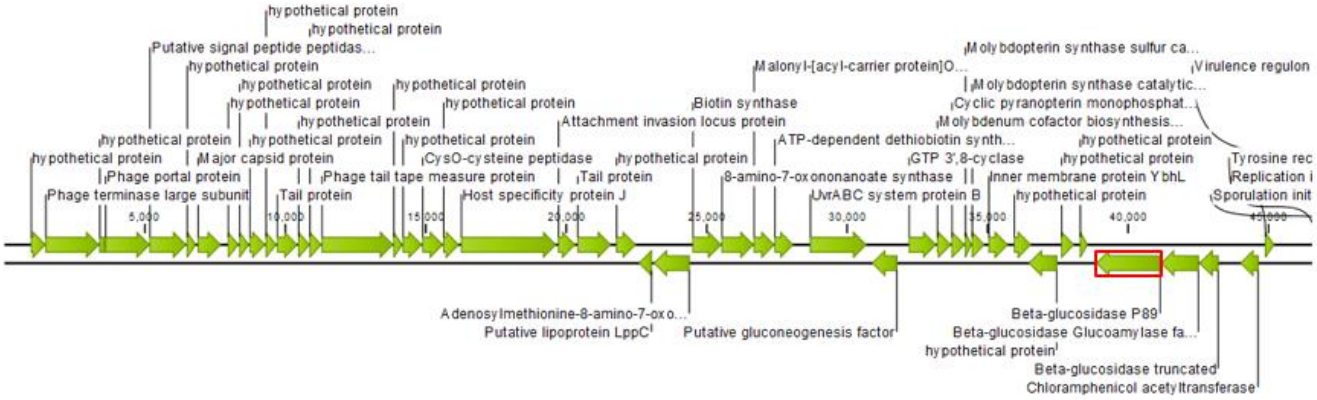

Figure 9: Gene organisation of ORFs identified in the fosmid clone P89A4 DNA sequence. The red box is the ORF coding for BglP89.

3.1.2 XylP55 sequence analysis

The XylP55 deduced amino acid sequence showed sequence similarity to GH43 proteins and specifically a characterized GH43 β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome (**Table 7**). Its closest hit against compost metagenomes on the IMG/M database was protein Ga0207873_100148011 from a cellulose-adapted compost microbial community from Newby Island Compost Facility, Milpitas, CA, USA (69.5% sequence identity / 99% query coverage). Pfam and dbCAN2 analysis also classified XylP55 into the GH43 family and SignalP analysis of the amino acid sequence showed that it contained one distinct GH43 catalytic domain located between amino acid residues 53-363 and a cleavage site between amino acid position 25 and 26. The GH43 family is one of the largest GH families and known to contain enzymes with various activities which include β-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55), arabinase (EC 3.2.99), and xylanase (EC 3.2.1.8), classified as per the CAZy database (www.cazy.org/GH43.html) with reports of multifunctional and bifunctional enzymes (Wongratpanya *et al.*, 2016). Due to the large number of proteins deposited in this family, it has been divided into 37 subfamilies with various enzyme activities, based on conserved amino acid sequences, biochemical properties, and structural analysis (Mewis *et al.*, 2016; Matsuzawa *et al.*, 2017).

Table 7: Closest sequence homologs of XylP55 based on a BLASTp NCBI database search.

Organism/Sample	Protein type/name	Query coverage	% Identity	Accession number
Flammeovirgaceae bacterium 311	GH43 family protein	100%	70.5	WP_061989802.1
Pontibacter lucknowensis	GH43 family protein	95%	70.0	SIR34655.1
Pontibacter roseus	GH43 family protein	95%	70.0	WP_157579052.1
Saccharophagus degradans	GH43 family protein	92%	72.2	WP-011467305.1
Compost metagenome	β-xylosidase / α-arabinofuranosidase	85%	76.4	BAS02081.1

The alignment of XylP55 (**Figure 10**) with other characterized GH43_1 protein sequences revealed that XylP55 contains a general base (Asp62) and general acid (Glu243) which are important in catalysis of GH43 β-xylosidases. The conserved Asp181 found in all GH43 enzymes is proposed to be involved in pKa modulation and orientation of the general acid residue as well as in substrate binding and transition state of the enzyme (Wagschal *et al*., 2009). The critical Trp272 and His321 amino acids, responsible for the coordination of calcium ions, are also conserved in the sequences. XylP55 also showed amino acid sequence similarities with a characterised GH43 β-xylosidase/α-arabinofuranosidase from a compost metagenome (Matsuzawa *et al*., 2015) which is classified into GH43 subfamily 1 (GH43_1), suggesting that XylP55 might also belong to the same subfamily. The evolutionary relationship or origin of XylP55 was determined using phylogenetic analysis performed with characterized GH43 β-xylosidase proteins classified into the GH43_1 subfamily from the CAZy database and included the five top hits based on sequence similarity through BLASTp analysis. Phylogenetic analysis of XylP55 was constructed with GH43_1 β-xylosidase and the

closest homologs (**Figure 11**) showed that XylP55 is closely related to a characterized β-xylosidase/α-L-arabinofuranosidase from a compost microbial metagenome.

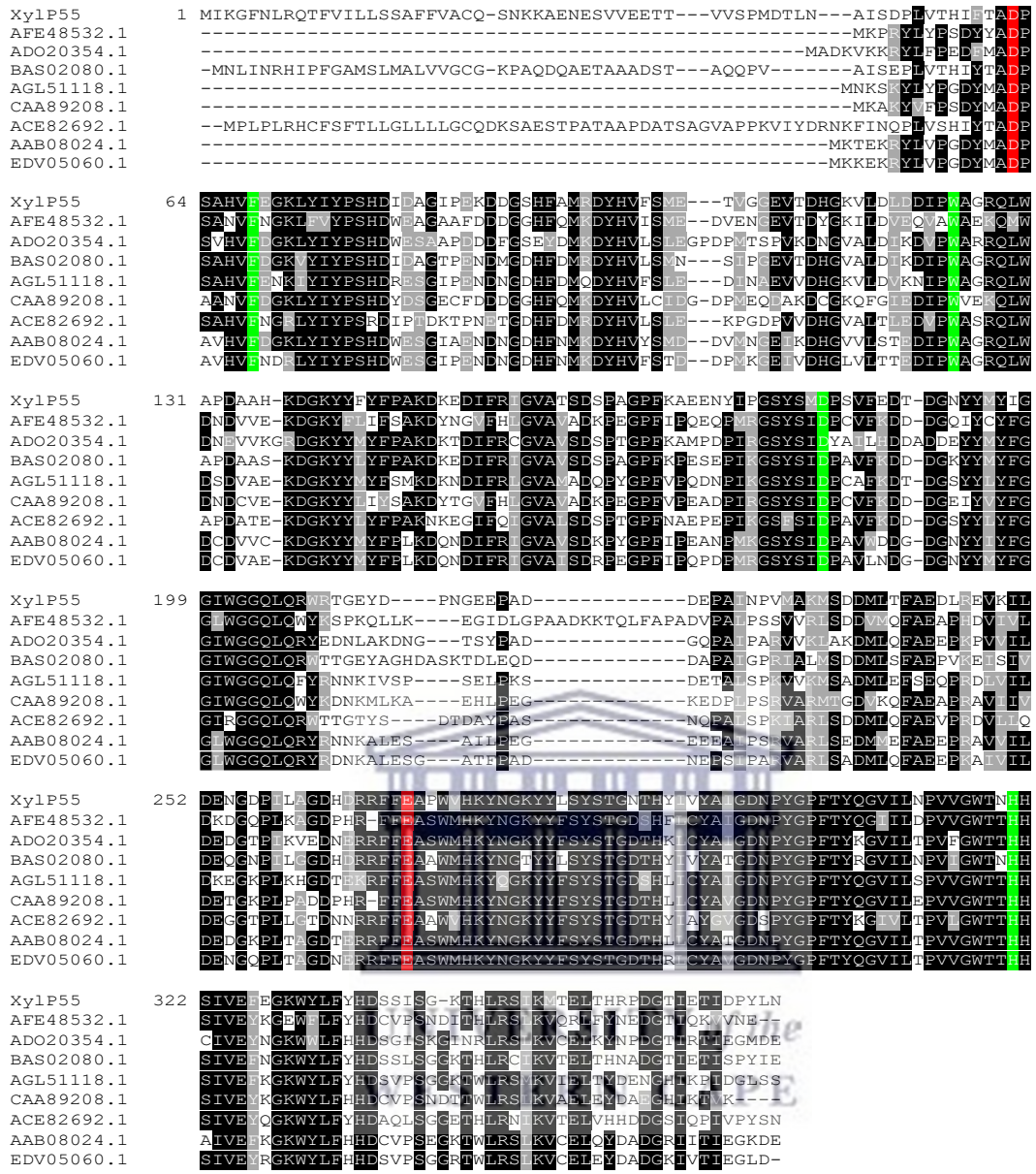

Figure 10: Multiple sequence alignment of XylP55 with characterized GH43 subfamily 1 proteins from the CAZy database. The sequences include; a β-xylosidase from uncultured rumen bacterium (AFE48532.1), β-xylosidase/ α-arabinofuranosidase from uncultured rumen bacterium (ADO20354.1), β-xylosidase/α-arabinofuranosidase from metagenome compost (BAS02081.1), β-xylosidase from *Sphingobacterium* sp. HP455 (AGL5118.1), β-xylanase/β-xylosidase from *Prevotella bryantii* B14 (CAA89208.1), β-xylosidase/ α-arabinofuranosidase from *Cellvibrio japonicus* Uedio7 (ACE82692.1), β-xylosidase/ α-arabinofuranosidase from *Bacteroides ovatus* (AAAB08024.1) and protein from *Bacteroides* intestinal DSM17393 (EDVO5060.1). The black and grey shaded areas represent 70% sequence identity and similarity. The conserved catalytic site residues D62 and E243 are shaded red and D181 with other amino acid residues important in catalysis of GH43 β-xylosidase enzymes shaded green**.**

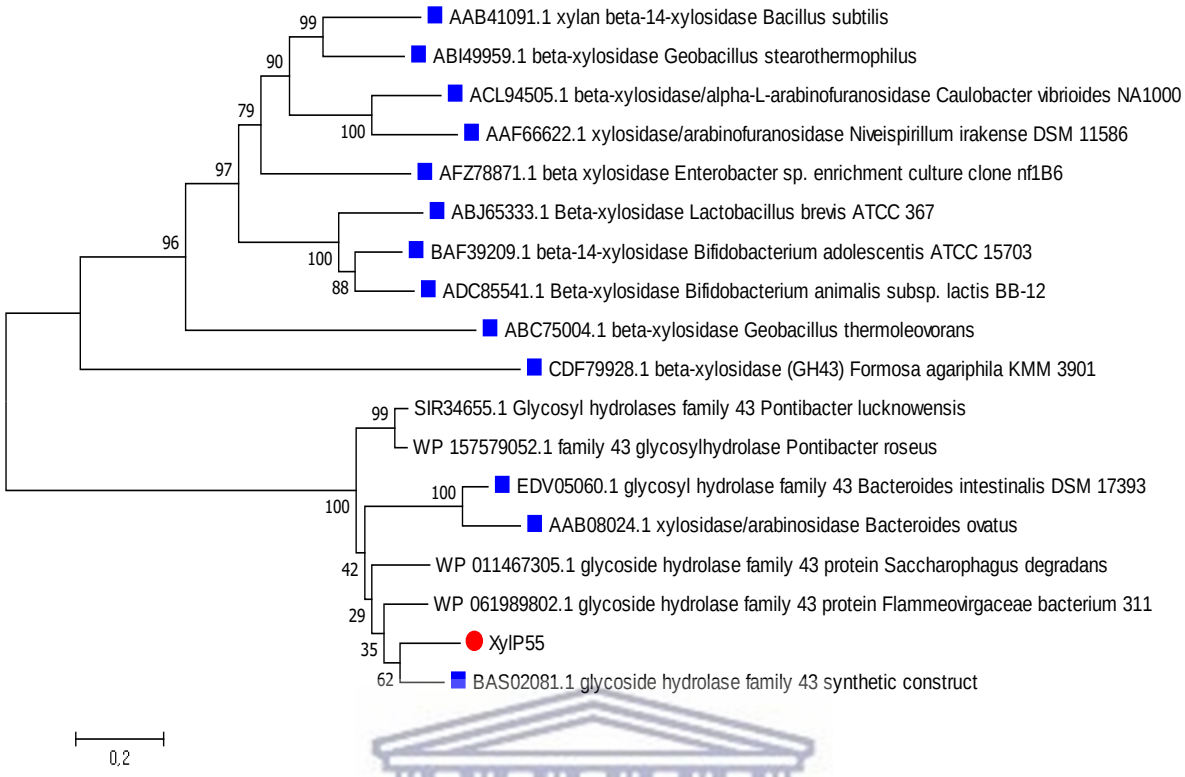

Figure 11: Evolutionary analyses of XylP55 GH43 subfamily 1 β-xylosidase from the CAZy database constructed in MEGA7 using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. XylP55 is represented with the red dot and the blue squares represent all the characterized GH43 β-xylosidase enzymes.

There are currently two crystal structures available from the GH43_1 subfamily. One being the structure of the β-xylosidase/α-arabinofuranosidase from a compost metagenome (BAS02081.1) which has (76.4%) sequence similarity to XylP55. We, therefore, took the opportunity to compare these two enzymes to see if there are notable differences (**Figure 12**). The high amino acid identity between these proteins allowed for a high-quality homology model to be constructed using the apo (no calcium) form of this protein (5glk; QMEAN = 0.78). Only residues 47-369 of the original coding sequence (LC025936) isolated from the metagenome were expressed for characterization and crystallization of that enzyme, thus the model built here for XylP55 only covers this region. The first portion of the protein was considered to be a secretion signal, which is why it was removed. Alignment of the protein structures resulted in an RMSD of 0.091, which should be expected given the nature of how the model was constructed (template-based) and high sequence similarity. Visual inspection of the XylP55 model compared with that of the β-xylosidase/α-arabinofuranosidase

(BAS02081.1) showed highly similar structures with no obvious deviations. Given the amino acid and presumed structural similarity to BAS02081.1, it would be expected that their biochemical characteristics would be highly similar.

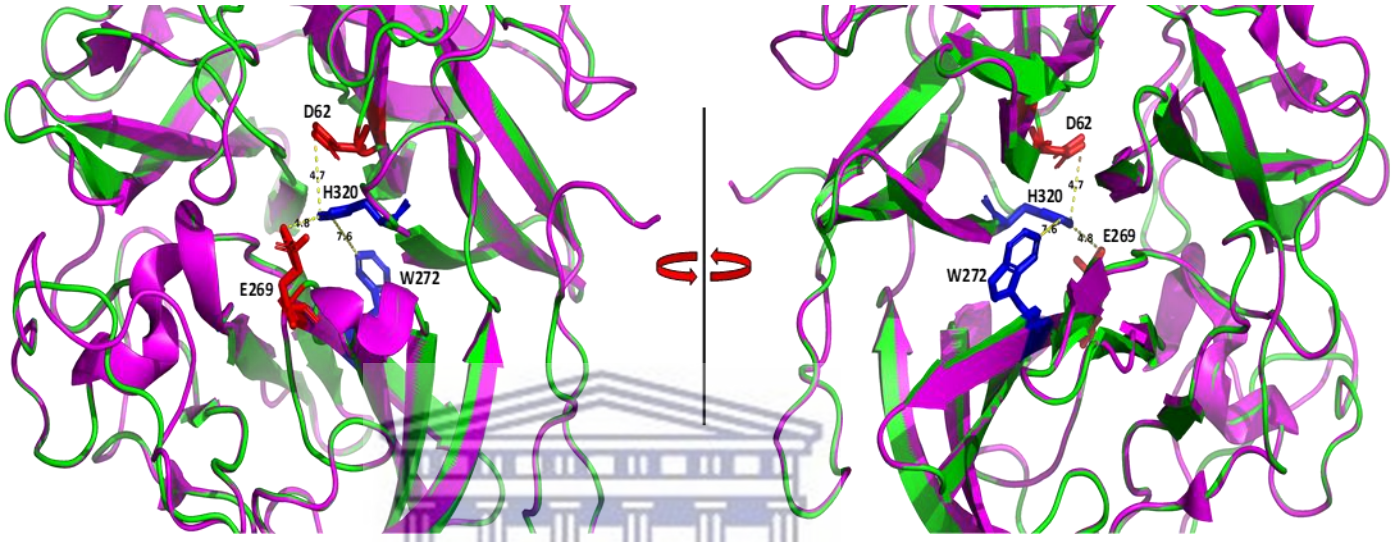

Figure 12: Alignment of XylP55 structure modelled on BAS02081.1 (5glk) as a template structure. The cartoon presentation of the protein backbone for BAS02081.1 is shown in magenta while P55 is in green. Catalytic residues are shown as sticks coloured red while those involved in coordinating calcium ions are in blue. Interatomic distances are indicated for scale.

3.1.3 XylP81 sequence analysis

The deduced amino acid sequence of XylP81 showed the highest sequence identity (79% identity / 100% query coverage) to a putative AraC-like regulator from a *Chloroflexi*-like bacterium (**Table 8**) assembled from metagenomic data of hydrothermal vents (Zhou *et al*., 2020). Lower hits indicate the relation to glycoside hydrolases and specifically β-1,4-xylosidase from a marine metagenome bacterium belonging to the *Anaerolineaceae* (Tully *et al*., 2018). To determine if there is any AraC-like transcriptional regulator-like motif present in XylP81, we used the Motif Search tool available through (expasy.org). No HTH motif could be identified, however, a GH39 domain that covers the full length of the proteins was suggesting that annotation of AraC *Chloroflexi* protein is misleading. In a study by

Maruthamutha *et al.*, 2017 a gene coding a β-1,4-xylosidase was originally annotated as an AraC family transcriptional regulator by the RAST server (http://RAST.nmpdr.org.), which also lacked a helix-turn-helix DNA binding motif. There is however an *Aeromonas caviae* gene (*xysA*) which codes for β-1,4-xylosidase (BAA95685.1) that has an AraC/Xyls transcription regulator family domain upstream of the GH domain, published only in the database. A BLASTp analysis of this *Aeromonas caviae* β-1,4-xylosidase revealed high sequence identity to another bacterial β-1,4-xylosidases which also harbour the AraC domain. None of them are characterized yet, limiting the information on the importance of the AraC domain on these putative β-1,4-xylosidases. Comparison of the domain architecture of XylP81 and the β-1,4-xylosidase from *Aeromonas* caviae further confirmed that XylP81 does not contain an AraC domain (**Figure 13**).

Table 8: Closest sequence homologs of XylP81 from the BLASTp search.

Organism	Protein name	Query coverage	% identity	Accession number
Chloroflexi bacterium	TPA: AraC family transcription regulator	100%	79	HDU42522.1
Anaerolineaceae bacterium	β-xylosidase	99%	74.4	MAU09869.1
Chloroflexi bacterium	TPA: β-xylosidase	98%	65.78	HGV25578.1
Anaerolineaceae bacterium	Glycoside hydrolase (partial)	63%	84.08	KAB2866579.1
Ruminiclostridium cellobioparum	β-xylosidase	98%	45.59	WP_004631070.1

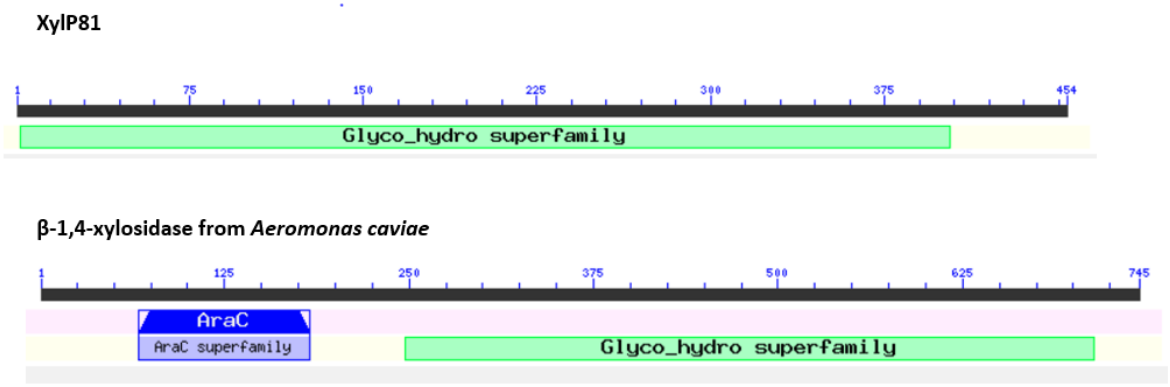

Figure 13: Comparison of the domain architecture of XylP81 and a β-1,4-xylosidase from *Aeromonas caviae* from NCBI conserved domain search tool.

Pfam and dbCAN2 analysis suggested that XylP81 belongs to the GH39 family with a single GH39 domain. The GH39 family is thus far known to contain β-xylosidases (EC 3.2.1.37) and α-L-iduronidases (EC 3.2.1.76) (www.cazy.org/GH39.html). Only 12 β-xylosidases are characterised in the GH39 family on the CAZy database limiting the functional understanding of β-xylosidases in the family. XylP81 showed relatively low sequence similarity with other functionally characterized GH39 β-xylosidases, therefore XylP81 might have novel biochemical characteristics in this GH family where specific activity and k_{cat}/K_M values can vary by several orders of magnitude. A XylP81 alignment (**Figure 14**) with other functionally characterized β-xylosidases from the GH39 family revealed that XylP81 contains the catalytic general acid/base (Glu102) and nucleophile (Glu225) predicted to be involved in catalysis of all GH39 β-xylosidases. Phylogenetic analysis of XylP81 (**Figure 15**) which included all the β-xylosidases from the GH39 family and the closets sequences homologs excluding the partial glycoside hydrolase from *Anaerolineaceae* bacterium. The tree showed that XylP81 formed a separate clade with all the uncharacterized protein sequences and is closely related to a β-xylosidases from *Anaerolineaceae* bacteria, members of the Chloroflexi phylum. Members of the Chloroflexi phylum are aerobic thermophiles and this phylum has been reported to be one of the most dominant phyla during the thermophilic composting stage, followed by *Proteobacteria* and *Actinobacteria* (Meng *et al*., 2019). This suggests that XylP81 could be a newly GH39 β-xylosidase which is probably derived from a *Chloroflexi*-like bacterium present during the thermophilic stage of composting.

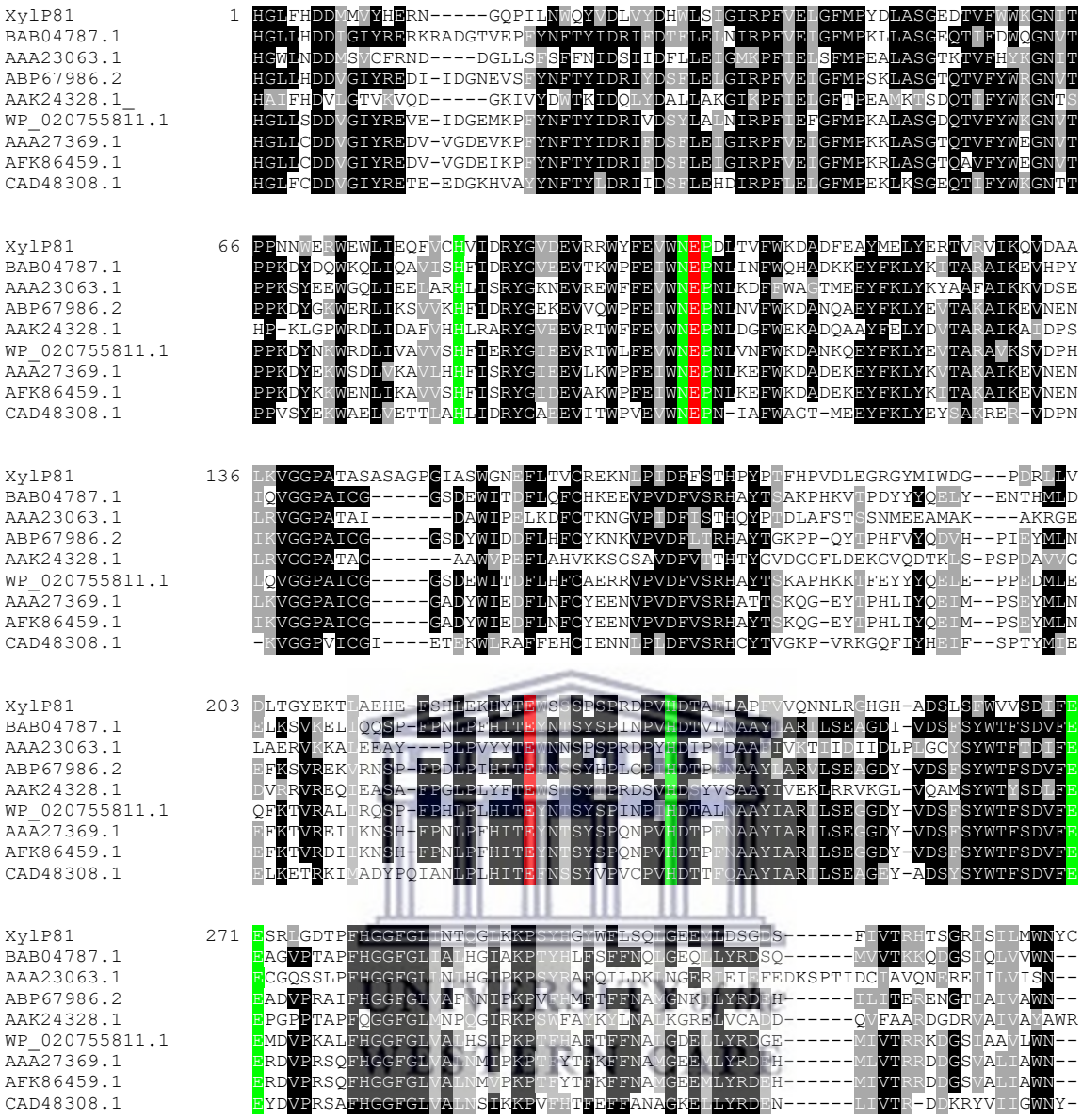

Figure 14: Multiple sequence alignment of the XylP81 protein with other GH39 β-xylosidases from the CAZy database. The sequences included; β-xylosidase from *Bacillus halodurans* C-125 (BAB04787.1), β-xylosidase from *Caldicellulosiruptor saccharolyticus* (AAA23063.1), xylan-1,4- β-xylosidase from *Caldicellulosiruptor saccharolyticus* DSM 8903 (ABP67986.2), β-xylosidase from *Caulobacter vibrioides* CB15 (AAK24321), β-xylosidase from *Geobacillus* sp. (WP_020755811.1), β-xylosidase from *Thermoanaerobacterium saccharolyticum* DSM 7060 (AAA27369.1), GH39 family protein from *Thermoanaerobacterium saccharolyticum* JW/SL-YS485 (AFK86459.1) and β-xylosidase from *Thermoclostridium stercorarium* (CAD48308.1). The black and grey shaded areas represent a 70% sequence identity and similarity amongst the proteins. The conserved catalytic amino acid residues E103 (general acid/base) and E225 (nucleophile) are shaded red colour and the green colour represents amino acid residues that are important in β-xylosidase activity.

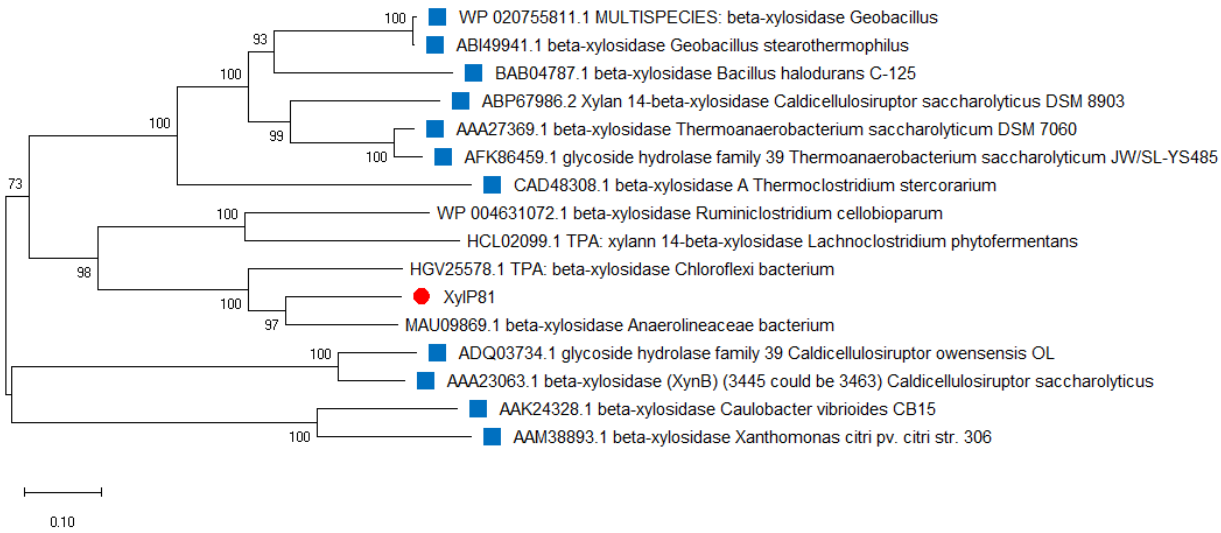

Figure 15: Evolutionary analyses and phylogenetic tree of XylP81 with other GH39 β-xylosidases from the CAZy database and the closest sequence homologs from a BLASTp search. The tree was constructed in MEGA7 using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. XylP81 is represented with the red dot and the blue squares represent all the characterized GH39 xylosidases.

3.1.4 BglP89 sequence analysis

The deduced amino acid sequence of BglP89 showed highest sequence identity with β-glucosidase protein sequences (**Table 9**) from the BLASTp analysis against the Genbank database, none of which are characterized. BglP89 is predicted to belong to the GH3 family according to Pfam and dbCAN2 analysis. The GH3 family is the largest GH family with about 300 functionally characterized enzymes with diverse enzyme activities which include; β-D-glucosidase (EC 3.2.1.21), β-D-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (3.2.1.55), glucan 1,3-β-glucosidase (EC 3.2.1.58), glucan 1,4-β-glucosidase (EC 3.2.21.75) and exo-1,3 / 1,4-glucanase (EC 3.2.1.1) (http://www.cazy.org/GH3.html). As with GH43 family members, several enzymes in the GH3 family have a bi or multi-functional activity. Further sequence analysis revealed that BglP89 is a three-domain β-glucosidase with an N-terminal domain of the GH3 family (Pfam: PF00933), C-terminal domain of GH3 family (Pfam PF101915), and Fibronectin-III-like domain (Pfam PRK1098; β-glucoside glucohydrolase). The function of the third domain is unclear as to date. This domain has

however been implicated to be involved in thermostability of GH3 β-glucosidases (Méndez-Líter *et al*., 2017) and similar to BglP89 the Fn-III like domain has been reported in other bacterial extracellular GH3 β-glucosidases (Li *et al*., 2018; Justo *et al*., 2015; Alves *et al*.,2018; Bergmann *et al*., 2014).

Table 9: Closest sequence homologs of BglP89 from a BLASTp search.

Organism	Protein name	Query coverage	% Identity	Accession number
Pontibacter korlensis	β-glucosidase BglX	99%	67.9	WP_148561686.1
Flammeovirgacea e bacterium 311	β-glucosidase BglX	99%	67.4	WP_081786901.1
Pontibacter spHB172049	β-glucosidase BglX	96%	70.2	WP_140621426.1
Rufibacter roseus	β-glucosidase BglX	95%	70.1	WP_066615335.1
Hymenobacter sp IS2118	β-glucosidase BglX	100%	66.45	WP_035564479.1

The alignment of BglP89 with other functionally characterized GH3 β-glucosidases from the CAZy database revealed that BglP89 contains the catalytic nucleophile (Asp292) catalytic base or acid (Glu724). These catalytic amino acid residues are conserved amongst GH3. The GH3 β-glucosidases usually have two major conserved motifs: the KHF which is proposed to be a proton donor and the SDW motif at the active site (Withers *et al*., 1990; (**Figure 16**). However, the tryptophan (W293) non-polar aromatic amino acid in GH3 β-glucosidases is replaced by a polar aromatic amino acid tyrosine (Y293) in the BglP89. Phylogenetic analysis of Bgl89 with other GH3 β-glucosidase and the five closest sequence homologs showed that BglP89 is possibly a novel β-glucosidase that forms a clade with uncharacterized β-glucosidases belonging to Bacteroidetes phylum reported to be the most dominant in the library used in this study (**Figure 17**).

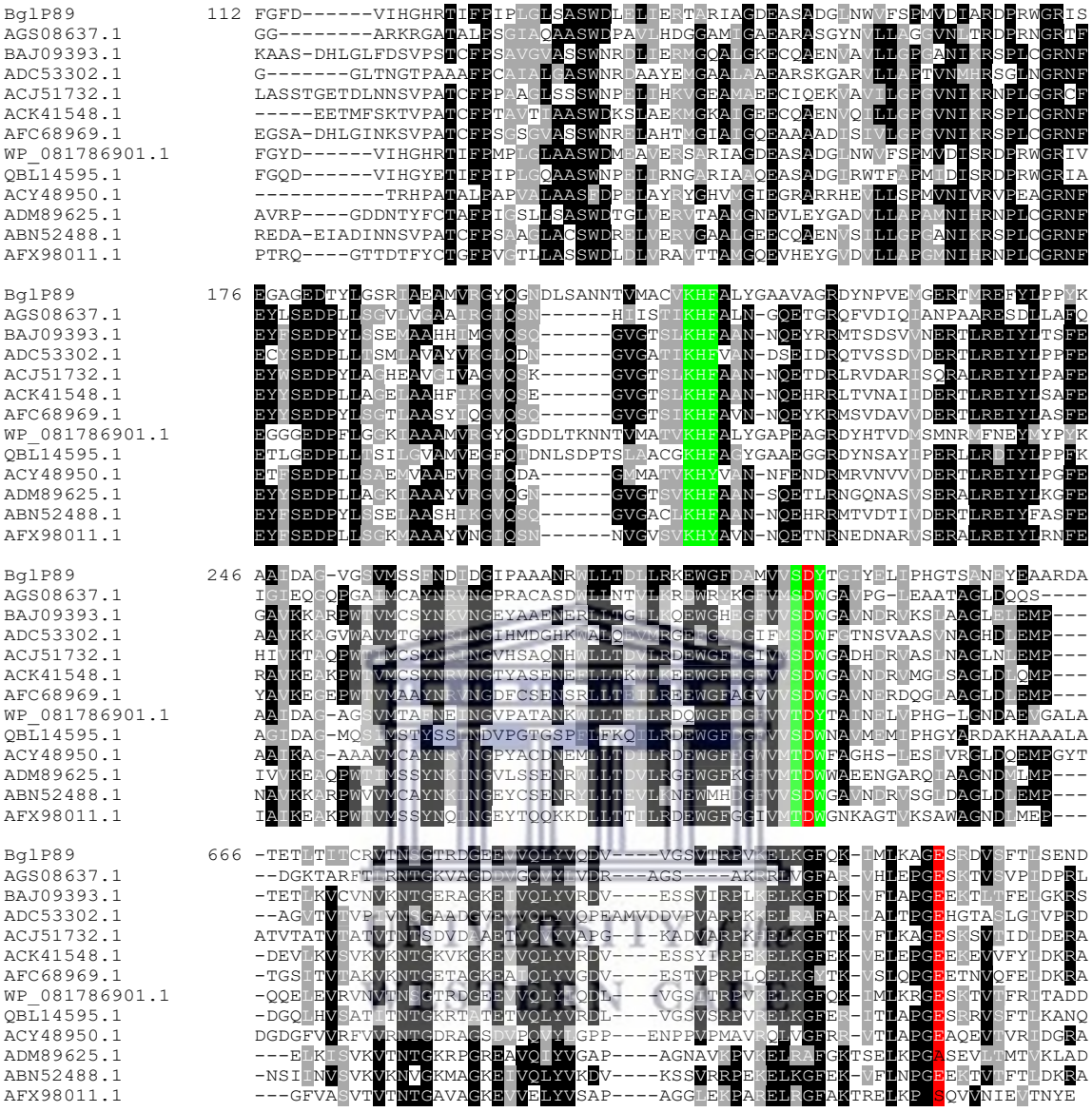

Figure 16: Multiple sequence alignment of the BglP89 protein sequence with other GH3 β-glucosidases. These sequences included; β-glucosidase from *Novosphingobium* sp. GX9 (AGS08637.1), β-glucosidase from *Paenibacillus relidesesami* (BAJ09393.1), β-glucosidase from *Martelella mediterranea* (ADC53302.1), β-glucosidase from *Bifidobacterium longum* subsp. infantis ATCC 15697 (ACJ51732.1), GH3 domain protein from *Dictyoglomus turgidum* DSM 6724 (ACK41548.1), β-glucosidase from *Paenibacillus xylanilyticus* (AFC689669.1), β-glucosidase BglX from *Flammeovirgaceae* bacterium 311 (WP_081786901.1), β-glucosidase from *Microbulbifer thermotolerans* (QBL14595.1), GH3 domain protein from *Rhodothermus marinus* DSM 4252 (ACY48950.1), β-glucosidase from an uncultured bacterium (ADM89625.1), GH3 domain protein from *Hungateiclostridium thermocellum* ATCC 27405 (ABN52488.1), and β-glucosidase from an uncultured microorganism (AFX98011.1). The black and grey shaded areas represent a 70% sequence identity and similarity amongst proteins. The predicted and conserved catalytic site residues D292 and E721 are shaded in red and substrate recognition amino acid residues are shaded green.

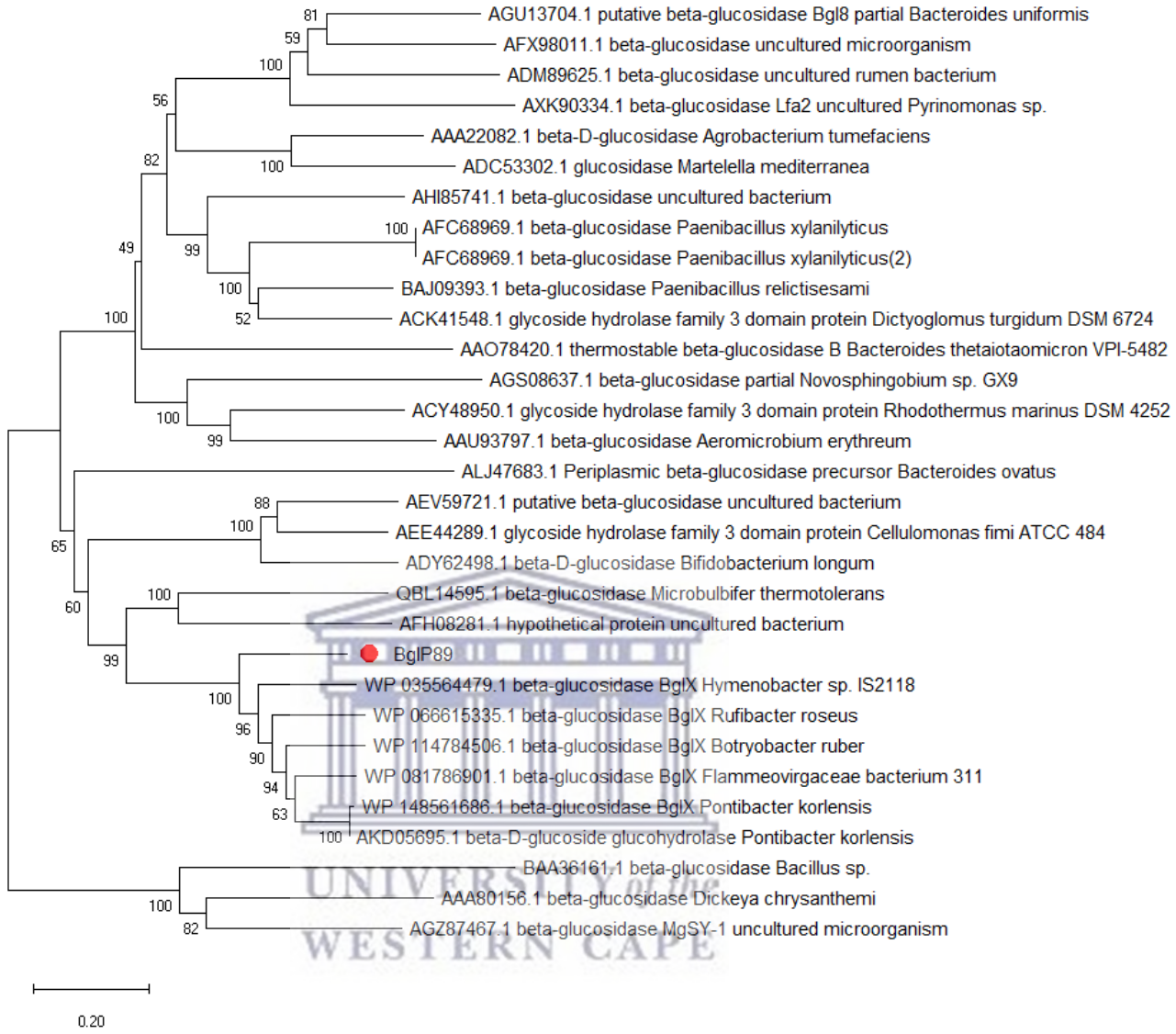

Figure 17: Evolutionary analyses and phylogenetic tree of BglP89 indicated by a red dot with other GH3 β-glucosidase enzymes from the CAZy database constructed in MEGA7 using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches.

3.2 Cloning GH encoding genes

The fosmid DNA was extracted from metagenomic clones and used as template DNA for PCR amplification. The genes encoding putative GHs in the P55E4 and P89A4 fosmids were amplified using high-fidelity DNA polymerase with gene-specific primers designed with incorporation of appropriate restriction enzyme sites NheI and XhoI for *xyl*P55 and NdeI and XhoI for *bgl*P89. PCR products of ~1,1 kb corresponding to the size of the *xyl*P55 gene, and ~2,3 kb corresponding to the size of the *bgl*P89 gene were observed by agarose gel electrophoresis (**Figure 18**).

Figure 18: Agarose gel (0.8% w/v) electrophoresis analysis showing *xylP55* and *bglP89* PCR amplicons from fosmid clones P55E4 and P89A4 respectively. Lane M: 1kb DNA marker, lane 1: negative control, lane 2: 1.1 kb xylP55 amplicon, and lane 3: 2.3 kb bglP89 amplicon.

The amplification products were purified and cloned into a pJET1.2/blunt vector and the ligations were used to transform chemically competent *E. coli* JM109 cells. Transformants harbouring the resulting constructs, pJET-*xyl*P55 and pJET-*bgl*P89, were selected and the DNA insert sizes were confirmed through restriction enzyme digestion using NheI and XhoI for pJET-*xylP55* and NdeI and XhoI for and pJET-*bglP89*. All four clones showed two distinct bands of 2.9 kb (pJET1.2/blunt vector backbone) and a band corresponding to either *xylP55* (~1,1 kb) *or bglP89* (~2,3 kb) (**Figure 19**).

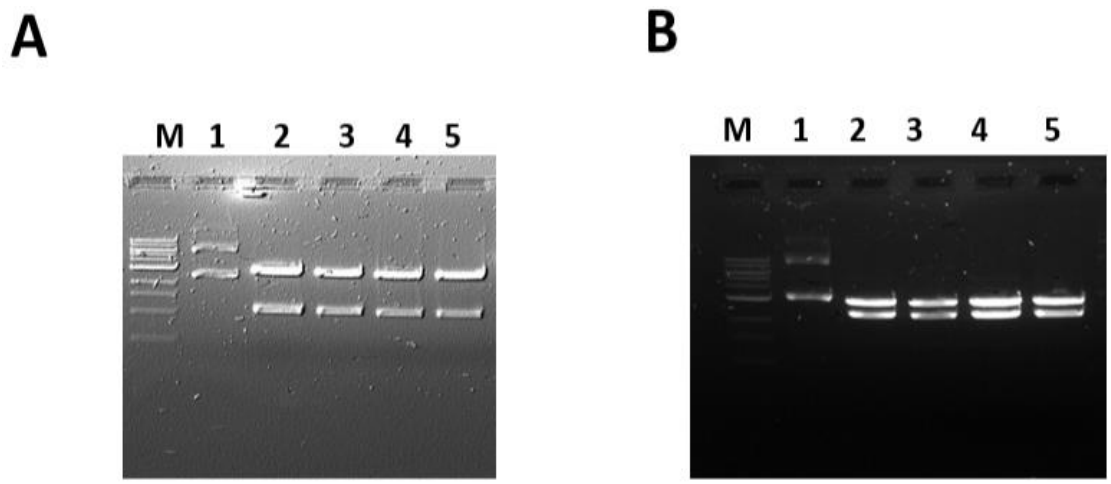

Figure 19: Agarose gel electrophoresis analysis of the double digest restriction of the recombinant pJET1.2/blunt constructs pJET-*xyl*P55 and pJET-*xyl*P89. A) Lane M, 1kb DNA marker ; lane 1, undigested pJET-*xyl*P55; lane 2-5 pJET-*xyl*P55 digested with NheI and XhoI. B) Lane M, 1kb DNA marker; lane 1 undigested pJET-*xyl*P89; lane 2-5 pJET-*xyl*P89 digested with NdeI and XhoI.

The *xyl*P55 and *bgl*P89 fragments excised from the pJET constructs were successfully cloned into a linearized pET21a(+) expression vector under transcriptional control of the strong T7 promoter to produce a C-terminal histidine-tagged sequence for downstream purification. The recombinant plasmids pET21-XylP55, pET21-XylP81 (synthesized by Biomatik), and pET21-BglP89 were transformed into two different *E. coli* hosts, BL21(DE3) and BL21(DE3)-Rosetta pLysS. The resultant pET21 constructs were confirmed as above and results are shown in (**Figure 20**).

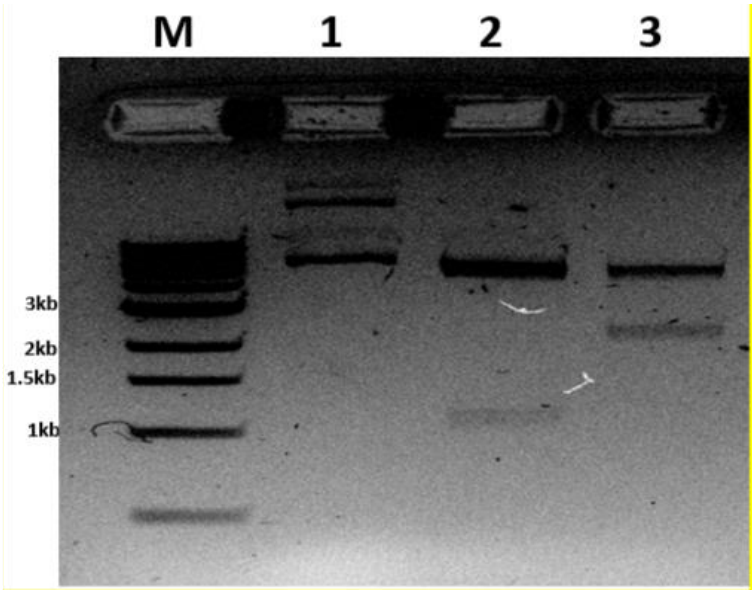

Figure 20: Agarose gel analysis of the restriction digest analysis of pET21a(+) constructs; pET21-XylP55 and pET21-BglP89. Lane M: 1kB DNA ladder, lane 1: Undigested pET21a(+), lane 2: NheI and XhoI digested pET21-XylP55 and lane 3: NdeI and XhoI digested pET21-BglP89.

3.3 Protein expression and purification

3.3.1 Protein expression

Protein expression was conducted at either 28°C or 37°C overnight for all the genes in two different *E. coli* hosts BL21(DE3) and BL21(DE3)-Rosetta pLysS. The XylP55 and XylP81 enzymes expressed in the soluble fraction when using the BL21(DE3)-Rosetta pLysS strain at 37°C (**Figure 21 A and C)**, however, the majority of XylP55 protein was expressed in the insoluble fraction. BglP89 was expressed mostly in the insoluble fraction under these conditions in both *E. coli* strains, and expression was improved by inducing at a lower temperature (28°C) as this has been shown to result in a higher yield of soluble protein expression (Vera *et al.*, 2007). BglP89 showed qualitatively higher expression in both the soluble and insoluble fractions at 28°C when using the BL21(DE3)-Rosetta pLysS (**Figure 21 B**). No expression was observed when using the *E. coli* BL21(DE3) as an expression host suggesting that the BL21(DE3)-Rosetta pLysS *E. coli* strain is a better expression host for BglP89 expression. The BL21(DE3)-Rosetta pLysS strain does enhance protein expression level because it carries a chloramphenicol-resistant plasmid called pRARE which encodes for several rare codon tRNAs that have lower abundance in *E. coli* (Kopanic *et al.*, 2013).

Preliminary enzyme activity testing of the crude extract showed that all the proteins were active on the assayed chromogenic *p*NP-linked substrates as a yellow colour was observed indicating the release of *p*-nitrophenol after hydrolysis.

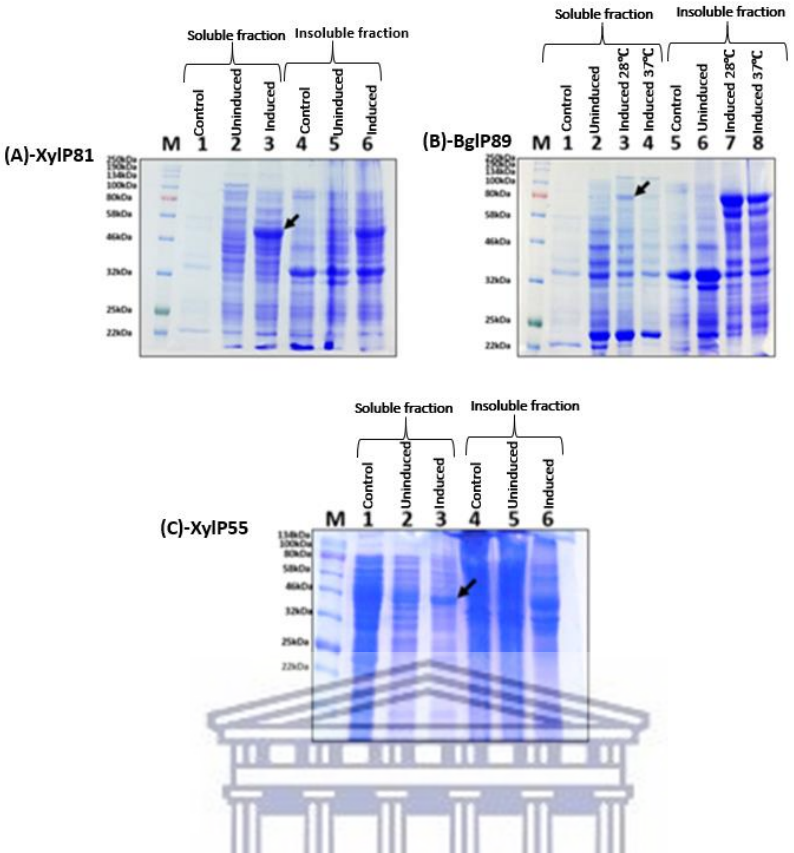

Figure 21: SDS-PAGE analysis of XylP81,XylP55 and BglP89 expression in *E. coli* Rosetta pLysS with 0.5mM IPTG. **A)** XylP81 protein expression at 37ºC. Lane M: ColorPlus prestained protein ladder, broad range (10-230 kDa); lane 1: *E. coli*-pET21a(+) no insert uninduced; lane 2: *E. coli*-pET21a-XylP81, uninduced soluble fraction; lane 3: *E. coli*-pET21a-XylP81 induced soluble fraction; lane 4: *E. coli*-pET21a(+) no insert induced; lane 5: *E. coli*-pET21a-XylP81 uninduced insoluble fraction; lane 6: *E. coli*-pET21a-XylP81 induced insoluble fraction. **B)** BglP89 protein expression at 28ºC and 37ºC. Lane M: ColorPlus prestained protein ladder; lane 1: *E. coli*-pET21a vector no insert uninduced; lane 2: *E. coli*-pET21a-BglP89 uninduced soluble fraction; lane 3: *E. coli*-pET21a-BglP89 induced soluble fraction at 28ºC, lane 4: *E. coli*-pET21a-BglP89 induced soluble fraction at 37ºC; lane 5: E. coli-pET21a(+) no insert induced; lane 6: *E. coli*-pETP89 uninduced insoluble fraction; lane 7: *E. coli*-pETP89 induced insoluble fraction at 28ºC and lane 8: *E. coli*-pET21a-BglP89 induced insoluble fraction at 37ºC. **C)** XylP55 protein expression at 37ºC. Lane M: ColorPlus prestained protein ladder, broad range (10-230 kDa); lane 1: *E. coli*-pET21a(+) no insert uninduced; lane 2: *E. coli*- pET21a-XylP55 uninduced soluble fraction; lane 3: *E. coli*-pET21a-XylP55 induced soluble fraction; lane 4: *E. coli*-pET21a(+) no insert induced; lane 5: *E. coli*-pET21a-XylP55 uninduced insoluble fraction; lane 6: *E. coli*-pET21a-XylP55 induced insoluble fraction. In panels A, B, and C arrows indicate the expressed protein band.

3.3.2 Protein purification

The recombinant proteins XylP55, XylP81, and BglP89 were all expressed in the *E. coli* Rosetta pLysS strain under optimised conditions for each recombinant protein. The histidine-tagged proteins were purified using nickel-affinity chromatography and the purity was analysed using SDS-PAGE analysis. The proteins were eluted with 1M imidazole and a single distinct band of ~58 kDa was observed for XylP81 and ~84 kDa for BglP89 (**Figure 22 A)**, corresponding to the deduced theoretical molecular weights (53 kDa and 84 kDa, respectively. Purification of XylP55 showed two bands, one corresponding to the deduced molecular weight of XylP55 (42 kDa) and an extra band of a slightly smaller size (40 kDa) (**Figure 22 B**). Analysis of the XylP55 protein sequence using SignalP 5.0 showed a potential proteolytic cleavage site at 22 amino acids from the start of XylP55. The double bands observed for XylP55 could therefore be the result of processing of the secretion signal by the Sec-dependent translocation pathway in *E. coli*, resulting in a mass difference of 2.5kDa (first 22 amino acids of XylP55). It is customary for the secretion signal to be deleted when designing an expression construct. However, activity was observed in the protein with the signal sequence, and therefore preliminary biochemical characterisation of the XylP55 protein was pursued. The high amino acid similarities of XylP55 to the characterized GH43 β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome (Matsuzawa *et al*., 2015) indicated that they might perform similarly. Initial biochemical characterization of XylP55 (**Figure 23**) indicated that it has the following biochemical characteristics (pH_{opt} 7.0; T_{opt} 50℃; active on both *p*NPX and *p*NPA), which is almost identical to BAS02081.1 (pH_{opt} 7.5; T_{opt} 55℃; active on both *p*NPX and *p*NPA). This suggests that the expression of the full-length protein did not seem to impede the performance of the purified enzyme. Furthermore, the similarity in biochemical characteristics together with high amino acid similarity to a characterized enzyme implied that little new information would be gained toward a better understanding of protein structure-function relationships. It was therefore decided that XylP55 would not be characterized further.

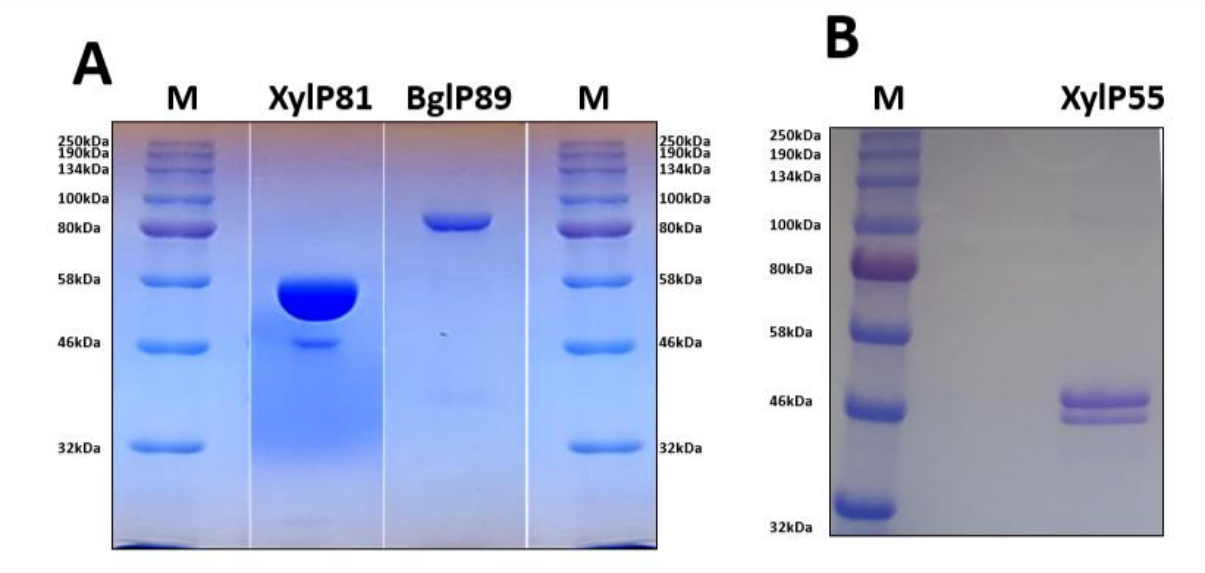

Figure 22: SDS-PAGE analysis of purified recombinant proteins, XylP81, BglP89, and XylP55 following nickel affinity chromatography purification. A) Lane M: ColorPlus prestained protein ladder, lane: 2 purified XylP81, lane: 3 purified BglPP89, and lane M: broad range protein marker. B) Lane M: ColorPlus prestained protein ladder and lane 1: purified XylP55.

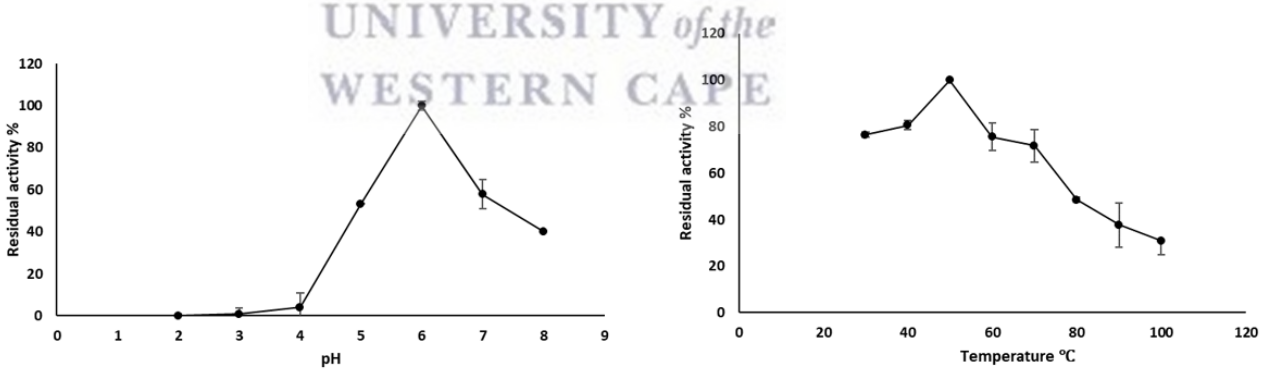

Figure 23: Initial biochemical characterization of XylP55 (pH and temperature optimum).

3.4 Biochemical characterization

3.4.1 pH/temperature optimum and thermostability

The pH and temperature optimum of the recombinant β-xylosidase Xyl81 and β-glucosidase BglP89 was determined using *p*NPX and *p*NPG chromogenic substrates, respectively. The substrates were selected based primarily on the sequence predictions. The enzymes were assayed at a pH range of (3-9; **Figure 24**). XylP81 showed optimum activity at pH 6 and maintained about 80% residual activity at pH 7 with a gradual decrease in activity as the pH increased. The XylP81 pH optimum is in the same range as other reported and functionally characterised GH39 β-xylosidase enzymes (**Table 11**). BglP89 also showed optimum activity at pH 6 and interestingly maintained about 80% activity at pH 7-8 with activity rapidly decreasing at pH 9. This suggests that both XylP81 and BglP89 are mostly active at neutral pH, consistent with other GHs isolated from a compost metagenome (Dougherty *et al*., 2012; Wagschal *et al*., 2009; Uchiyama *et al*., 2013).

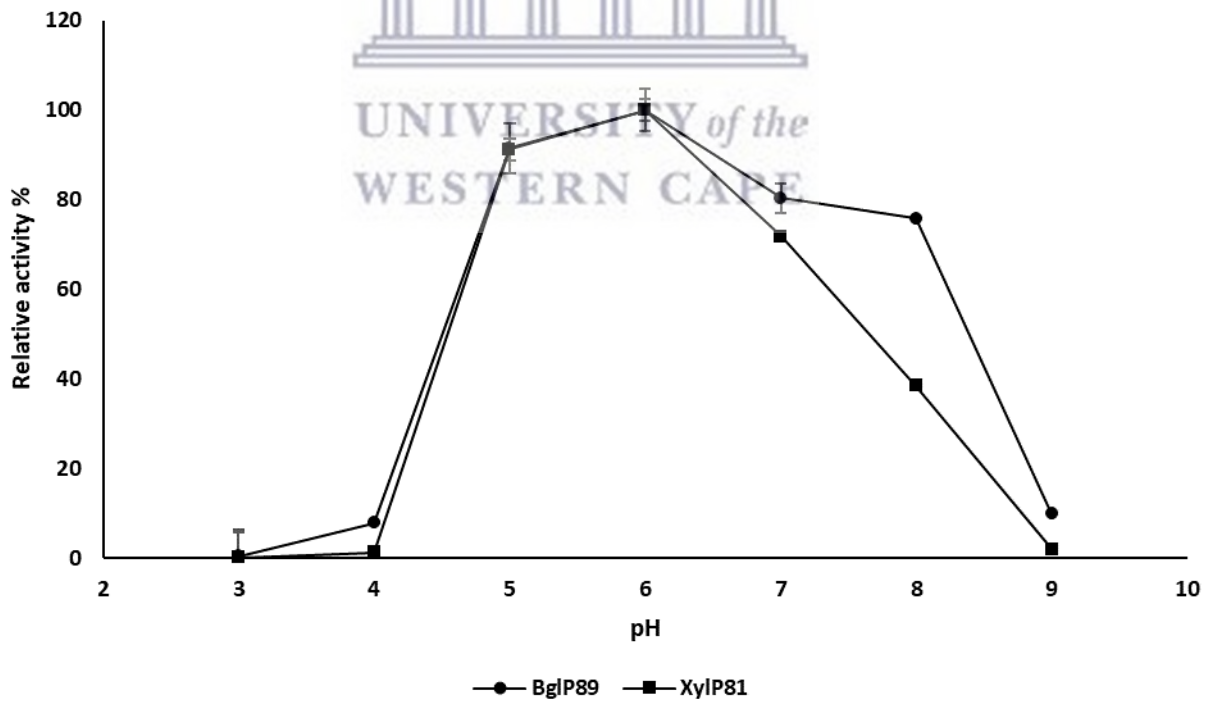

Figure 24: pH optimum of BglP89 and XylP81. Error bars represent the standard deviation of an average of three biological replicates.

The recombinant enzymes were assayed at various temperatures from 20-80℃ (**Figure 25**) at the optimum pH 6 and both XylP81 and BglP89 showed optimum activity at 50℃. XylP81 maintained about 80% activity at 60℃ with a rapid decrease inactivity thereafter. The optimum temperature of XylP81 is relatively low in contrast to some characterised GH39 β-xylosidases (**Table 11**). BglP89 lost 80% activity after 50℃ and β-glucosidases in the GH3 family (**Table 11**) showed they have varied temperature optimum depending on the source of the enzyme.

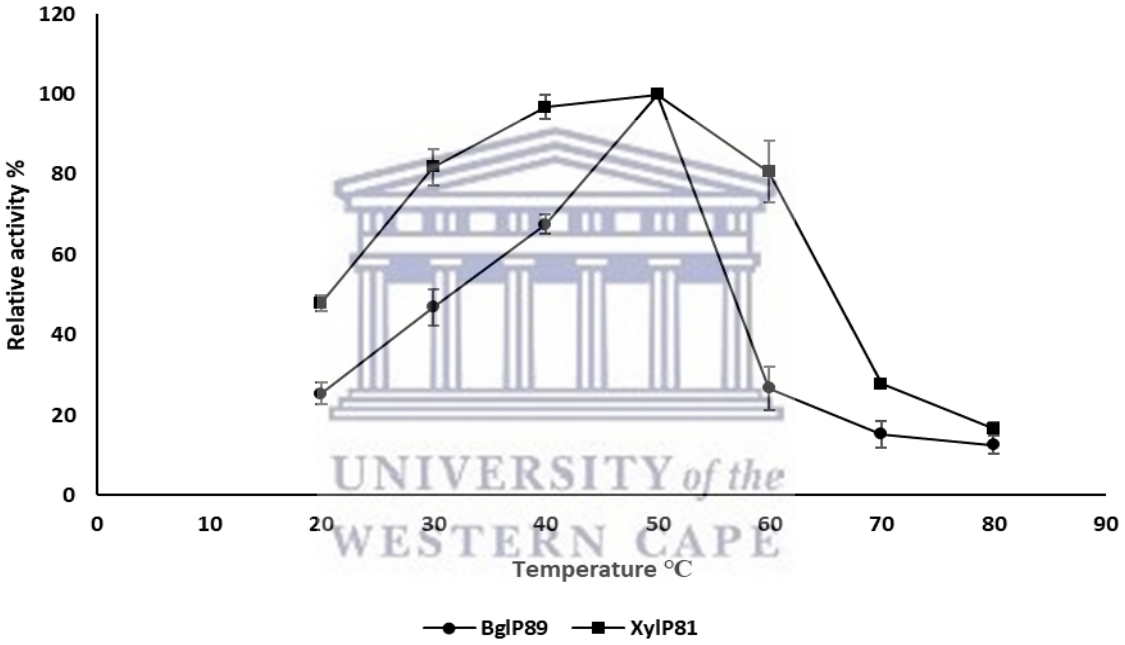

Figure 25: Temperature optimum of BglP89 and XylP81. Error bars represent the standard deviation of an average of three biological replicates.

Thermal stability is an important characteristic of enzymes as thermostable enzymes offer specific advantages for use in industry; therefore, the thermostability of the recombinant XylP81 and BglP89 was determined. These proteins were isolated from a metagenomic library constructed from thermophilic horse manure compost, an environment expected to harbour thermophilic microorganisms and thus thermostable enzymes. The recombinant enzymes were incubated at 50, 55, 60, and 70℃ for 1 hour, and stability was determined by measuring residual activity. Enzyme kept at 4ºC for the duration of the assay was assayed as control and

regarded as 100% residual enzyme activity. XylP81 showed good stability at 50℃ and 55℃ retaining 100% of its activity after incubation for 1 hour. At a higher temperature of 60℃, the enzyme lost about 80% activity after 1-hour incubation and lost 80% activity at 70℃ after just 10 minutes (**Figure 26**). Other functionally characterised GH39 β-xylosidases reported and regarded as thermostable are shown in **Table 11**. BglP89 showed high stability at 50 and 55℃ as it retained 100% of its activity after the 1 hour incubation but lost 50% activity after 1 hour at 60℃ (**Figure 27**). This suggests that XylP81 and BglP89 are produced by moderately thermophilic bacteria as they have good thermostability in the temperature range of 50-60℃. This correlates with the sequence analysis of XylP81 and BglP89 which showed that they share close similarity with sequences from bacterial species from the Chloroflexi and Bacteroidetes phyla which are reported as major phyla active in the composting process (Meng *et al*., 2019; Wang *et al.*, 2016). Chloroflexi is known to contain aerobic thermophiles whilst a variety of bacterial species in the Bacteroidetes phylum are reported to be slightly thermophilic (Albuquerque *et al*., 2018) explaining the moderate thermophilic nature of the XylP81 and BglP89.

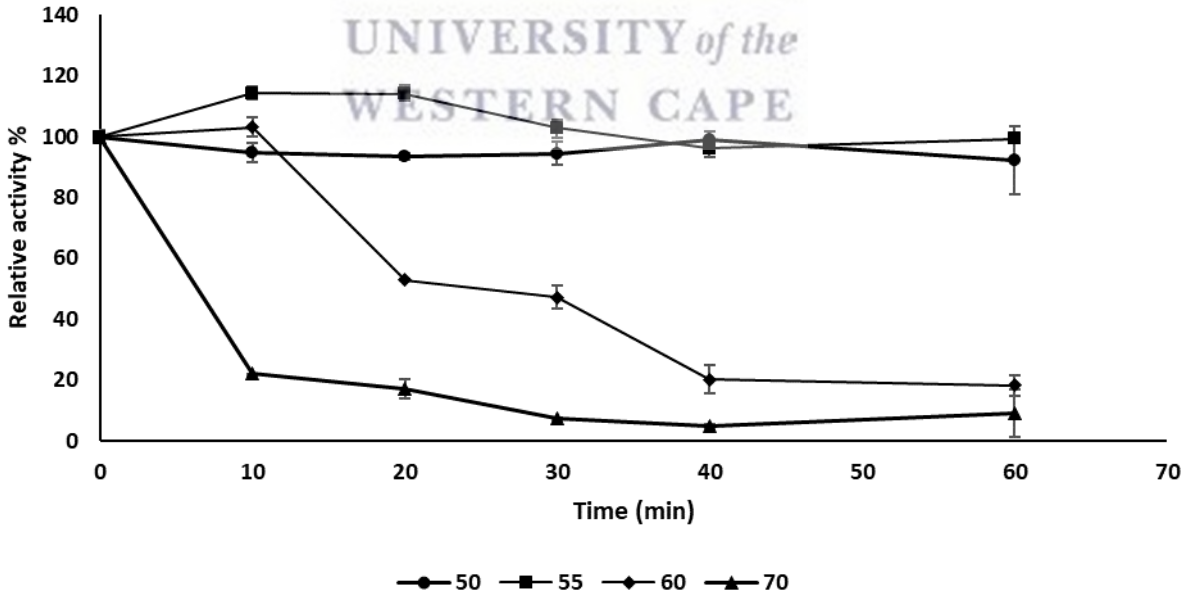

Figure 26: XylP81 thermostability at 50, 55, 60, and 70℃. Enzyme was incubated for 1 hour at different temperatures without the substrate before assaying residual activity. Residual activity at 100% is the unincubated enzyme. Error bars represent the standard deviation of the average of three biological replicates.

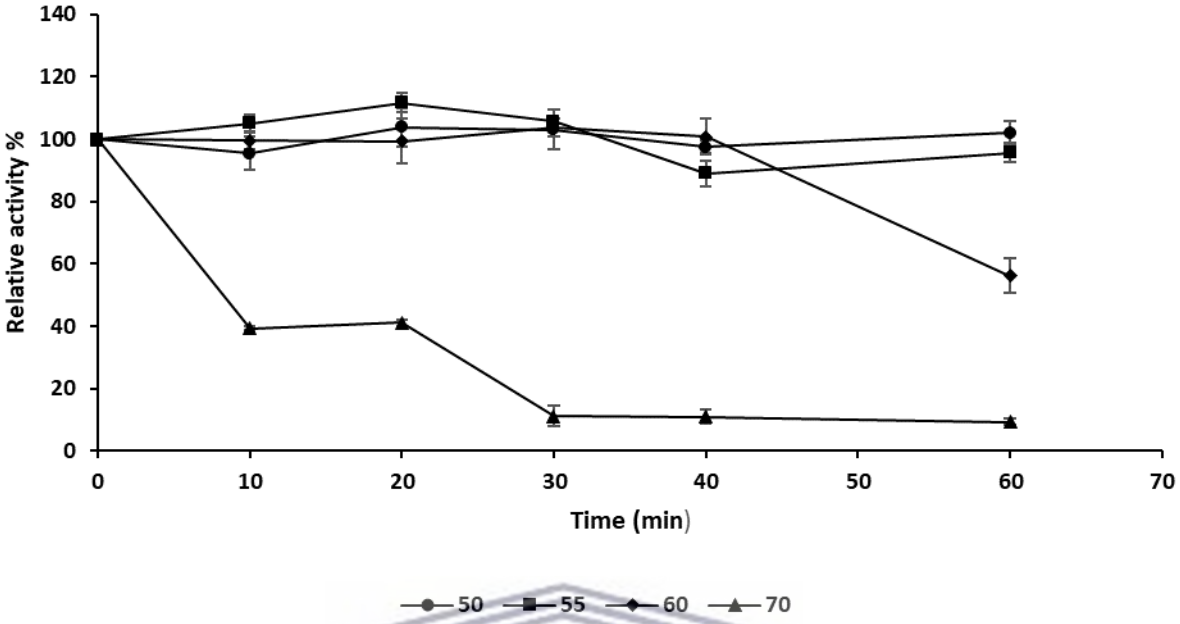

Figure 27: BglP89 thermostability at 50, 55, 60, and 70°C. Enzyme was incubated for 1 hour at different temperatures without substrate before assaying residual activity. Residual activity at 100% is the unincubated enzyme. Error bars represent the standard deviation of the average of three biological replicates.

3.4.2 Effect of metal ions on purified XylP81 and BglP89

The effect of metal ions (Ag^{+}, Ca^{2+}, Cu^{2+}, Fe^{3+}, K^{+}, Mg^{2+}, $Mn^{2+,\ Na+}$, Ni^{2+}, and Zn^{2+}) was tested on the activity of XylP81 and BglP89. The assays were performed in the presence of 5 mM of each metal. The activity recorded for an assay without the addition of metal ion was considered 100%, and relative activities were calculated against it (**Table 10**). The effect of metal ions was tested to determine if the recombinant proteins have a metal-ion binding loop or site that would potentially increase their activity or if they are metal independent. XylP81 was slightly activated by Mg^{2+} and Mn^{2+} to 105% and 117% respectively and was significantly inhibited by Ag^{+} (43%) and Cu^{2+} (67%). Similarly, a GH39 β-xylosidase from *Geobacillus* sp was reported to lose 73% of its activity in the presence of Cu^{2+} (Bhalla *et al*., 2014). The tested metal ions had no significant effect on BglP89 enzyme activity with Mn^{2+} showing slight activation. This indicated that BglP89 activity doesn't necessarily require the addition of metal ions as cofactors

during hydrolysis. Similarly, Mn^{2+} has been reported to enhance β-glucosidase in *Thermoanaerobacterium thermosaccharolyticum* (Pei *et al*., 2012).

Table 10: The effect of various metal ions on the activity of XylP81 and BglP89 using *p*NPX and *p*NPG substrates, respectively.

Metal ion (5mM)	Relative activity (%)	
	XylP81	BglP89
Control	100	100
Ag^{+}	52.6±3.3	98.1±0.78
Ca^{2+}	94.7±3.9	106.1±4.9
Cu^{2+}	43±3.8	96.3±1.6
Fe^{3+}	99±2.7	100.2±0.6
K^{+}	98.5±3.01	99.9±2.2
Mg^{2+}	105.1±4.48	104±3.7
Mn^{2+}	117.5±4.7	119.5±3.1
Na^{+}	102.5±4.8	98.4±4.6
Ni^{2+}	94.7±6.7	106.4±3.9
Zn^{2+}	71±5.1	99.8±2.5

Note: XylP81 and BglP89 activity without the addition of metal ions was used as 100%.

13.4.3 Enzyme kinetics

The substrates *p*NPX and *p*NPG were used to determine the kinetic parameters (K_M and V_{max}) for XylP81 and BglP89, respectively, from the Michaelis-Menten (MM) plots (**Figure 28 and Figure 29**). The K_M and V_{max} of XylP81 were 5.3 mM and 0.43 µmol/min^{-1} respectively. The K_M value of XylP81 is within the same range (1.66 mM – 28 mM) of other functionally characterized GH39 β-xylosidases in (**Table 11**). The calculated k_{cat} was 107 s^{-1} and k_{cat}/ K_M was 20.3 $s^{-1}mM^{-1}$. The k_{cat} value for XylP81 was within the (5.26 s^{-1}- 189.29 s^{-1}) range of the GH39 β-xylosidases present in (**Table 11**) and is higher compared to β-xylosidases from

Dictoglomus thermophilum and *Bacillus halodurans* C-125. BglP89 showed a K_M value of 8.4 mM and V_{max} of 0.55 µmol/min^{-1}, the K_M value of BglP89 is significantly higher than most GH3 β-glucosidases (**Table 11**). Furthermore, the k_{cat} and k_{cat}/ K_M of BglP89 were 186.9 s^{-1} and 22 $s^{-1}mM^{-1}$, respectively. The comparison of BglP89 with other β-glucosidases presented in the table showed that it has the second highest turnover number k_{cat}. The higher and apparent K_M values for XylP81 and BglP89 showed that the enzymes have a lower affinity for the artificial *p*-nitrophenol substrates (*p*NPX and *p*NPG) although they show high k_{cat} values. The kinetic parameters of enzymes depend on a specific substrate, as well as the concentration used, temperature, and pH values. Comparison of the enzyme parameters of XylP81 and BglP89 with other GH39 β-xylosidases and GH3 β-glucosidases shows that β-xylosidases and β-glucosidases from different microorganisms have different kinetic parameters, justifying continued biodiscovery efforts.

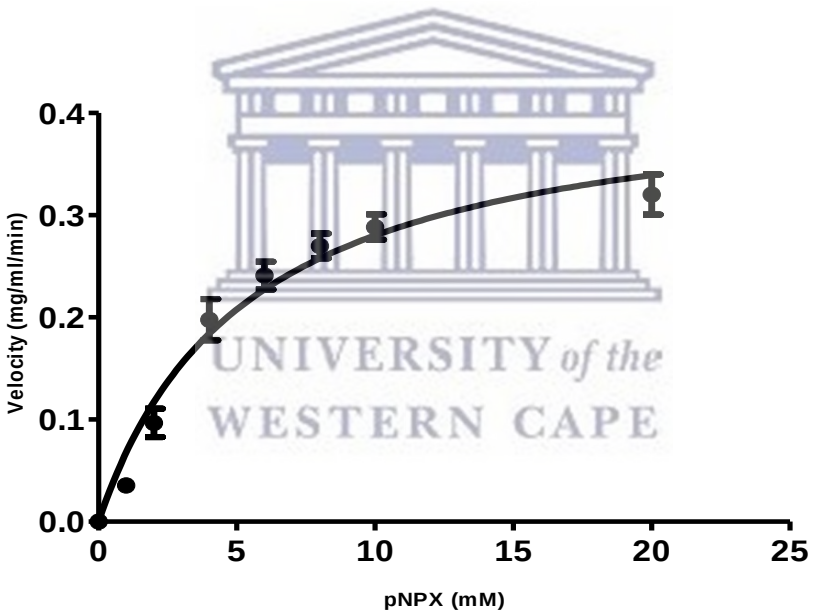

Figure 28: Michaelis-Menten nonlinear regression plot for XylP81 using pNPX as substrate**.** The graph was generated using GraphPad Prism 5.0 (GraphPad Software, San Diego, USA). Error bars represent the standard deviation of the average of three biological replicates.

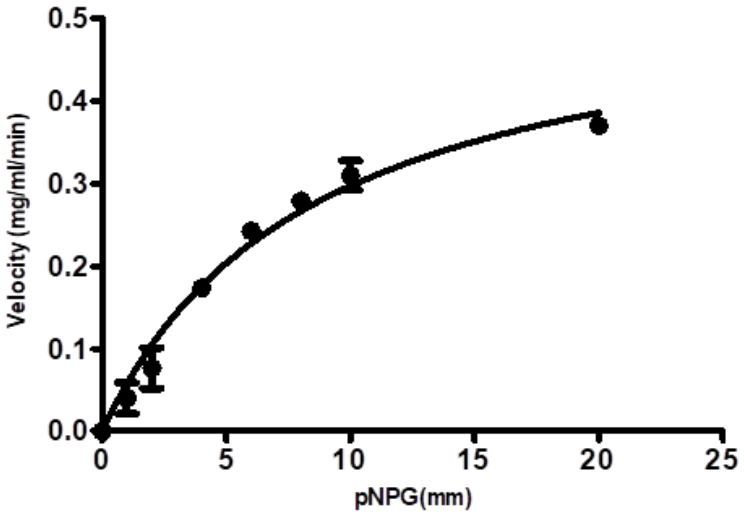

Figure 29: Michaelis-Menten nonlinear regression plot for BglP89 using pNPG as substrate. The graph was generated using GraphPad Prism 5.0 (GraphPad Software, San Diego, USA). Error bars represent the standard deviation of the average of three biological replicates.

Table 11: Comparison of XylP81 and BglP89 biochemical properties with GH39 family β xylosidase and GH3 family β-glucosidase from literature.

Enzyme and Source	Family	pH optimum	Temperature optimum	Thermostability	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($s^{-1}mM^{-1}$)	Xylose/ Glucose inhibition	Reference
XylP81	**GH39**	**6**	**50°C**	**100% activity at 55°C**	**5.3**	**107**	**20.3**	**42% at 3M**	**This study**
Dictoglomus thermophilum (β-xylosidase)	GH39	6	75°C	60% activity at 85°C for 120 min	1.66	76.7	56.24	60% at 3M	Li *et al*., 2018
Geobacillus sp. WSCUF1 (β-xylosidase)	GH39	6.5	70°C	50% activity at 70°C for 9 days	2.8	128.57	45.9	50% at 300 mM	Bhalla *et al*., 2014
Thermoanaerobacterium saccharolyticum JW/SLY (β-xylosidase)	GH39	6	65°C	50% activity at 67°C for 60min	28	189.292	6.76	70% at 200 mM	Shao *et al*., 2011
Bacillus halodurans C-125 (β-xylosidase)	GH39	7.5	55°C	50% activity at 60°C for 144 min	8.6	5.26	$6x10^5$	50% at 300 mM	Smaali *et al*., 2006
Caulobacter crescentus (β-xylosidase)	GH39	6	55°C	50% activity at 50°C for 40 min	9.3	-	-	ND	Corrêa *et al*., 2012
BglP89	**GH3**	**6**	**50°C**	**50% activity at 60°C for 40 min**	**8.49**	**186.9**	**22**	**46% at 150 mM**	**This study**
Rumen cattle feeding on Miscanthus sinensis (β-glucosidase)	GH3	5	38°C	90% activity at 30°C for 60 min	0.309 mmol/L	-	-	20% at 650 mM	Li *et al*., 2014
Yak rumen metagenome (RuBG3A β-glucosidase)	GH3	4.6	40°C	50% activity at 45°C for 60 min	1.06	126.6	$11.94x10^4$	ND	Bao *et al*., 2012
Microbulbifer thermotolarance (β-glucosidase)	GH3	7	60°C	ND	0.29	0.50	1.724	59% at 10 mM	Pyeon *et al*., 2019
Amazon soil metagenome (AmBGL17, β-glucosidase)	GH3	6	45°C	50% activity at 45°C for 120 min	0.30	38.5	$12.86x10^4$	ND	Bergmann *et al*., 2014

Bifidobacterium adolescentis (BaBgl, β-glucosidase)	GH3	6	45°C	ND	1.1	94	87	50% at 67 mM	Michlmayr *et al*., 2015
Metagenomic derived (rMlBgl, β-glucosidase)	GH3	7	40°C	50% activity at 35°C	0.69	581.52	842.78	ND	Mai *et al*., 2016
Ruminal ($GlyA_1$ β-glucosidase)	GH3	6.5	55°C	ND	10.7	1.63	$15.23x10^4$	ND	Mercedes *et al*.,2016
Brazilian forest soil (β-glucosidase)	GH3	5.5	50°C	50% activity at 40-55°C	0.76	13.2	$17.4x10^6$	50% at 300 mM	Alves *et al*., 2018
Thermotoga thermarum (β-glucosidase)	GH3	5	80°C	80% activity at 80°C	2,41	8.25	3.42	50% at 500 mM	Long *et al*., 2016

The biochemical characteristics of GH39 β-xylosidases and GH3 β-glucosidase in this table were determined using *p*NPX and *p*NPG substrates respectively

3.5 Substrate specificity

The substrate specificity of the purified recombinant XylP81 and BglP89 enzymes was tested on various chromogenic *p*-nitrophenol substrates, although not an exhaustive range, and natural substrates (**Table 12**). XylP81 showed broad substrate specificity with the highest activity against *p*NPX (122 U/mg) with some activity against *p*NPA whilst showing low but detectable activity against *p*NPG and *p*NPGal. XylP81 could also, therefore, be considered a bifunctional enzyme with both β-xylosidase/α-L-arabinofuranosidase activities. This dual activity is observed in many other β-xylosidases and is likely due to the spatial similarity of sugar ring hydroxyl groups in D-xylopyranose and L-arabinofuranose sugars (Wagschal *et al*., 2009). This dual or bifunctional activity is known and mostly reported for β-xylosidases from the GH3 and GH43 families, and not yet reported for the GH39 family. Furthermore, XylP81 was tested against more complex (longer chains) natural substrates which included beechwood xylan, birchwood xylan, arabinan (sugar beet), and arabinoxylan (rye). It unexpectedly displayed very little to no activity towards beechwood and birchwood xylan, which could be an indication of low β-xylanase activity. A GH39 β-xylosidase from a fungal strain *Orpinomyces* sp. has β-xylanase activity of 10.8 U/mg on beechwood xylan (Morrison *et al*., 2016). No activity was detected on arabinan (sugar beet) and arabinoxylan (rye). Xylan from beechwood and birchwood is reported to have a β-1,4-linked xylose backbone substituted with α-1,2 linked 4-*O*-methylglucuronic acid (Dodd and Cann 2009). The presence of these sidechains affects the accessibility of the main xylan chain, and therefore the efficient and complete degradation of xylan requires the action of multiple enzymes. Although XylP81 showed activity on *p*NPA the DNS assay indicated that it had no activity against arabinan which is an α-1,5 linked arabinofuranosyl backbone with α-1,2 and α-1,3 linked- arabinofuranosyl side chains, and also no activity on arabinoxylan, which is composed of a β-1,4 xylose backbone with α-1,2 or α-1,3 linked arabinofuranosyl side chains. Although the GH39 family is reported to consist of enzymes with β-xylosidase and α-L-iduronidase activities on the CAZy database (http://www.cazy.org/GH39.html), XylP81 showed promising broad substrate activity with α-L-arabinofuranosidase, β-galactosidase, and low β-glucosidase and traces of low β-xylanase activity which then increases the known activities for this family. XylP81 showed the highest β-xylosidase activity compared to the majority of GH39 β-xylosidases; except for a GH39 β-xylosidases from *Caulobacter crescentus* (215 U/mg; Corrêa *et al*., 2012) and *Geobacillus* sp. (133 U/mg) for the *p*NPX substrate.

BglP89 showed the highest specific activity against *p*NPG (133.5 U/mg) with significant activity against *p*NPA and *p*NPGal and low activity on *p*NPX. This indicates that BglP89 is a multifunctional enzyme and sequence analysis suggested that BglP89 belongs to the GH3 family, known to represent enzymes that are bi- or multifunctional, hence it was selected for characterization in this study. BglP89 also showed low but detectable activity against cellobiose, lichenan (moss), and β-glucan (barley), substrates which a mixed-orientation of β-1,3 and β-1,4 glycosidic bonds. This indicated that BglP89 has a low β-1,3-1,4-exoglucanase activity and has previously been reported for GH3 family enzymes. Furthermore, BglP89 showed low activity on Avicel PH-10 microcrystalline cellulose and CMC amorphous cellulose. Avicel and CMC mainly contain β-1,4 glycosidic bonds but due to the complexity of their structural composition the cellulose backbone or the main chain becomes less accessible **(**Sørensen *et al*., 2013). Therefore β-glucosidases are not expected to have activity on these substrates. β-glucosidase enzymes are classified into three major groups: i) the aryl β-glucosidase, with high affinity for aryl β-glucoside, for example, chromogenic *p*NP linked substrates; ii) cellobiase β-glucosides, with an affinity for only oligosaccharides, for example, cellobiose; and iii) broad specificity β-glucosidases, which have an affinity for both cellobiose and aryl-β-glucoside substrates (Bao *et al*., 2012). BglP89 demonstrates both aryl-β-glucosidase and cellobiase activity and could, therefore, be considered a broad specificity β-glucosidase.

Table 12: Specific activity of XylP81 and BglP89 on various substrates.

Substrate(s)	XylP81 (U/mg)	BglP89 (U/mg)
*p*NPX	122±4.5	6.2±0.2
*p*NPG	1,1±1.16	133.5±0.32
*p*NPGal	3.6±0.95	25.7±0.05
*p*NPA	17.4±0.18	23.7±0.06
Arabinan (sugar beet)	0.0	-
Arabinoxylan (rye)	0.0	-
Beechwood xylan	0.0	-
Birchwood xylan	0.0	-
Avicel	-	0.62±0.31
Cellobiose (+)	-	1.3±0.02
CMC	-	0.36±0.11
β-Glucan (Barley)	-	1.53±0.14
Lichenan (Moss)	-	1.98±0.32

3.5.1 TLC hydrolytic products analysis

The hydrolytic products formed by recombinant BglP89 on complex natural polysaccharides which include Avicel, Cellobiose (+), CMC, β-glucan (barley), and lichenan (moss) were analysed using thin-layer chromatography (TLC). The hydrolytic products formed after hydrolysis of D-(+)-cellobiose, β-glucan (barley), lichenan (moss), Avicel PH101, and CMC by BglP89 were also analysed using TLC. BglP89 was able to hydrolyse cellobiose, β-glucan (barley), and lichenan (moss) generating glucose as an end product after 1 hour. Prolonging the reaction on β-glucan and lichenan showed a significant increase in the amount of glucose produced **(Figure 30**). The ability of BglP89 to hydrolyse cellobiose to glucose confirms that BglP89 is indeed a β-glucosidase with hydrolytic activity on short cello-oligosaccharides as expected. Lichenan and β-glucan have a mixture of β-1,3 and β-1,4 glycosidic linkages in

different ratios, with β-glucan having 1:3 or 1:4 ratio of β-1,3 and β-1,4 glycosidic bonds. A future study assessing the activity on various β-1,3 glycosidic linkage substrates can also confirm or elucidate if these bonds are cleaved by BglP89. TLC analysis of the hydrolysis products of Avicel PH-101 and CMC cellulosic substrates showed a spot which at first was thought to represent cellobiose. However, if BglP89 conferred β-1,4-endoglucanase activity resulting in cellobiose, it would stand to reason that it would then convert the cellobiose to glucose, given its demonstrated activity on cellobiose in Figure 30. Furthermore, it is questionable whether the spots in lanes 13-15 (Figure 30) actually align with the cellobiose standard; therefore it is difficult to conclude whether it represents BglP89 activity., Prolonging the reaction incubation time to 24 hours did not show an increase in the amount of product produced (**Figure 30**). Therefore, the resultant spot on the TLC requires further analysis to determine the action of BglP89 on Avicel PH-101 and CMC cellulosic substrates.

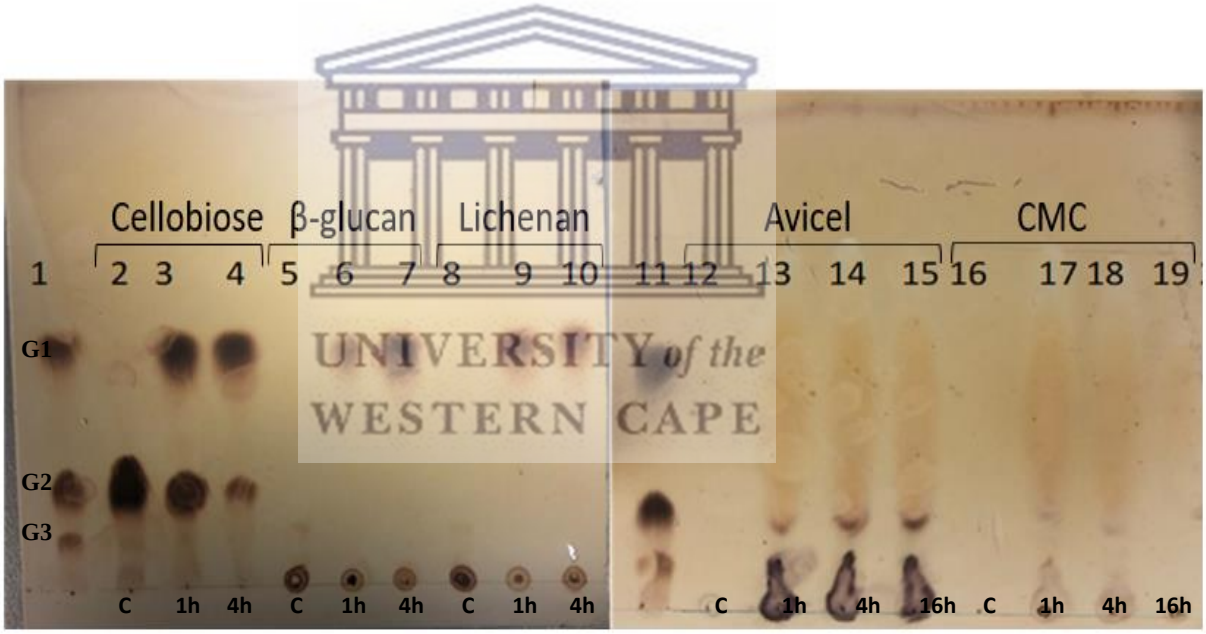

Figure 30: Thin-layer chromatography (TLC) analysis of products after BglP89 hydrolysis on cellobiose, β-glucan, and lichenan for 1h and 4h incubation, Avicel and CMC for 1h, 4h, and 16 hours respectively. Lane1: Standards mixture: (G1 - Glucose, G2 - Cellobiose and G3 - Cellotriose). Lane 2: cellobiose control (without enzyme). Lane 3-4: cellobiose with BglP89. Lane 5: β-glucan control (without enzyme). Lane 6-7: β-glucan with BglP89. Lane 8: lichenan control (without enzyme). Lane 9-10: lichenan with BglP89. Lane 11: Standards mixture (G1-Glucose, G2, Cellobiose, and G3-Cellotriose). Lane 12: Avicel PH-101 control (without enzyme). Lane 13-15: Avicel PH-101 with BglP89, Lane 16: CMC control (without enzyme). Lane 17-19: CMC with BglP89.

3.6 End product inhibition

Lignocellulose degrading enzymes have been reported to be competitively inhibited by the accumulation of their end products. Therefore, discovering novel lignocellulose degrading enzymes with a high tolerance to their end product would be an important property for various industrial applications. The effect of xylose and glucose was tested on the activity of XylP81 and BglP89, respectively. XylP81 showed remarkable tolerance to xylose as it retained about 42% activity in 3 M xylose (**Figure 31**). In comparison, a GH39 β-xylosidase from *Geobacillus* sp retained 50% activity at 300 mM xylose (Bhalla *et al*., 2014) and β-xylosidase from *Clostridium clariflavum* showed 28,6% inhibition at 200 mM xylose (Geng 2017). There is only one reported GH39 β-xylosidase, from *Dictytoglomus thermophilum*, with comparable end product tolerance (Li *et al*., 2018), retaining 40% relative activity in 3 M xylose.

BglP89 retained 46% activity in 150 mM glucose (**Figure 32**) and gradually lost activity with an increase in glucose concentration, which represents a relatively high tolerance in comparison to other reported β-glucosidases. A GH3 β-glucosidase from *Penicillum brasilianum* (Krogh *et al*., 2010) retained 50% activity in 2.3 mM glucose, whereas the β-glucosidase from *Myceliophthora thermophilia* (Zhao *et al*., 2015) retained 50% activity in just 282 µM. Most β-glucosidases are reported to be sensitive to and inhibited by their end product glucose (Yang *et al*., 2015) which binds competitively to the enzymes active site (Teugjas and Valjamae, 2013). This inhibition leads to the accumulation of cellobiose which is an end product inhibitor of endo/exo-glucanase activity thereby limiting and decreasing the rate of cellulose degradation (Costa *et al*., 2019). GH3 β-glucosidases are reported to be more susceptible to their end-product glucose compared to GH1 β-glucosidases. A study comparing GH1 and GH3 β-glucosidase structures suggested that the deeper narrow substrate channel in GH1 β-glucosidases make the active site inaccessible to glucose in contrast with the GH3 β-glucosidases which have a shallow pocket channel (Santos *et al*., 2019; Sun *et al*., 2018).

Both XylP81 and BglP89 showed high product inhibition tolerance compared with other enzymes in their respective families. High-end product tolerance is a desirable property for many industrial applications.

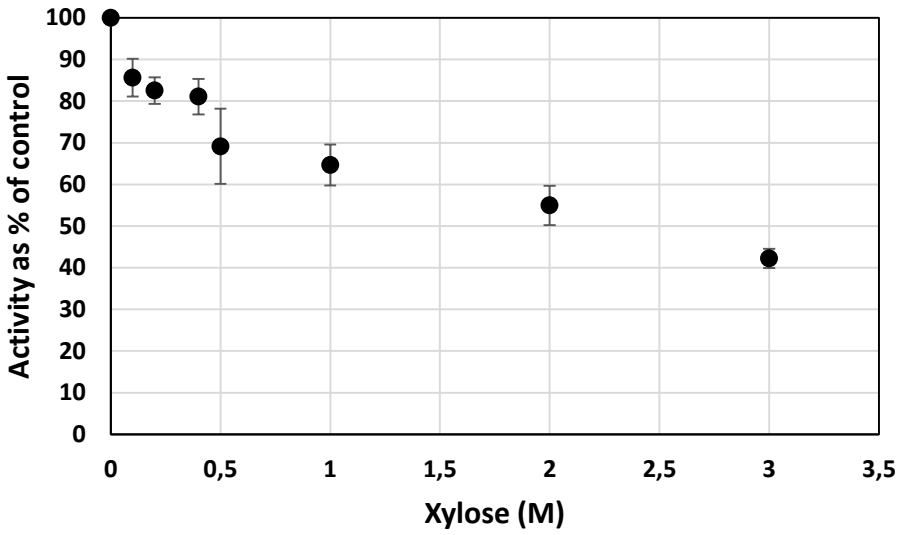

Figure 31: The effect of different concentrations of xylose on the activity of XylP81. Error bars represent the standard deviation of the average of three replicates.

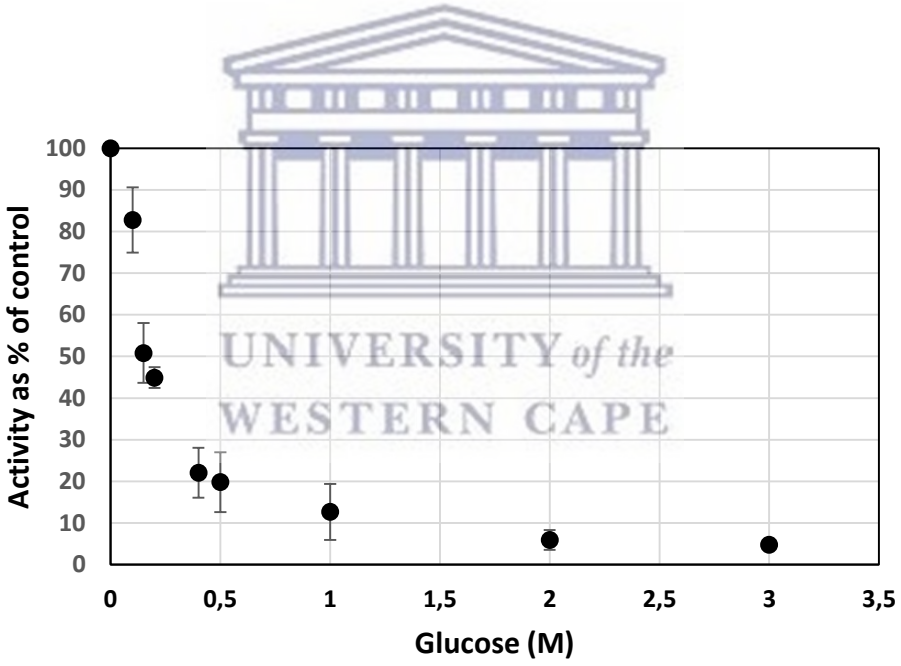

Figure 32: The effect of different concentrations of glucose on the activity of BglP89. Error bars represent the standard deviation of the average of three replicates.

3.7 Conclusion

In this study, three genes encoding glycoside hydrolases XylP55, XylP81, and BglP89 were successfully cloned from a horse manure composted metagenomic library which was

previously constructed. Sequence domain analysis classified these enzymes into GH43, GH39, and GH3 families, respectively. XylP81 and BglP89 were successfully characterised. Optimum activity was measured at pH 6 and 50°C with moderate thermostability indicating that they are moderately thermostable enzymes. Both enzymes showed activity against a broad substrate range and could have industrial and biotechnological applications.

Chapter 4: General conclusion

Lignocellulose plant biomass is the most abundant biomass on Earth, and it is composed of cellulose, hemicellulose, and lignin. The recalcitrant nature of plant biomass hinders its complete enzymatic degradation to fermentable sugars for industrial and biotechnological processes (Thapa *et al*., 2020; Zoghlami and Paës 2019). GHs, which include cellulases and xylanases, play an important role in the degradation of lignocellulosic substrates, as they hydrolyse the β-1,4-glycosidic bonds in cellulose and hemicellulose polymers. Natural environments such as soil (Alves *et al*., 2018), rumen (Wang *et al*., 2019), and compost (Dougherty *et al*., 2012) where lignocellulosic material is biodegraded, harbour bacteria and fungi that produce lignocellulolytic enzymes. This then makes these microbial communities a promising source for discovery of new enzymes of industrial and biotechnological importance. Thus, researchers have focused on investigating such environments to discover more GH enzymes that possess higher catalytic activity to degrade the recalcitrant lignocellulose biomass synergistically and efficiently.

Composting is a process that involves the degradation of organic waste materials. This is achieved by complex and diverse microbial communities which produce a variety of extracellular hydrolytic enzymes that synergistically decompose organic matter under controlled conditions (Partanen *et al*., 2010; Meng *et al*., 2019). The temperature in compositing rises to 80℃ due to the microbial activity and this stage is dominated by thermophilic microbial communities producing thermostable enzymes. These enzymes can withstand the high temperatures that are in demand in industrial processes as they reduce the possibility of microbial contamination, increase substrate and overall degradation efficiency (Sato *et al*., 2017). The full exploration of these glycoside hydrolase producing microbes is hindered by the inability to culture most of them by traditional culturing methods in the laboratory. To accelerate the discovery of novel, biotechnologically, and industrially important biocatalysts, metagenomics has been used to access the uncultured microorganisms in composting environments. Metagenomics is a culture-independent approach that involves the direct isolation of DNA from environmental samples and screening is achieved by using sequence-based and function-based approaches. Sequence-based screening identifies genes by comparing sequences to known sequence information in databases. The functional-based

screening relies on the preparation of a metagenomic library, the heterologous expression of environmental DNA clones followed by activity screening. This approach has great potential for revealing novel genes that could have been missed using sequence-based screening alone (Cheng *et al*., 2017 Lam *et al*., 2015). Function-based screening has been successfully reported in the literature for the discovery and identification of several novel GH encoding genes from compost samples. For example, novel glycoside hydrolases with unique properties which include esterase, β-glucosidase, endoxylanase, and xylanase were identified from a compost microbial metagenomic library by functional screening (Ohlhoff *et al.,* 2015; Uchiyama *et al*., 2013; Verma *et al*., 2013; Ellilä *et al* 2019). This suggests that compost is indeed a promising source for the discovery of new hydrolytic enzymes.

In a previous study, a metagenomic library was constructed from the thermophilic stage of horse manure compost and three fosmid clones (P55E4, P81G1, and P89A4) encoding GH encoding ORFs were selected for further study. The main aim of this study was to functionally characterise the designated XylP55, XylP81, and BglP89 enzymes encoded on these fosmids. Amino acid sequence analysis indicated that they belong to the GH43 subfamily 1, GH39, and GH3 family, respectively. Due to the low amino acid sequence similarity with other homologues in the CAZy database, they could potentially represent novel clades in their respective GH families. Further analysis also revealed that XylP55 is closely related to a β-xylosidase/α-L-arabinofuranosidase from a compost microbial metagenome, and therefore suggested that this enzyme could potentially display bi-functional activity (Matsuzawa *et al*., 2015). All three proteins were successfully expressed in the pET21a(+) expression vector and used to transform *E. coli* BL21(DE3)-Rosetta pLysS.

XylP81 and BglP89 were successfully purified and subsequently characterised. They both showed activity at a pH range of 5-8 a wide temperature range of 20-70°C and exhibited optimum activity at pH 6 and 50°C with moderate thermostability. These enzymes also showed broad substrate specificity as they have shown hydrolysing activity on a broad substrate range with high specific activity on the chromogenic or artificial substrates and low to no activity on more complex natural substrates. XylP81 showed activity on *p*NPX confirming that it's a β-xylosidase with activity on *p*NPA and *p*NPGal. BglP89 showed high activity on *p*NPG confirming that it's a β-glucosidase with significant activity on *p*NPA and *p*NPGal.

Furthermore, the enzymatic activity of the recombinant XylP81 and BglP89 was not negatively affected by the presence of metal ions tested indicating that they are metal-independent. Enzymes are often reported to be susceptible to inhibition by their end-product slowing down enzyme hydrolysis. The majority of β-glucosidases are often inhibited by an increase in glucose concentration and BglP89 lost 80% activity in the presence of 500 mM glucose. GH3 β-glucosidases are more prone to glucose inhibition compared to GH1 β-glucosidases and studies have predicted that this might be due to structural differences (Santos *et al*., 2019). On the other hand, XylP81 retained about 40% activity in the presence of 3 M xylose.

Future works would look at testing the activity of XylP81 on additional substrates and activity of BglP89 on Avicel and CMC followed by HPLC quantitative analysis of the hydrolysis products by XylP81. Characterising the enzyme kinetic parameters of XylP81 and BglP89 on more complex substrates would provide insight on the affinity and catalytic efficiency of the enzymes on naturally occurring substrates. Furthermore, testing the synergistic effect of β-xylosidase XylP81 with β-1,4-endoxylanases and β-glucosidase BglP89 with β-1,4-endoglucanase to enhance the hydrolysis of xylan and cellulose hydrolysis in industrial applications would be warranted.

In conclusion, this study successfully characterised a new β-xylosidase (XylP81) and β-glucosidase (BglP89), both of which are rate-limiting enzymes for the complete degradation of xylan and cellulose respectively. Both enzymes were identified through functional-based metagenomic screening. The biochemical characteristics of these enzymes make them promising candidates for use in various industrial and biotechnological applications. This also confirms that the composting microbial community is a rich, valuable source for prospecting GHs which are potentially useful in enzymatic degradation of lignocellulose biomass.

References

Adelsberger, H.H.C., Glawischnig, E., Zverlov, V. V., & Schwarz, W. H. (2004). Enzyme system of Clostridium stercorarium for hydrolysis of arabinoxylan: Reconstitution of the in vivo system from recombinant enzymes. *Microbiology*, 150(7), pp. 2257–2266. https://doi.org/10.1099/mic.0.27066-0.

Ahmad, T., Singh, RS., Gupta, G., Sharma, A., Kaur, B. (2019). Metagenomics in the search for industrial enzymes. *Advances in Enzyme Technology*, 1(15), pp. 419-451. https://doi.org/10.1016/B978-0-444-64114-4.00015-7.

Ahmed, S., Luis, A. S., Bras, J. L. A., Ghosh, A., Gautam, S., Gupta, M. N., … Goyal, A. (2013). A Novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (Ct43Araf) from *Clostridium thermocellum*. *PLoS ONE*, 8(9), pp. 1-10. https://doi.org/10.1371/journal.pone.0073575.

Albuquerque, L., Polónia, A. R. M., Barroso, C., Froufe, H. J. C., Lage, O., Lobo-Da-Cunha, A., … Da Costa, M. S. (2018). Raineya orbicola gen. nov., sp. nov. a slightly thermophilic bacterium of the phylum bacteroidetes and the description of raineyaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 68(4), pp. 982–989. https://doi.org/10.1099/ijsem.0.002556.

Alessi, A. M., Gray, V., Farquharson, F. M., Flores-López, A., Shaw, S., Stead, D., … Louis, P. (2020). β-Glucan is a major growth substrate for human gut bacteria related to Coprococcus eutactus. *Environmental Microbiology*, 22(6), pp. 2150-2164. https://doi.org/10.1111/1462-2920.14977.

Allgaier, M., Reddy, A., Park, J. I., Ivanova, N., D'Haeseleer, P., Lowry, S., … Hugenholtz, P. (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS ONE*, 5(1), p. 8812. https://doi.org/10.1371/journal.pone.0008812.

Alves, L. de F., Meleiro, L. P., Silva, R. N., Westmann, C. A., & Guazzaroni, M.-E. (2018). Novel ethanol- and 5-hydroxymethyl furfural-stimulated β-glucosidase retrieved from a Brazilian secondary atlantic forest soil metagenome. *Frontiers in Microbiology*, 9, p. 2556. https://doi.org/10.3389/fmicb.2018.02556.

Armstrong, Z., Liu, F., Kheirandish, S., Chen, H.-M., Mewis, K., Duo, T., … Withers, S. G. (2019). High-throughput recovery and characterization of metagenome-derived glycoside hydrolase-containing clones as a resource for biocatalyst development. *MSystems*, 4(4), pp. 1-19. https://doi.org/10.1128/msystems.00082-19.

Bao, L., Huang, Q., Chang, L., Sun, Q., Zhou, J., & Lu, H. (2012). Cloning and characterization of two β-glucosidase/xylosidase enzymes from yak rumen metagenome. *Applied Biochemistry and Biotechnology*, 166(1), pp. 72–86. https://doi.org/10.1007/s12010-011-9405-x.

Bergmann, J. C., Costa, O. Y. A., Gladden, J. M., Singer, S., Patrik, D., Simmons, B. A., & Quirino, B. F. (2014). Discovery of two novel β-glucosidases from an Amazon soil metagenomic library. *FEMS Microbiology Letters*, 351(2), pp. 147–155. https://doi.org/10.1111/1574-6968.12332.

Bhalla, A., Bischoff, K. M., & Sani, R. K. (2014). Highly thermostable GH39 β-xylosidase from a *Geobacillus* sp. strain WSUCF1. *BMC Biotechnology*, 14(1), pp. 7–11. https://doi.org/10.1186/s12896-014-0106-8.

Boers, S. A., Jansen, R., & Hays, J. P. (2019). Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *European society of clinical microbiology*, 38(6), pp. 1059-1070. https://doi.org/10.1007/s10096-019-03520-3.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1), pp. 248-254. https://doi.org/10.1016/0003-2697(76)90527-3.

Bravman, T., Zolotnitsky, G., Shulami, S., Belakhov, V., Solomon, D., Baasov, T., … Shoham, Y. (2001). Stereochemistry of family 52 glycosyl hydrolases: A β-xylosidase from *Bacillus stearothermophilus* T-6 is a retaining enzyme. *FEBS Letters*, 495(1–2), pp. 39–43. https://doi.org/10.1016/S0014-5793(01)02360-2.

Chang, L., Ding, M., Bao, L., Chen, Y., Zhou, J., & Lu, H. (2011). Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Applied Microbiology and Biotechnology*, 90(6), pp. 1933–1942. https://doi.org/10.1007/s00253-011-3182-x.

Cheng, F., Sheng, J., Dong, R., Men, Y., Gan, L., & Shen, L. (2012). Novel xylanase from a holstein cattle rumen metagenomic library and its application in xylooligosaccharides and ferulic acid production from wheat straw. *Journal of agricultural and food chemistry*, 60(51), pp. 12516-24. https://doi.org/10.1021/JF302337w.

Cheng, J., Romantsov, T., Engel, K., Doxey, A. C., Rose, D. R., Neufeld, J. D., & Charles, T. C. (2017). Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS ONE*, 12(3), pp. 1–20. https://doi.org/10.1371/journal.pone.0172545.

Collins, T., Gerday, C., & Feller, G. (2005). Xylanases, xylanase families, and extremophilic xylanases. *FEMS Microbiology Reviews*, 29(1), pp. 3-23. https://doi.org/10.1016/j.femsre.2004.06.005.

Colombo, L. T., de Oliveira, M. N. V., Carneiro, D. G., de Souza, R. A., Alvim, M. C. T., dos Santos, J. C., … Passos, F. M. L. (2016). Applying functional metagenomics to search for novel lignocellulosic enzymes in a microbial consortium derived from a thermophilic composting phase of sugarcane bagasse and cow manure. *Antonie van Leeuwenhoek*, 109(9), pp. 1217-1233. https://doi.org/10.1007/s10482-016-0723-4.

Corrêa, J. M., Graciano, L., Abrahão, J., Loth, E. A., Gandra, R. F., Kadowaki, M. K., … Simão, R. D. C. G. (2012). Expression and characterization of a GH39 β-xylosidase II from *Caulobacter crescentus*. *Applied Biochemistry and Biotechnology*, 168(8), pp. 2218–2229. https://doi.org/10.1007/s12010-012-9931-1.

Costa, L. S. C., Mariano, D. C. B., Rocha, R. E. O., Kraml, J., da Silveira, C. H., Liedl, K. R., … De Lima, L. H. F. (2019). Molecular dynamics gives new insights into the glucose tolerance and inhibition mechanisms on β-glucosidases. *Molecules*, 24(18), p. 3215. https://doi.org/10.3390/molecules24183215.

Culligan, E. P., Sleator, R. D., Marchesi, J. R., & Hill, C. (2014). Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence*, 5(3), pp. 399–412. https://doi.org/10.4161/viru.27208.

Datta, S., Rajnish, K. N., Samuel, M. S., Pugazlendhi, A., & Selvarajan, E. (2020). Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme, and drug discovery. A review. *Environmental Chemistry Letters*, 18(4), pp. 1229-1241. https://doi.org/10.1007/s10311-020-01010-z.

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3(9), pp. 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9.

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, pp. 1–21. https://doi.org/10.3389/fmicb.2016.01521.

Dodd, D., & Cann, I. K. O. (2009). Enzymatic deconstruction of xylan for biofuel production. *GCB Bioenergy*, 1(1), pp. 2–17. https://doi.org/10.1111/j.1757-1707.2009.01004.x.

Dougherty, M. J., D'haeseleer, P., Hazen, T. C., Simmons, B. A., Adams, P. D., & Hadi, M. Z. (2012). Glycoside hydrolases from a targeted compost metagenome, activity-screening, and functional characterization. *BMC Biotechnology*, 12 (2012), p. 38. https://doi.org/10.1186/1472-6750-12-38.

Ellilä, S., Bromann, P., Nyyssönen, M., Itävaara, M., Koivula, A., Paulin, L., & Kruus, K. (2019). Cloning of novel bacterial xylanases from lignocellulose-enriched compost metagenomic libraries. *AMB Express*, 9(1), p. 124. https://doi.org/10.1186/s13568-019-0847-9.

Escobar-Zepeda, A., De León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6, pp. 1–15. https://doi.org/10.3389/fgene.2015.00348.

Escuder-Rodríguez, J. J., DeCastro, M. E., Becerra, M., Rodríguez-Belmonte, E., & González-Siso, M. I. (2018). Advances of functional metagenomics in harnessing thermozymes. *Metagenomics*, pp. 289-307. https://doi.org/10.1016/B978-0-08-102268-9.00015-X.

Fisher, A., & S Fong, S. (2014). Lignin biodegradation and industrial implications. *AIMS Bioengineering*, 1(2), pp. 92–112. https://doi.org/10.3934/bioeng.2014.2.92.

Geng, A., Wang, H., Wu, J., Xie, R., & Sun, J. (2017). Characterization of a β-xylosidase from *Clostridium clariflavum* and its application in xylan hydrolysis. *BioResources*, 12(4), pp. 9253-9262. https://doi.org/10.15376/biores.12.4.9253-9262.

Gomes-Pepe, E. S., Sierra, E. G. M., Pereira, M. R., Castellane, T. C. L., & De Lemos, E. G. M. (2016). Bg10: A novel metagenomics alcohol-tolerant and glucose-stimulated GH1 β-glucosidase suitable for lactose-free milk preparation. *PLoS ONE*, 11(12), pp. 1–25. https://doi.org/10.1371/journal.pone.0167932.

Gruninger, R. J., Gong, X., Forster, R. J., & McAllister, T. A. (2014). Biochemical and kinetic characterization of the multifunctional β-glucosidase/β-xylosidase/α-arabinosidase, Bgxa1. *Applied Microbiology and Biotechnology*, 98(7), pp. 3003–3012. https://doi.org/10.1007/s0025-013-5191-4.

Gumerov, V. M., Rakitin, A. L., Mardanov, A. V., & Ravin, N. V. (2015). A novel highly thermostable multifunctional beta-glycosidase from *Crenarchaeon Acidilobus saccharovorans*. *Archaea*, 2015(976832). https://doi.org/10.1155/2015/978632.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280(2), pp. 309–316. https://doi.org/10.1042/bj2800309.

Hess M, Sczyrba A, Egan R, et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331, pp. 463-467. https://doi.org/10.1126/science.1200387.

Horn, S. J., Vaaje-Kolstad, G., Westereng, B., & Eijsink, V. G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels*, 5, p. 45. https://doi.org/10.1186/1754-6834-5-45.

Hua, M., Zhao, S., & Zhang, L. (2015). Direct detection, cloning, and characterization of a glucoside hydrolase from forest soil. *Biotechnology Letters*, 37(6), pp. 1227–1232. https://doi.org/10.1007/s10529-015-1777-5.

Huang, Y., Zheng, X., Pilgaard, B., Holck, J., Muschiol, J., Li, S., & Lange, L. (2019). Identification and characterization of GH11 xylanase and GH43 xylosidase from the chytridiomycetous fungus, *Rhizophlyctis rosea*. *Applied Microbiology and Biotechnology*, 103(2), pp. 777–791. https://doi.org/10.1007/s00253-018-9431-5.

Hubbe, M. A., Nazhad, M., & Sánchez, C. (2010). Composting as a way to convert cellulosic biomass and organic waste into high-value soil amendments: A review. *BioResources*, 5(4), pp. 2808–2854. https://doi.org/10.15376/biores.5.4.2808-2854.

Hwang, E. J., Lee, Y. S., & Choi, Y. L. (2018). Cloning, purification, and characterization of the organic solvent tolerant β-glucosidase, OaBGL84, from *Olleya aquimaris* DAU311. *Applied Biological Chemistry*, 61(3), pp. 325–336. https://doi.org/10.1007/s13765-018-0361-9.

Isikgor, F. H., & Becer, C. R. (2015). Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polymer Chemistry*, 6(25), pp. 4497–4559. https://doi.org/10.1039/c5py00263j.

Jordan, D. B., & Wagschal, K. (2010). Properties and applications of microbial β-D-xylosidases featuring the catalytically efficient enzyme from *Selenomonas ruminantium*. *Applied Microbiology and Biotechnology*, 86(6), pp. 1647–1658. https://doi.org/10.1007/s00253-010-2538-y.

Joshi, N., Sharma, M. & Singh, S.P., (2020). Characterization of a novel xylanse from an extreme termperature hot spring metagemnome for xylooligosaccharides production. *Applied Microbiology and Biotechnology*, 104, pp. 4889-4901. https://doi.org/10.1007/s00253-020-10562-7.

Jünemann, S., Kleinbölting, N., Jaenicke, S., Henke, C., Hassa, J., Nelkner, J., … Stoye, J. (2017). Bioinformatics for NGS-based metagenomics and the application to biogas research. *Journal of Biotechnology*, 261, pp. 10–23. https://doi.org/10.1016/j.jbiotec.2017.08.012

Justo, P. I., Corrêa, J. M., Maller, A., Kadowaki, M. K., da Conceição-Silva, J. L., Gandra, R. F., & Simão, R. de C. G. (2015). Analysis of the xynB5 gene encoding a multifunctional GH3-BglX β-glucosidase-β-xylosidase-α-arabinosidase member in *Caulobacter crescentus*. *Antonie van Leeuwenhoek*, 108(4), pp. 993–1007. https://doi.org/10.1007/s10482-015-0552-x.

Kaur, J. (2007). Purification and characterization of β-glucosidase from *Melanocarpus* sp. MTCC 3922. *Electronic Journal of Biotechnology*, 10(2), pp. 1015-1104. https://doi.org/10.2225/vol10-issue2-fulltext-4.

Knapik, K., Becerra, M., & González-Siso, M. I. (2019). Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios Hot Spring in Spain. *Scientific Reports*, 9(1), pp. 1–12. https://doi.org/10.1038/s41598-019-47637-z.

Knob, A., & Carmona, E. C. (2009). Cell-associated acid β-xylosidase production by *Penicillium sclerotiorum*. *New Biotechnology*, 26(1–2), pp. 60–67. https://doi.org/10.1016/j.nbt.2009.03.002.

Ko, K. C., Lee, J. H., Han, Y., Choi, J. H., & Song, J. J. (2013). A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal bacteria. *Biochemical and Biophysical Research Communications*, 441(3), pp. 567–572. https://doi.org/10.1016/j.bbrc.2013.10.120.

Kodzius, R., & Gojobori, T. (2015). Marine metagenomics as a source for bioprospecting. *Marine Genomics*, 24, pp. 21–30. https://doi.org/10.1016/j.margen.2015.07.001.

Kopanic, J. L., Al-mugotir, M., Zach, S., Das, S., Grosely, R., & Sorgen, P. L. (2013). An *Escherichia coli* strain for expression of the connexin45 carboxyl terminus attached to the 4th transmembrane domain. *Frontiers in pharmacology* 4(106), pp. 4–11. https://doi.org/10.3389/fphar.2013.00106.

Kousar, S., Mustafa, G., & Jamil, A. (2013). Microbial xylosidases: Production and biochemical characterization. *Pakistan Journal of Life and Social Sciences*, 11(2), pp. 85–95.

Krogh, K. B. R. M., Harris, P. V., Olsen, C. L., Johansen, K. S., Hojer-Pedersen, J., Borjesson, J., & Olsson, L. (2010). Characterization and kinetic analysis of a thermostable GH3 β-glucosidase from *Penicillium brasilianum*. *Applied Microbiology and Biotechnology*, 86(1), pp. 143–154. https://doi.org/10.1007/s00253-009-2181-7.

Kucharska, K., Rybarczyk, P., Hołowacz, I., Łukajtis, R., Glinka, M., & Kamiński, M. (2018). Pretreatment of lignocellulosic materials as substrates for fermentation processes. *Molecules*, 23(11), pp. 1–32. https://doi.org/10.3390/molecules23112937.

Kuhad, R. C., Gupta, R., & Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Research*, 2011, 280696. https://doi.org/10.4061/2011/280696.

Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives. *Journal of Industrial Microbiology and Biotechnology*, 35(5), pp. 377–391. https://doi.org/10.1007/s10295-008-0327-8.

Kumar, A., & Chandra, R. (2020). Ligninolytic enzymes and its mechanisms for degradation of lignocellulosic waste in environment. *Heliyon*, 6(2), e03170. https://doi.org/10.1016/j.heliyon.2020.e03170.

Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., & Charles, T. C. (2015). Current and future resources for functional metagenomics. *Frontiers in Microbiology*, 6, pp. 1–8. https://doi.org/10.3389/fmicb.2015.01196.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), pp. 680–685. https://doi.org/10.1038/227680a0.

Lemos, L. N., Pereira, R. V., Quaggio, R. B., Martins, L. F., Moura, L. M. S., da Silva, A. R., … Setubal, J. C. (2017). Genome-centric analysis of a thermophilic and cellulolytic bacterial consortium derived from composting. *Frontiers in Microbiology*, 8, p. 644. https://doi.org/10.3389/fmicb.2017.00644.

Li, Y., Liu, N., Yang, H., Zhao, F., Yu, Y., Tian, Y., & Lu, X. (2014). Cloning and characterization of a new β-Glucosidase from a metagenomic library of Rumen of cattle feeding with *Miscanthus sinensis*. *BMC Biotechnology*, 14(1), pp. 1–9. https://doi.org/10.1186/1472-6750-14-85.

Li, Q., Wu, T., Qi, Z., Zhao, L., Pei, J., & Tang, F. (2018). Characterization of a novel thermostable and xylose-tolerant GH 39 β-xylosidase from *Dictyoglomus thermophilum*. *BMC Biotechnology*, 18, p. 29. https://doi.org/10.1186/s12896-018-0440-3.

Linares-Pastén, J., Andersson, M., & Karlsson, E. (2014). Thermostable glycoside hydrolases in biorefinery technologies. *Current Biotechnology*, 3(1), pp. 26–44. https://doi.org/10.2174/2211550111302666004 1.

Liu, D., Zhang, R., Yang, X., Zhang, Z., Song, S., Miao, Y., & Shen, Q. (2012). Characterization of a thermostable β-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. *Microbial Cell Factories*, 11, pp. 1–15. https://doi.org/10.1186/1475-2859-11-25.

Liu, N., Li, H., Chevrette, M. G., Zhang, L., Cao, L., Zhou, H., … Wang, Q. (2019). Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *ISME Journal*, 13(1), pp. 104–117. https://doi.org/10.1038/s41396-018-0255-1.

Long, L., Shi, H., Li, X., Zhang, Y., Hu, J., & Wang, F. (2016). Cloning, Purification, and Characterization of a thermostable β-glucosidase from *Thermotoga thermarum* DSM 5069. *BioResources*, 11(2), pp. 3165–3177.

Mai, Z., Su, H., & Zhang, S. (2016). Characterization of a metagenome-derived β-glucosidase and its application in conversion of polydatin to resveratrol. *Catalysts*, 6(3), 35. https://doi.org/10.3390/catal6030035.

Maitan-Alfenas, G. P., Oliveira, M. B., Nagem, R. A. P., de Vries, R. P., & Guimarães, V. M. (2016). Characterization and biotechnological application of recombinant xylanases from *Aspergillus nidulans*. *International Journal of Biological Macromolecules*, 91, pp. 60–67. https://doi.org/10.1016/j.ijbiomac.2016.05.065.

Maruthamuthu, M., and Van Elsas, J. D. (2017). Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library. *Biotechnology for Biofuels*, 10(1), pp. 1–17. https://doi.org/10.1186/s13068-017-0808-y.

Matsuzawa, T., Kaneko, S., & Yaoi, K. (2015). Screening, identification, and characterization of a GH43 family β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome. *Applied Microbiology and Biotechnology*, 99(21), pp. 8943–8954. https://doi.org/10.1007/s00253-015-6647-5.

Matsuzawa, T., Kaneko, S., Kishine, N., Fujimoto, Z., Yaoi, K., (2017). Crystal structure of metagenomic β-xylosidase/ α-L-arabinofuranosidase activated by calcium, *The Journal of Biochemistry*, 162(3), pp. 173–181. https://doi.org/10.1093/jb/mvx012.

Mehta, R., Singhal, P., Singh, H., & Damle, D. (2016). Insight into thermophiles and their wide-spectrum applications. *Biotechnology*, 6(1), pp. 1–9. https://doi.org/10.1007/s13205-016-0368-z.

Méndez-Líter, J. A., Gil-Muñoz, J., Nieto-Domínguez, M., Barriuso, J., De Eugenio, L. I., & Martínez, M. J. (2017). A novel, highly efficient β-glucosidase with a cellulose-binding domain: Characterization and properties of native and recombinant proteins. *Biotechnology for Biofuels*, 10(1), pp. 1–15. https://doi.org/10.1186/s13068-017-0946-2.

Meng, Q., Yang, W., Men, M., Bello, A., Xu, X., Xu, B., … Zhu, H. (2019). Microbial community succession and response to environmental variables during cow manure and corn straw composting. *Frontiers in Microbiology*, 10, pp. 1–13. https://doi.org/10.3389/fmicb.2019.00529.

Mercedes, R. E., Julia, M. N., Del Pozo, M. V., González, B., Golyshin, P. N., Polaina, J., … Julia, S. A. (2016). Structural and functional characterization of a ruminal β-glycosidase defines a novel subfamily of glycoside hydrolase family 3 with permuted domain topology. *Journal of Biological Chemistry*, 291(46), pp. 24200–24214. https://doi.org/10.1074/jbc.M116.747527.

Mewis, K., Lenfant, N., Lombard, V., & Henrissat, B. (2016). Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Applied and Environmental Microbiology*, 82(6), pp. 1686–1692. https://doi.org/10.1128/AEM.03453-15.

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry*, 31(3), pp. 426–428. https://doi.org/10.1021/ac60147a030.

Michlmayr, H., Varga, E., Malachova, A., Nguyen, N. T., Lorenz, C., Haltrich, D., … Adam, G. (2015). A versatile family 3 glycoside hydrolase from *Bifidobacterium adolescentis* hydrolyzes β-glucosides of the Fusarium mycotoxins deoxynivalenol, nivalenol, and HT-2 toxin in cereal matrices. *Applied and Environmental Microbiology*, 81(15), pp. 4885–4893. https://doi.org/10.1128/AEM.01061-15.

Mladenov, M. (2018). Chemical composition of different types of compost. *Journal of Chemical Technology and Metallurgy*, 53(4), pp. 712–716.

Mohammad, B. T., Daghistani, H. I. Al, Jaouani, A., Abdel-latif, S., & Kennes, C. (2017). Isolation and characterization of thermophilic bacteria from Jordanian hot springs: *Bacillus licheniformis* and *Thermomonas hydrothermalis* isolates as potential producers of thermostable enzymes. *International Journal of Microbiology*, 2017, 6943952. https://doi.org/10.1155/2017/6943952.

Mohsin, I., Poudel, N., Li, D. C., & Papageorgiou, A. C. (2019). Crystal structure of a GH3 β-glucosidase from the thermophilic fungus *Chaetomium thermophilum*. *International Journal of Molecular Sciences*, 20(23), pp. 1–15. https://doi.org/10.3390/ijms20235962.

Montella, S., Ventorino, V., Lombard, V., Henrissat, B., Pepe, O., & Faraco, V. (2017). Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. *Scientific Reports*, 7, pp. 1–15. https://doi.org/10.1038/srep42623.

Morrison, J. M., Elshahed, M. S., & Youssef, N. (2016). A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A. *PeerJ*, 3, pp. 1–25. https://doi.org/10.7717/PEERJ.2289.

Murphy, J., & Walsh, G. (2019). Purification and characterization of a novel thermophilic β-galactosidase from *Picrophilus torridus* of potential industrial application. *Extremophiles*, 23(6), pp. 783–792. https://doi.org/10.1007/s00792-019-01133-4.

Nevondo W. (2016). Development of a high throughput cell-free metagenomic screening platform. PhD thesis UWC IMBM.

Nieto-Domínguez, M., de Eugenio, L. I., Barriuso, J., Prieto, A., de Toro, B. F., Canales-Mayordomo, ángeles, & Martínez, M. J. (2015). Novel pH-stable glycoside hydrolase family 3 β-xylosidase from *Talaromyces amestolkiae*: An enzyme displaying regioselective transxylosylation. *Applied and Environmental Microbiology*, 81(18), pp. 6380–6392. https://doi.org/10.1128/AEM.01744-15.

Ohlhoff, C. W., Kirby, B. M., Van Zyl, L., Mutepfa, D. L. R., Casanueva, A., Huddy, R. J., … Tuffin, M. (2015). An unusual feruloyl esterase belonging to family VIII esterases and displaying a broad substrate range. *Journal of Molecular Catalysis B: Enzymatic*, 118, pp. 79-88. https://doi.org/10.1016/j.molcatb.2015.04.010.

Panda, M. K., & Sahu, M. K. (2013). Isolation and characterization of a thermophilic *Bacillus* sp. with protease activity isolated from hot spring of Tarabalo, Odisha, India. *Iranian journal of microbiology*, 5(2), pp. 159–165.

Park, Y. J., Jeong, Y. U., & Kong, W. S. (2018). Genome sequencing and carbohydrate-active enzyme (CAZyme) repertoire of the white rot fungus *Flammulina elastica*. *International Journal of Molecular Sciences*, 19(8), p. 2379. https://doi.org/10.3390/ijms19082379.

Partanen, P., Hultman, J., Paulin, L., Auvinen, P., & Romantschuk, M. (2010). Bacterial diversity at different stages of the composting process. *BMC Microbiology*, 10, p. 94. https://doi.org/10.1186/1471-2180-10-94.

Pei, J., Pang, Q., Zhao, L., Fan, S., & Shi, H. (2012). *Thermoanaerobacterium thermosaccharolyticum* β-glucosidase: A glucose-tolerant enzyme with high specific activity for cellobiose. *Biotechnology for Biofuels*, 5, pp. 1–10. https://doi.org/10.1186/1754-6834-5-31.

Pérez, J., Muñoz-Dorado, J., De La Rubia, T., & Martínez, J. (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *International Microbiology*, 5(2), pp. 53–63. https://doi.org/10.1007/s10123-002-0062-3.

Pyeon, H. M., Lee, Y. S., & Choi, Y. L. (2019). Cloning, purification, and characterization of GH3 β-glucosidase, MtBgl85, from *Microbulbifer thermotolerans* DAU221. *Peer J*, (7), pp. 1–18. https://doi.org/10.7717/peerj.7106.

Rånby, B. (2001). Natural cellulose fibres and membranes: Biosynthesis. *Encyclopedia of Material: Science and Technology*, pp.5938-5943. http://doi.org/10.1016/B0-08-043152-6/01034-2.

Ratnadewi, A. A. I., Fanani, M., Kurniasih, S. D., Sakka, M., Wasito, E. B., Sakka, K., … Puspaningsih, N. N. T. (2013). β-D-xylosidase from *Geobacillus thermoleovorans* IT-08: Biochemical characterization and bioinformatics of the enzyme. *Applied Biochemistry and Biotechnology*, 170(8), pp. 1950–1964. https://doi.org/10.1007/s12010-013-0329-5.

Rohman, A., Dijkstra, B. W., & Puspaningsih, N. N. T. (2019). β-xylosidases: Structural diversity, catalytic mechanism, and inhibition by monosaccharides. *International Journal of Molecular Sciences*, 20(22), pp. 7–11. https://doi.org/10.3390/ijms20225524.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*, 454(7206), pp. 841–845. https://doi.org/10.1038/nature07190.

Sahoo, K., Sahoo, R. K., Gaur, M., & Subudhi, E. (2020). Cellulolytic thermophilic microorganisms in white biotechnology: a review. *Folia Microbiologica*, 65(1), pp. 25–43. https://doi.org/10.1007/s12223-019-00710-6.

Santos, C. A., Morais, M. A. B., Terrett, O. M., Lyczakowski, J. J., Zanphorlin, L. M., Ferreira-Filho, J. A., … Souza, A. P. (2019). An engineered GH1 β-glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials. *Scientific Reports*, 9(1), pp. 1–10. https://doi.org/10.1038/s41598-019-41300-3.

Sathya, T. A., & Khan, M. (2014). Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry. *Journal of Food Science*, 79(11), pp. 2149–2156. https://doi.org/10.1111/1750-3841.12677.

Sato, M., Suda, M., Okuma, J., Kato, T., Hirose, Y., Nishimura, A., … Shibata, D. (2017). Isolation of highly thermostable β-xylosidases from a hot spring soil microbial community using a metagenomic approach. *DNA Research*, 24(6), pp. 649–656. https://doi.org/10.1093/dnares/dsx032.

Shao, W., Xue, Y., Wu, A., Kataeva, I., Pei, J., Wu, H., & Wiegel, J. (2011). Characterization of a novel β-xylosidase, XylC, from *thermoanaerobacterium saccharolyticum* JW/SL-YS485. *Applied and Environmental Microbiology*, 77(3), pp. 719–726. https://doi.org/10.1128/AEM.01511-10.

Shen, Y., Li, Z., Huo, Y. Y., Bao, L., Gao, B., Xiao, P., … Li, J. (2019). Structural and functional insights into CmGH1, a novel GH39 family β-glucosidase from deep-sea bacterium. *Frontiers in Microbiology*, 10, p. 2922. https://doi.org/10.3389/fmicb.2019.02922.

Smaali, I., Rémond, C., & O'Donohue, M. J. (2006). Expression in *Escherichia coli* and characterization of β-xylosidases GH39 and GH-43 from *Bacillus halodurans* C-125. *Applied Microbiology and Biotechnology*, 73(3), pp. 582–590. https://doi.org/10.1007/s00253-006-0512-5.

Sørensen, A., Lübeck, M., Lübeck, P. S., & Ahring, B. K. (2013). Fungal beta-glucosidases: A bottleneck in industrial use of lignocellulosic materials. *Biomolecules*, 3(3), pp. 612–631. https://doi.org/10.3390/biom3030612.

Srivastava, N., Rathour, R., Jha, S., Pandey, K., Srivastava, M., Thakur, V. K., … Mishra, P. K. (2019). Microbial beta glucosidase enzymes: Recent advances in biomass conversation for biofuels application. *Biomolecules*, 9(6), pp. 1–23. https://doi.org/10.3390/biom9060220.

Strakowska J, Błaszczyk L, and Chełkowski J. (2018). The significance of cellulolytic enzymes produced by *Trichoderma* in the opportunistic lifestyle of this fungus. *Journal of basic microbiology*, 54, pp. 2-13. https://doi.org/10.1002/jobm.201300821.

Sun, J., Wang, W., Yao, C., Dai, F., Zhu, X., Liu, J., & Hao, J. (2018). Overexpression and characterization of a novel cold-adapted and salt-tolerant GH1 β-glucosidase from the marine bacterium *Alteromonas* sp. L82. *Journal of Microbiology*, 56(9), pp. 656–664. https://doi.org/10.1007/s12275-018-8018-2.

Sung, M., Kim, H., Bae, J., Rhee, S., Jeon, C. O., Kim, K., … Baek, D. (2002). *Geobacillus toebii* sp. nov., a novel thermophilic bacterium isolated from hay compost. *International journal of systematic and evolutionary microbiology*, 52, pp. 2251–2255. https://doi.org/10.1099/ijs.0.02181-0.

Tandon, G. D. (2015). Bioproducts from residual lignocellulosic biomass. *Advances in Biotechnology*, pp. 52–75. http://schoolar.google.com/scholar?q.

Teugjas, H., & Väljamäe, P. (2013). Selecting β-glucosidases to support cellulases in cellulose saccharification. *Biotechnology for Biofuels*, 6(1), pp. 1–13. https://doi.org/10.1186/1754-6834-6-105.

Thapa, S., Mishra, J., Arora, N., Mishra, P., Li, H., O′Hair, J., … Zhou, S. (2020). Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. Reviews. *Environmental Science and Biotechnology*, 19(3), pp. 621–648. https://doi.org/10.1007/s11157-020-09536-y.

Thornbury, M., Sicheri, J., Slaine, P., Getz, L. J., Finlayson-Trick, E., Cook, J., … McCormick, C. (2018). Discovery and characterization of novel lignocellulose-degrading enzymes from the porcupine microbiome by synthetic metagenomics. *PLoS ONE*, 14(1), e0209221. https://doi.org/10.1371/journal.pone.0209221.

Tully, B. J., Graham, E. D., & Heidelberg, J. F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data*, 5, pp. 1–8. https://doi.org/10.1038/sdata.2017.203.

Uchiyama, T., Miyazaki, K., & Yaoi, K. (2013). Characterization of a novel β-Glucosidase from a compost microbial metagenome with strong transglycosylation activity. *Journal of Biological Chemistry*, 288(25), pp. 18325–18334. https://doi.org/10.1074/jbc.M113.471342.

Usui, K., Ibata, K., Suzuki, T., & Kawai, K. (1999). XynX, a possible exo-xylanase of *Aeromonas caviae* ME-1 that produces exclusively xylobiose and xylotetraose from xylan. *Bioscience, Biotechnology and Biochemistry*, 63(8), pp. 1346–1352. https://doi.org/10.1271/bbb.63.1346.

Valášková, V., & Baldrian, P. (2006). Degradation of cellulose and hemicelluloses by the brown rot fungus Piptoporus betulinus. Production of extracellular enzymes and characterization of the major cellulases. *Microbiology*, 152(12), pp. 3613–3622. https://doi.org/10.1099/mic.0.29149-0.

Valenzuela, S. V, Lopez, S., Biely, P., Sanz-aparicio, J., & Pastor, F. I. J. (2016). The glycoside hydrolase family 8 reducing-end xylose-releasing exo-oligoxylanase Rex8A from *Paenibacillus barcinonensis* BP-23 is active on branched xylooligosaccharides. *Applied and Environmental Microbiology*, 82, pp. 5116–5124. https://doi.org/10.1128/AEM.01329-16.

Vera, A., González-Montalbán, N., Arís, A., Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering*, 96(6), pp. 1101-1106. https://doi.org/10.1002/bit.21218.

Verma, D., Kawarabayasi, Y., Miyazaki, K., & Satyanarayana, T. (2013). Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. *PLoS ONE*, 8(1), e52459. https://doi.org/10.1371/journal.pone.0052459.

Vuong, T. V., & Wilson, D. B. (2010). Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnology and Bioengineering*, 107(2), pp. 195–205. https://doi.org/10.1002/bit.22838.

Wagschal, K., Heng, C., Lee, C. C., & Wong, D. W. S. (2009). Biochemical characterization of a novel dual-function arabinofuranosidase/ xylosidase isolated from a compost starter mixture. *Applied Microbiology and Biotechnology*, 81(5), pp. 855–863. https://doi.org/10.1007/s00253-008-1662-4.

Walker, J. A., Pattathil, S., Bergeman, L. F., Beebe, E. T., Deng, K., Mirzai, M., … Fox, B. G. (2017). Determination of glycoside hydrolase specificities during hydrolysis of plant cell walls using glycome profiling. *Biotechnology for Biofuels*, 10(1), pp. 1–19. https://doi.org/10.1186/s13068-017-0703-6.

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., & Wu, W. (2016). Metagenomic analysis of microbial consortia enriched from compost: New insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnology for Biofuels*, 9(1), pp. 1–17. https://doi.org/10.1186/s13068-016-0440-2.

Wang, H., Hart, D. J., & An, Y. (2019). Functional metagenomic technologies for the discovery of novel enzymes for biomass degradation and biofuel production. *Bioenergy Research*, 12(3), pp. 457–470. https://doi.org/10.1007/s12155-019-10005-w.

Wierzbicka-Woś, A., Henneberger, R., Batista-García, R. A., Martínez-Ávila, L., Jackson, S. A., Kennedy, J., & Dobson, A. D. W. (2019). Biochemical characterization of a novel monospecific endo-β-1,4-glucanase belonging to GH family 5 from a rhizosphere metagenomic library. *Frontiers in Microbiology*, pp. 1–19. https://doi.org/10.3389/fmicb.2019.01342.

Wilkens, S. (2015). Structure and mechanism of ABC transporters. *Prime Reports*, 9, pp. 1–9. https://doi.org/10.12703/P7-14.

Wilkens, C., Busk, P. K., Pilgaard, B., Zhang, W. J., Nielsen, K. L., Nielsen, P. H., & Lange, L. (2017). Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. Biotechnology Biofuels, 10, p. 158. https://doi.org/10.1186/s13068-017-0840-y.

Withers, S.G., Warren, A.J., Street, I.P., Rupitz, K., Kempton, J.B and Aebersold, R. (1990). Unequivocal demonstration of the involvement of a glutamate residue as a nucleopjoile in the mechanism of a retaining glycosidase. *Journal of the American Chemical Society* 112(15), pp. 5887-5889.

Wongratpanya, K., Sermsathanaswadi, J. N. T. et al. (2016). Glycoside hydrolase family 43 from *Paenibacillus curdlanolyticus* Strain B-6; An accessory enzyme to enhance the hydrolysis of pretreated rice straw. *Advances in Life Sciences*, 6(1), pp. 1–6. https://doi.org/10.5923/j.als.20160601.01.

Yang, B., Dai, Z., Ding, S. Y., & Wyman, C. E. (2011). Enzymatic hydrolysis of cellulosic biomass. *Biofuels*, 2(4), pp. 421–449. https://doi.org/10.4155/bfs.11.116.

Yang, Y., Zhang, X., Yin, Q., Fang, W., Fang, Z., Wang, X., … Xiao, Y. (2015). A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases. *Scientific Reports*, 5, 17296. https://doi.org/10.1038/srep17296.

Yun, J., & Ryu, S. (2005). Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell Factories*, 4, p. 8. https://doi.org/10.1186/1475-2859-4-8.

Zhao, J., Guo, C., Tian, C., & Ma, Y. (2015). Heterologous expression and characterization of a GH3 β-Glucosidase from Thermophilic Fungi *Myceliophthora thermophila* in *Pichia pastoris*. *Applied Biochemistry and Biotechnology*, 177(2), pp. 511–527. https://doi.org/10.1007/s12010-015-1759-z.

Zang, X., Liu, M., Fan, Y., Xu, J., Xu, X., & Li, H. (2018). The structural and functional contributions of β-glucosidase-producing microbial communities to cellulose degradation in composting. *Biotechnology for Biofuels*, 11(1), pp. 1–13. https://doi.org/10.1186/s13068-018-1045-8.

Zhang, L., Fu, Q., Li, W., Wang, B., Yin, X., Liu, S., … Niu, Q. (2017). Identification and characterization of a novel β-glucosidase via metagenomic analysis of Bursaphelenchus xylophilus and its microbial flora. *Scientific Reports*, 7(1), pp. 1–11. https://doi.org/10.1038/s41598-017-14073-w.

Zhou, Z., Liu, Y., Xu, W., Pan, J., & Luo, Z. (2020). Genome- and Community-Level Interaction Insights into Carbon Utilization and Element Cycling Functions of *Hydrothermarchaeota* in Hydrothermal Sediment. *mSystems*, 5, pp. e00795-19. https://doi.org/10.1128/mSystems.00795-19.

Zoghlami, A., & Paës, G. (2019). Lignocellulosic biomass: Understanding recalcitrance and predicting hydrolysis. *Frontiers in Chemistry*, 7, 874. https://doi.org/10.3389/fchem.2019.00874.