TUM School of Life Sciences

Identification and analysis of novel bacterial enzymes for pectin degradation

Yajing Liu

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer **Doktorin der Naturwissenschaften** genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Wilfried Schwab

Prüfende der Dissertation:

1. Prof. Dr. Wolfgang Liebl
2. Prof. Dr. Johan Philipp Benz

Die Dissertation wurde am 29.03.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 20.06.2022 angenommen.

I Content

II List of figures

III List of tables

IV List of abbreviations

APS	Ammonium persulfate
AA	Auxiliary activity
A_{600}	Absorbance at 600 nm
A_{595}	Absorbance at 595 nm
A_{540}	Absorbance at 600 nm
A_{405}	Absorbance at 600 nm
b/bp	Base/base pair
bar	Pressure (10^5 Pascal)
BLASTP	*Basic Local Alignment Search Tool* for Amino Acid Sequences
BLASTN	*Basic Local Alignment Search Tool* for Nucleic Acid Sequences
CAZymes	Carbohydrate-active enzymes
CBM	Carbohydrate binding module
CE	Carbohydrate esterase
CTAB	Cetyltrimethylammonium bromide
Da	Dalton
ddH_2O	Double demineralized, deionized water
DNA	Deoxyribonucleic acid
DNSA	3,5-Dinitrosalicylic acid
dNTPs	Deoxyribonucleotide triphosphate
EC-number	Enzyme commission number
EDTA	Ethylenediaminetetraacetic acid
eDNA	Environmental DNA
et al.	and others
FPLC	Fast protein liquid chromatography
g	Gram
g	Acceleration due to gravity
gDNA	Genomic DNA
GH	Glycoside hydrolase
GT	Glycosyl transferase
h	Hour
HPLC	High performance liquid chromatography
HPAEC-PAD	High performance anion exchange chromatography with pulsed amperometric detection
I	Liter
IMNGS	Integrated microbial next generation sequencing
ISTD	International standard
LfL	Bayerische Landesanstalt für Landwirtschaft
m	Meter
M	mol/l (Molar)
MALDI-TOF	matrix assisted laser desorption ionization-time of flight mass spectrometry
Min	Minute
Mol	Mol (6.022×10^{23}particular)
MW	Molecular weight
NGS	Next-Generation Sequencing
OTU	*Operational taxonomic units*
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
PCR	Polymerase chain reaction
PL	Polysaccharide lyase

*p*NP	*para*-Nitrophenol
rpm	Rotations per minute
RT	Room temperature
s	second
SDS-PAG	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
T	Temperature
TEMED	Tetramethylethylenediamine
TRIS	Hydroxymethyl aminomethan
TLC	Thin-layer chromatography
UV	Ultraviolet
V	Volt
vs.	versus
v/v	Volume per volume
w/v	Weight per volume
°C	Degree Celsius
Mono-, Oligo- and Polysaccharide	
AG I	Arabinogalactan
Ara	L-Arabinose
AX	Arabinoxylan
AXOS	Arabinoxylooligosaccharide
CMC	Carboxymethylcelluloses
Gal	D-Galactose
GalA	D-Galacturonic acid
Glu	D-Glucose
GL	Galactan (ex.Lupin)
GP	Galactan (ex.potato)
HM Pectin	High methoxyl pectin
LAG	Larch arabinogalactan
LM Pectin	Low methoxyl pectin
Man	Mannose
MeGlcA	4-O-Methly-D-Glucuronic acid
PGP	Pectic galactan (Potato)
PGL	Pectic galactan (Lupin)
PGA	Polygalacturonic acid
Rha	L-Rhamnose
WAX-RS	Wheat arabinoxylan reducing sugar
XOS	Xylooligosaccharide
Xyl	L-Xylose
Size unit	
M	Mega (10^{6})
k	Kilo (10^{3})
c	Centi (10^{-2})
m	Milli (10^{-3})
µ	Micro (10^{-6})
n	Nano (10^{-9})
p	Piko (10^{-12})

V Zusammenfassung

Pektinsubstanzen, die neben Pektin auch Arabinan und Arabinogalaktan enthalten, sind wichtige Bestandteile von Zuckerrübenschnitzeln, die als Klebstoff für Zellulosemikrofibrillen dienen und so die Steifigkeit und Integrität der Pflanzenzellwand verbessern. Die vollständige Verwertung von Pflanzenzellwand-Biomasse erfordert eine möglichst vollständige Hydrolyse von Cellulose, Hemicellulose und Pektinsubstanzen. Um interessante bakterielle Enzyme mit lignocellulolytischer Aktivität zu finden, wurden verschiedene Methoden angewandt. Zu den häufig verwendeten Methoden gehört die Sequenzierung von Genomen einzelner, isolierter Bakterienstämme. Dieser Ansatz stößt jedoch an seine Grenzen, da bei weitem nicht alle Mikroorganismen mit den heute verfügbaren mikrobiologischen Methoden kultiviert werden können. Die Entwicklung der Metagenomik hat dieses Hindernis teilweise überwunden, da sie sich auf die Erforschung des gesamten genetischen Materials von Umweltproben konzentriert. In unseren Studien konzentrierten wir uns auf die Identifizierung von bakteriellen Enzymen mit kultivierungsabhängigen/-unabhängigen Methoden. Darüber hinaus wurde untersucht, wie einzelne Polysaccharide und komplexe Biomasse die Zusammensetzung von Abbaupopulationen in mikrobiellen Konsortien bestimmen.

Bei einem kürzlich beschriebenem neuem, anaerobem, cellulolytischem Bakterium, *Acetivibrio mesophilus* Stamm N2K1, wurde die Abbauleistung an bestimmten Polysacchariden aus Lignocellulosematerial, nämlich an verschiedenen Hemicellulose- und Pektinsubstraten, untersucht. Die hydrolytischen Aktivitäten von N2K1 gegenüber schwer abbaubaren Substraten oder pflanzlicher Biomasse, einschließlich unlöslichem Weizen-Arabinoxylan (WAX), Hafer-Dinkel-Xylan (OSX) und Zuckerrübenschnitzel (SBA), lassen vermuten, dass dieser Organismus eine vielversprechende Quelle für Enzyme für industrielle Anwendungen ist. Es wurden die Gene für vier mutmaßliche Glykosidhydrolasen verschiedener Glykosidhydrolase (GH) Familien (darunter Gene für ein mutmaßliches GH51-Enzym und zwei mutmaßliche GH43-Enzyme aus einem Gencluster, sowie für ein weiteres mutmaßliches GH43-Enzym und ein mutmaßliches GH28-Enzym) heterolog in *Escherichia coli* exprimiert und charakterisiert. Zwei der GHs, *Am*Araf51 und *Am*Araf43, zeigten Arabinofuranosidase-Aktivität. Darüber hinaus zeigte *Am*Araf51 auch β-Xylosidase-Aktivität. *Am*Araf51 und *Am*Araf43 zeigten maximale Aktivität bei 70°C, pH 5,0-6,5 bzw. 57°C, pH 4,0-5,0. Die Messung der spezifischen Aktivitäten dieser beiden Enzyme mit den Polysaccharidsubstraten WAX und SBA zeigte deren

gegensätzliche Präferenzen gegenüber den beiden Arabinose-haltigen Substraten. *Am*Araf51 spaltete vorzugsweise das Arabinan-Polymer, während *Am*Araf43 eine starke Präferenz für Arabinoxylan aufwies. Bei verschiedenen Oligosacchariden als Substraten hatte *Am*Araf51 nicht nur die Eigenschaft, Arabinosereste an O-3-, O-2- und terminalen O-5-Positionen von arabinosylierten Arabinose- oder Xylo-Oligosacchariden zu entfernen, sondern zeigte auch eine exo-β-Xylosidase-Nebenaktivität. Im Vergleich dazu spaltet *Am*Araf43 vorzugsweise die 1,3-Arabinosyl-Seitenkette von doppelt substituierten Xylose- oder Arabinoseresten ab. Die Anwesenheit von *Am*Araf51 stimulierte zusammen mit M_Xyn10 (GH10 Endo-Xylanase, deren Gen aus dem Metagenom einer Biogasreaktorprobe stammt) den erreichten Verzuckerungsgrad beim WAX-RS-Abbau bzw. zusammen mit *Pp*Abn43 (GH43 Endo-Arabinanase aus *Paenibacillus polymyxa* DSM292) beim SBA-Abbau dramatisch, wobei die Ausbeute an Arabinose insbesondere beim WAX-RS Abbau mit der Kombination aus *Am*Araf51 und M_Xyn10 um das 13,77-fache im Vergleich zu *Am*Araf51 allein verbessert wurde. Die Ausbeute an Xylose im Vergleich zu M_Xyn10 alleine stieg fast um das Fünffache.

In dieser Studie wurde weiterhin eine funktionsbasierte Screening-Methode angewandt, um neue, interessante Enzyme mit lignocellulolytischen Aktivitäten zu identifizieren. Mit dieser Methode wurden metagenomische Fosmid-Bibliotheken aus mit SBP angereicherten mikrobiellen Gemeinschaften aus Biogasfermentern, Elefantenkot oder anderen Quellen Polysaccharid-abbauender Bakterien durchforstet, indem selbst hergestellte, CPH- (Chromogenes Polysacchid-Hydrogel) oder ICB- (Unlösliche chromogene Biomasse) markierte Substrate eingesetzt wurden. Durch Illumina-Zufallssequenzierung der Fosmide aus enzymatisch aktiven Fosmidklonen kombiniert mit Sanger-Endsequenzierung der Fosmid-Insertionen (so genannte "Pooled-Sequencing"-Strategie), sowie Annotation der erhaltenen Sequenzen mit dbCAN CAZyme (*C*arbohydrate-*A*ctive *E*nzymes) wurden 71 CAZyme-kodierende Gene aus 55 aktiven Fosmidklonen identifiziert. Schließlich wurde ein ungefähr 56 kbp großes, assembliertes DNA-Fragment, das vermutlich von einem *Xylanivirga thermophila*-Stamm oder einem eng verwandten Stamm stammt, gründlich analysiert. Es enthielt 48 Open Reading Frames (ORFs), von denen 31 für Proteine kodieren, die (mutmaßlich) mit Zuckerbindung und -Transport, Abbau arabinosehaltiger Polysaccharide und Pentosestoffwechsel zu tun haben. Die biochemische Charakterisierung von sieben mutmaßlichen Arabinosylhydrolasen aus den Glykosidhydrolase (GH)-Familien GH51 und GH43 ergab, dass zwei davon Endo-Arabinanase-Aktivität und vier Exo-L-Arabinofuranosidase-Aktivität aufwiesen, aber mit unterschiedlichen, zueinander komplementären

Spaltungseigenschaften. Diese Enzyme arbeiten synergistisch und können SBA und entzweigeschnittenes Arabinan (DA) vollständig hydrolysieren. Auf der Grundlage der Enzymcharakterisierung und der annotierten Funktionen konnte eine Modellvorstellung für den Arabinan-Abbau, die Aufnahme von Spaltprodukten und den Arabinose-Stoffwechsel in dem bisher wenig charakterisierten Bakterium *X. thermophila* vorgeschlagen werden.

Außerdem wurden die katalytische Eigenschaften einer Endo-Xylanase (bezeichnet als M_Xyn10), die vom Fosmid-Klon "BM1T-2 H7" kodiert wurde, untersucht. Die biochemische Charakterisierung von M_Xyn10 zeigte, dass die Hydrolyseprodukte, die aus verschiedenen Xylansubstraten (WAX-I, WAX-RS, BWX) freigesetzt wurden, Xylose, Xylobiose und endständige einfach/doppelt substituierte Xylo-Oligosaccharide umfassten. Mit linearen und verzweigten Xylooligosacchariden als Substrat konnten erstere effizient zu Xylose und Xylobiose abgebaut werden, während aus letzteren Produkte mit terminaler Substitution entstanden. Darüber hinaus zeigte dieses Enzym eine starke Transglykosylierungsaktivität mit *p*NP-Xylobiosid. Die Thermostabilität und das breite Substratspektrum von M_Xyn10 gegenüber verschiedenen Xylan-Substraten könnten für die industrielle Anwendung von Interesse sein.

In der Natur gehören Bakterien zu den Hauptproduzenten von Enzymen zum Abbau von Lignozellulose. Doch in der Regel kann ein einzelner Bakterienstamm Pflanzenbiomasse bzw. schwer abbaubare Anteile der Pflanzenzellwände nicht alleine abbauen. Der Abbau dieser Verbindungen erfordert die synergistische Wirkung mehrerer Bakterien. Um mikrobielle Gemeinschaften mit starken lignocellulolytischen Aktivitäten zu selektieren, wurde eine zweimonatige Anreicherung von mikrobiellen Konsortien auf verschiedenen Substraten (Filterpapier, Xylan, Zuckerrübenschnitzel) nach Animpfung aus einem mit Zuckerrübensilage gefütterten Biogasfermenter der Bayerischen Landesanstalt für Landwirtschaft (LfL) durchgeführt. Durch die Prüfung der Aktivitäten der sekretierten Enzyme und die Analyse der Bakteriengemeinschaft über 16S rRNA-Amplikon-Sequenzierungsdaten der Langzeitanreicherung konnten Informationen über die enthaltenen Xylan-/Zuckerrübenschnitzel-abbauenden Bakterien erhalten werden. Die extrazellulären Enzyme dieser mikrobiellen Gemeinschaften zeigten ein breites Spektrum lignocellulolytischer Aktivitäten gegen Xylan, Arabinoxylan, CMC, Zuckerrübenschnitzel, Arabinan (nur SBP-Anreicherung) und *mixed-linkage* β-Glucan, wobei *Firmicutes* und *Bacteroidetes* die wichtigsten Phyla waren. Bei den Cellulose-abbauenden Bakterien handelte es sich hauptsächlich um *Firmicutes*, das sekretierte Proteom wies

dabei ein relativ enges Spektrum der Substratabbaufähigkeit auf. Insgesamt tragen die Ergebnisse dieser Arbeit zu einem verbesserten Verständnis der an der Verzuckerung wichtiger Bestandteile komplexer Lignozellulose-Biomasse beteiligten bakteriellen Gemeinschaften bei.

VI Abstract

Pectic substances, including pectin, arabinan and arabinogalactan, are significant components of sugar beet pulp, which serves as the adhesive of cellulose microfibrils in plant cell walls, thus improving cell wall rigidity and integrity. Full utilization of plant cell wall biomass requires hydrolysis of as much of the cellulose, hemicellulose and pectic substances as complete as possible. To find interesting bacterial enzymes with efficient lignocellulolytic activity, different methods have been applied. Commonly used methods include a sequence-based approach by sequencing the whole genomes of isolated bacteria. However, it has limitations due to the limited cultivation success for many bacteria from natural habitats with today's microbiological techniques. The development of metagenomics has partially overcome this barrier, which focuses on analysis of the whole genetic material of environmental samples. In our studies, we focused on the identification of bacterial enzymes with cultivation-dependent/independent methods. Further, we investigated how single polysaccharides and complex biomass drive the community structure of degraders in microbial consortia.

The lignocellulolytic degradation ability of *Acetivibrio mesophilus* strain N2K1, which is a recently described novel anaerobic cellulolytic bacterium, its degradative potential towards different hemicellulose substrates and pectic substrates was investigated. The substantial hydrolytic activities of N2K1 on recalcitrant substrates or plant biomass, including insoluble wheat arabinoxylan (WAX-I), oat spelt xylan (OSX), sugar beet pulp (SBA), suggested that N2K1 may be an attractive source of enzymes for industrial applications. Four putative glycoside hydrolases from different glycoside hydrolase (GH) enzyme families (including one putative GH51 and two putative GH43 enzymes encoded by a gene cluster, as well as a further putative GH43 and a putative GH28 enzyme) were heterologously expressed in *Escherichia coli* and characterized. Among them, two GHs, *Am*Araf51 and *Am*Araf43, showed arabinofuranosidase activity. Additionally, *Am*Araf51 also showed β-xylosidase side activity. *Am*Araf51 and *Am*Araf43 exhibited maximum activity at 70°C, pH 5.0-6.5 and 57°C, pH 4.0-5.0, respectively. Analysis of the specific activities of these enzymes against WAX and SBA revealed opposite priorities of both enzymes towards the arabinose-based substrates. *Am*Araf51 preferably cleaved the polysaccharide arabinan, while *Am*Araf43 displayed a strong preference toward arabinoxylan. With various oligosaccharides as substrates, *Am*Araf51 not only had the action mode of removing arabinose moieties at the O-3, O-2 and terminal O-5 positions of arabinosylated arabinose-/xylo-oligosaccharides

but also had exo-β-xylosidase side activity. In comparison, *Am*Araf43 preferably cleaved 1,3-arabinosyl side chains from double substituted xylose or arabinose residues. The presence of *Am*Araf51 dramatically stimulated the saccharification level of AX and SBA along with M_Xyn10 (GH10 endo-xylanase whose gene was isolated from the metagenome of a biogas reactor sample) for WAX-RS degradation or along with *Pp*Abn43 (GH43 endo-arabinanase from *Paenibacillus polymyxa* DSM292) for SBA degradation. To this end, the yield of arabinose, particularly from WAX-RS degradation with the combination of *Am*Araf51 and M_Xyn10, improved 13.77-fold compared with *Am*Araf51 working alone. The xylose yield increased almost five-fold compared with M_Xyn10 working alone.

A function-based screening method was furthermore used in our study to identify novel, interesting enzymes with lignocellulolytic activities. With this method, by using "house-made" CPH (Chromogenic Polysaccharide Hydrogel) or ICB (Insoluble Chromogenic Biomass) labeled substrates, we screened metagenomic fosmid libraries constructed from SBP-enriched biogas fermenter microbial communities, elephant feces, or other sources of polysaccharide-decomposing bacteria. Using a combination of random Illumina sequencing of Fosmids from enzymatically active Fosmid clones and Sanger end-sequencing of the Fosmid inserts (so-called "Pooled-sequencing" strategy), as well as dbCAN CAZyme (carbohydrate-active enzyme) annotation, 71 CAZyme-encoding genes were identified from 55 active fosmid clones. Finally, an around 56 kbp assembled DNA fragment putatively originating from a *Xylanivirga thermophila* strain or a closely related strain was analyzed in detail. It contained at least 48 Open Reading Frames (ORFs), 31 of which encode proteins with (putative) functions in sugar-binding, transport, arabinose-containing polysaccharides degradation and further pentose metabolism. Biochemical characterization of seven putative arabinosyl hydrolases from the GH families GH51 and GH43 revealed two with endo-arabinanase activities and four with exo-L-arabinofuranosidase activities but different, complementary cleavage modes. These enzymes worked synergistically and could completely hydrolyze SBA and debranched arabinan (DA). Based on the enzyme characterization and annotated functions, a potential model for arabinan degradation, transport and arabinose metabolism in the so-far little characterized bacterium *X. thermophila* is proposed.

Besides, the catalytic mechanism of an endo-xylanase (designated M_Xyn10) from fosmid clone "BM1T-2 H7" was investigated. Biochemical characterization of M_Xyn10 showed it released xylose, xylobiose and terminal single/double substituted xylo-oligosaccharides from different xylan substrates (WAX-I, WAX-RS, BWX). With linear and branched xylooligosaccharides as substrates, the former was

efficiently degraded to xylose and xylobiose, while the latter led to hydrolysis products with terminal substitution(s). Furthermore, this enzyme also showed a strong transglycosylation activity with *p*NP-xylobioside. The thermostability and broad cleavage specificity of M_Xyn10 towards different xylan substrates suggested that this enzyme may have potential for industrial application.

In nature, bacteria belong to the major producers of enzymes decomposing lignocellulose. But usually, a single bacterial strain cannot degrade plant biomass and its highly recalcitrant components of plants cell walls alone. The degradation of these compounds necessitates the synergistic action of multiple bacteria. To select microbial communities with strong lignocellulolytic activities, a two-month enrichment of microbial consortia was conducted on different substrates (filter paper, xylan, sugar beet pulp) after inoculation with a sample from a biogas fermenter of the Bayerische Landesanstalt für Landwirtschaft (LfL) fed with sugar beet silage. By testing the activities of the secreted enzymes and by analysis of the bacterial community structure with 16S rRNA amplicon sequencing, long-term enrichment data and information about the xylan-/sugar beet pulp-degrading bacteria were obtained. *Firmicutes* and *Bacteroidetes* were the most prominent phyla, and the secreted enzymes of these microbial communities exhibited a broad range of lignocellulolytic activities against xylan, arabinoxylan, CMC, sugar beet pulp, arabinan (only SBP enrichment) and mixed-linkage β-glucan. The cellulose-degrading bacteria mainly constituted *Firmicutes*, and the secreted enzymes showed a relatively narrow range of substrate degrading ability. Overall, the results from this study contribute to an improved understanding of the bacterial communities involved in the saccharification of the important components of complex lignocellulose biomass.

VII Publications

Publications based on results from this study:

Liu Y, Vanderhaeghen S, Feiler W, Angelov A, Baudrexl M, Zverlov V, Liebl W: **Characterization of two α-L-arabinofuranosidases from *Acetivibrio mesophilus* and their synergistic effect in degradation of arabinose-containing substrates.** *Microorganisms* 2021, 9(7), 1467.

Liu Y, Angelov A, Feiler W, Baudrexl M, Zverlov V, Liebl W, Vanderhaeghen S: **Characterization of metagenome-derived arabinosyl hydrolases from biogas fermenter microbial communities.** (submitted)

Further publications:

Fang R, Lei G, Jiang J, Du H, Liu Y, Lei Z, Ye C, Li N, Peng Y: **High- and low-virulent bovine *Pasteurella multocida* induced differential NLRP3 inflammasome activation and subsequent IL-1beta secretion**. *Vet Microbiol* 2020, **243**:108646.

Fang R, Du H, Lei G, Liu Y, Feng S, Ye C, Li N, Peng Y: **NLRP3 inflammasome plays an important role in caspase-1 activation and IL-1beta secretion in macrophages infected with *Pasteurella multocida***. *Vet Microbiol* 2019, **231**:207-213.

Fang R, Wu R, Du H, Jin M, Liu Y, Lei G, Jiang B, Lei Z, Peng Y, Nie K *et al.*: **Pneumolysin-Dependent Calpain Activation and Interleukin-1alpha Secretion in Macrophages Infected with *Streptococcus pneumoniae***. *Infect Immun* 2017, **85**(9).

Conferences

3rd Global Bioeconomy Alliance, 30th September 2019- 4th October 2019, Sao Paulo and Ubatuba, Brazil, Poster Presentation**: Identification and analysis of bacterial/fungal enzymes for pectin degradation.** Liu Y, Angelov A, Überlacker M, Vanderhaeghen S, Liebl W.

Annual Meeting 2020 of the Association for General and Applied Microbiology (VAAM), 8th March 2020- 11th March 2020, Leipzig, Germany, Poster Presentation: **Metabolic properties of a novel cellulolytic consortium with high potential in lignocellulose degradation.** Rettenmaier R [1], Liu Y [1], Ludwig C, Lombard V, Liebl W, Vanderhaegen S, Zverlov V.

1 Introduction

1.1 Sugar beet pulp

Sugar beet pulp (SBP) is the fibrous residue left over after the extraction of the sugar fraction from the root of sugar beet. It is an energy-rich by-product of industrial production. In 2019, around 278.5 million metric tons of sugar beet were harvested worldwide (https://www.statista.com/statistics/249609/sugar-beet-production-worldwide/, access on 29.07.2021). After juice extraction, it leads to 17 million tons of dehydrated beet pulp (Spagnuolo *et al.*, 1997). As an agricultural waste, SBP is mainly disposed of as animal feed in countries with widespread cattle-raising manufacturing. In other counties, it is discarded in landfills (Hutnan *et al*., 2000). However, SBP can serve as a valuable, sustainable resource and its biotransformation is of great importance in industrial applications. The dried SBP comprises 19-23% glucose, 16-24% arabinose and 19-23% of galacturonic acid (Kracher *et al.*, 2014; Leijdekkers *et al.*, 2013). The different heterogeneous sugar contents in SBP form a complicated macromolecular structure (Table 1.1). The cellulose microfibrils contribute to the rigidness of the cell wall, while pectic substances and hemicellulose serve as the adhesion agent for the cellulose network. The intertwined polysaccharide chains increase the rigidity of SBP, which makes the enzymatic hydrolysis of SBP for its use as a feedstock for biotechnological production processes such as the production of advanced biofuels a significant challenge (Oumer, 2017).

Table 1.1 Chemical composition of sugar beet pulp (SBP).

Component	Content (g kg^{-1})
Dry matter	210.1 ± 1.2*
Total protein (N·6.25)	88.4 ± 0.5
Diethyl ether extractable substance	55.2 ± 0.5
Ash	23.7 ± 0.4
Hemicellulose	272.3 ± 2.4
Cellulose	140.0 ± 2.3
Pectin	134.8 ± 0.8
Lignin	88.6 ± 3.4
Remaining sucrose as saccharide after inversion	11.2 ± 1.2
Monosaccharides **:	
Arabinose	190.4 ± 0.5
Glucose	189.3 ± 0.9
Xylose	16.6 ± 1.0
Sum of galactose, mannose and rhamnose	80.1 ± 2.8

Component	Content (g kg^{-1})
Uronic acids	188.3 ± 1.3
Methanol	4.2 ± 0.9
Acetic acid	1.7 ± 0.9

* Standard deviation. ** Determined in liquid fraction obtained after sulfuric acid hydrolysis of SBP. The chemical composition of SBP is shown, taken from Pińkowska *et al.*, (2019).

Enzymatic saccharification of SBP, rather than chemical degradation, has a promising prospect since it is eco-friendly, technologically feasible and economically beneficial (Micard *et al.*, 1996; Spagnuolo *et al.*, 1997). The complete degradation process is highly specific when using defined biocatalysts, and the energy required to maintain mild enzymatic reaction conditions is relatively low. Thus, it results in a low cost for the industrial process, while avoiding chemical waste (Micard *et al.*, 1996; Spagnuolo *et al.*, 1997). An earlier study has shown that SBP was hydrolyzed into soluble sugars by the synergism of pectinase and cellulases. Pectinase can disintegrate the cell wall structure, thus generating more permeable matrix that allows cellulase to access crystalline and amorphous structures (Micard *et al.*, 1997). Furthermore, Leijdekkers *et al*. showed that combining Viscozyme L (a commercial enzyme mixture from *Aspergillus aculeatus*) with an experimental pectinase (from *Aspergillus niger*) was able to release 82% arabinose, 79% mono-galacturonic acid and 17% glucose from SBP (Leijdekkers *et al.*, 2013). Although much improvement has been achieved in enzymatic synergism for SBP degradation, there are still uncertainties regarding important aspects of industrial utilization of enzyme cocktails, such as unclear cleavage mechanisms of different enzymes on recalcitrant biomass. Therefore, more efforts are required to fully understand the whole hydrolysis process. Additionally, more knowledge about the structure of each component of SBP, especially pectic substances, and the specificity of various carbohydrate-active enzymes has to be developed.

1.2 Structure and enzymatic hydrolysis of the major components of SBP

1.2.1 Cellulose and cellulolytic enzymes

Cellulose is a simple polysaccharide composed of β-1,4-linked glucose subunits. Each glucose shows a steric 180° rotation relative to its neighboring glucose resides, so the constitutional unit in cellulose is the cellobiose. The length of the cellulose backbones can vary between 100 and 14,000 residues (Biswabandhu *et al.*, 2019). In the plant cell walls, cellulose usually forms several layers of the sheets organized in parallel. The spaces between the “sheets” are embedded in other polymers, including

hemicellulose (mainly xylans and glucomannans), lignin and pectin (Pierre Beguin, 1994). Even though cellulose has a chemically relatively simple structure, it still displays resistance to enzymatic hydrolysis due to its insolubility, crystalline structure, and heterogeneous nature (Lakhundi *et al.*, 2015).

Given these characteristics, it is difficult for a single GH to clasp cellulose into its cleavage site directly. Therefore, at least three different enzymes, including endo-glucanase activity, exo-glucanase activity (also called cellodextrinase or cellobiohydrolase) and β-glucosidases activity, are required for complete degradation of polymeric cellulose into its monomeric unit glucose (Den *et al.*, 2007; Lynd *et al.*, 2002; Teeri, 1997). According to the Carbohydrate-Active Enzymes database (http://www.cazy.org/), cellulase can be classified into 115 GH families based on amino acid sequence similarities and crystal structures. Endo-glucanases randomly cleave internal bonds of the cellulose backbone. Thereby the exposed new chain ends are more accessible for the exo-glucanases. Cellobiohydrolases, for example, bind the chain end and break every alternative glycosidic linkage, resulting in cellobiose (Shrotri *et al.*, 2017). In comparison, β-glucosidase can degrade cellobiose to glucose from the non-reducing ends (Figure 1.1) (Ratnayake, 2019). The structures of cellulases include catalytic modules, which have been classified into various GH families. Non-catalytic modules like carbohydrate-binding modules (CBMs) and/or other functionally known or unknown modules located N- or C- terminally to the catalytic module can improve the binding ability of the enzyme. For example, Nakamura *et al*. found two cellobiohydrolases produced by the fungus *Trichoderma reesei* (TrCel6A) and the bacterium *Cellulomonas fimi* (CfCel6B) share a similar catalytic domain (CD) but have different cellulose binding domain (CBDs) and linkers (between CD and CBDs). The CBDs in CfCel6B and linker in TrCel6A are mainly involved in the initial substrate binding (Nakamura *et al.*, 2020). After enzymatic hydrolysis of cellulose, the resulting glucose can be used as a feedstock for microbial fermentation processes, e. g., it can be fermented to ethanol, isopropanol, or butanol. Therefore, cellulose is an attractive renewable carbon source. Its bioconversion is of significant importance in biotechnology industries (Pierre Beguin, 1994).

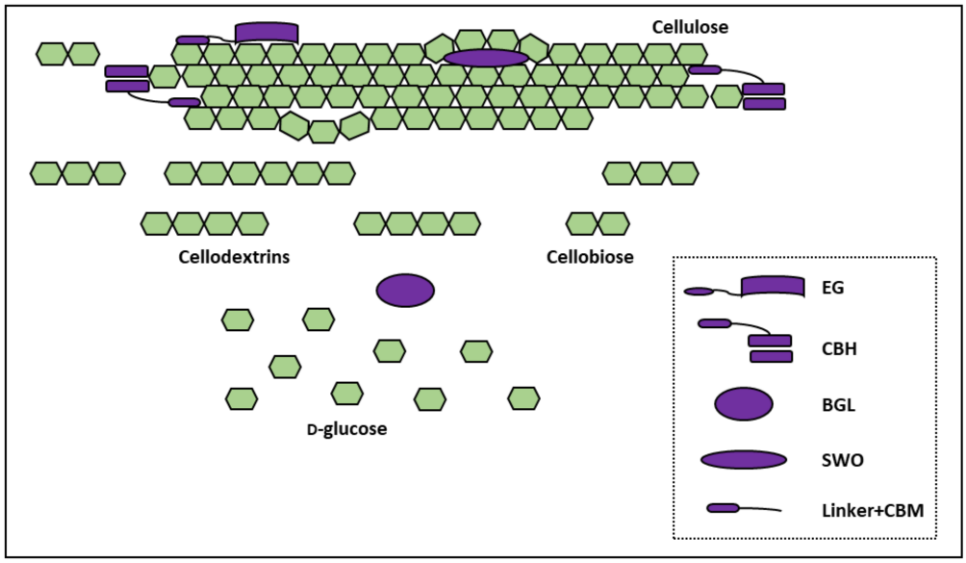


Figure 1.1 Hydrolysis of cellulose by cellulases and non-catalytic proteins of *Trichoderma reesei*. Multiple members of the different types of cellulase enzymes, including CBHs (cellobiohydrolase), EGs (endo-glucanase) and BGLs (β-glucosidases), degrade the crystalline cellulose synergistically to glucose. CBHs and some of the EGs are composed of a catalytic domain and a smaller carbohydrate-binding module (CBMs) connected via a flexible linker region. In addition, non-enzymatic proteins such as swollenin (SWO) are involved in the degradation of cellulose by disrupting the crystalline structure and thus increasing the accessibility of cellulose for enzymatic proteins. The degradation of cellulose is shown, based on Seiboth *et al.* (2011).

In plants, cellulose primarily serves as a structural polysaccharide. Cellulose forms crystalline microfibrils to maintain the plant's cell wall strength, which enables plant cells to resist osmotic pressure and mechanical stress. Commercially available pure cellulose includes filter paper, cotton and avicel. They have different physical heterogeneity regarding the pore size, available surface area and degree of crystallinity. Thus, they can be used to assess the enzymatic efficiency of cellulases. In comparison, carboxymethylcellulose (CMC), a derivative of cellulose produced by digesting the cellulose with chloroacetic acid and sodium hydroxide, has an amorphous structure (Anita *et al.*, 2013). Due to its solubility and easily digestible property, CMC is often used for assays for endo-glucanases (Pierre Beguin *et al.*, 1994). Additionally, artificial chromogenic and fluorogenic substrates, such as *p*-nitrophenyl-β-D-cellobioside and methylumbelliferyl-β-D-cellobioside (MUC) have been developed which are convenient substrates for enzymological studies. Chromogenic or fluorogenic compounds can be used to dye polysaccharides or to be linked to the anomeric carbon of cellobiose or cellodextrins by a β-glucosidic bond (van Tilbeurgh *et al.*, 1982).

1.2.2 Hemicellulose and hemicellulolytic enzymes

Hemicellulose is mostly a heterogenic polysaccharide and accounts for up to 30% of lignocellulose components. It has a smaller molecular weight but a relatively more complex structure than cellulose since the backbone of hemicelluloses can be substituted by various sugar chains and other substituents, usually unseen in cellulose. Compared with cellulose, which is only comprised of hexose (glucose), hemicelluloses can be composed of multiple sugar monomers such as pentoses (xylose and arabinose), hexoses (mannose and galactose) and the six-carbon deoxy sugar rhamnose (Perez *et al.*, 2002) (Figure 1.2). According to the structural varieties such as backbone linkages and the attached side chains, hemicelluloses can be subdivided into four types, including xylans (arabinoxylan, beechwood xylan, birchwood xylan), xyloglucans, mixed linkage β-glucan and mannans, which vary in abundance as well as distribution in different plants (Heinze, 2005). Xylan has a backbone of β-1,4-linked xylose residues or occasionally of β-1,3- or mixed β-1,3- and β-1,4- glycosidic bonds. Xylans only consisting of D-xylopyranose residues are designated homoxylans. In contrast, glucuronoxylans and glucuronoarabinoxylans belong to heteroxylans. They always have the same backbone of D-xylopyranose residues but different sugar chain substituents. For example, glucuronoxylan has a backbone of β-1,4-linked xylopyranose residues decorated with 4-*O*-methyl glucuronosyl residues and α-1,2-linked glucuronosyl residues (Heinze *et al.*, 2012). Arabinoxylan has the same backbone, substituted with arabinofuranosyl side chains at O-2 or O-3 of the xylose residues (Barron *et al.* 2006). The arabinofuranosyl side chains can be further esterified by aromatic acids such as *p*-coumaric acid and ferulic acid (Smith *et al.*, 2017). Besides, other groups such as the acetyl group can also modify the xylan backbones to modify the properties of the plant cell walls (Smith *et al.*, 2017).

Xylan biodegradation requires the concerted action of different enzymes. Endo-1,4-β-xylanase cleaves the xylan backbone, thus generating xylooligosaccharides. Exo-β-1,4-xylosidase can cleave the glycosidic bonds from reducing/non-reducing ends, thus releasing xylose. Due to the presence of various side-chain substitutions, more enzymes, such as feruloyl esterases, α-L-arabinofuranosidases and glucuronosidases, are also required to fully hydrolyse heteroxylan (Figure 1.2). Xylanases are grouped into families GH5, GH7, GH8, GH10, GH11 and GH43 in the CAZyme database. Among all these families, GH10 and GH11 xylanase are the most studied enzymes. It has been found that GH10 xylanase have relatively high activity towards short xylo-oligosaccharides, which suggests that these enzymes usually have a small substrate-binding site (Biely *et al.*, 1997a). Various studies showed that family

GH10 xylanases commonly have four to seven substrate-binding positions in the active-site cleft (Derewenda *et al.*, 1994; Biely *et al.*,1997a; Biely *et al.*, 1981), Highly conserved -2, -1 and +1 subsites play an important role in substrate binding. However, the cleavage site usually takes place at the second xylosidic bond after a 1,3-β glycoside bond present in rhodynenan (β-1,3-β-1,4-xylan) with isomeric xylotriose (Xylβ1-3Xylβ1-4Xyl) as final products or at the third xylosidic bond after the substituted side chain, the latter one can be proved by the presence of aldotetrauronic acid as the hydrolysis product from glucuronoxylan (Biely *et al.*,1997a). In contrast to GH10, GH11 xylanases can hydrolyze xylan into larger products than GH10 xylanases. Presumably, the enzyme activity is restricted by the hindrance of β-1,3-linkages or substituents (Biely *et al.*, 1997a). Therefore, hydrolysis of glucuronoxylan by GH11 xylanase can generate aldopentauronic acid as product, with a terminally unsubstituted xylose residue at the non-reducing end (Biely *et al.*, 1997a; Christakopoulos *et al.*, 2003).

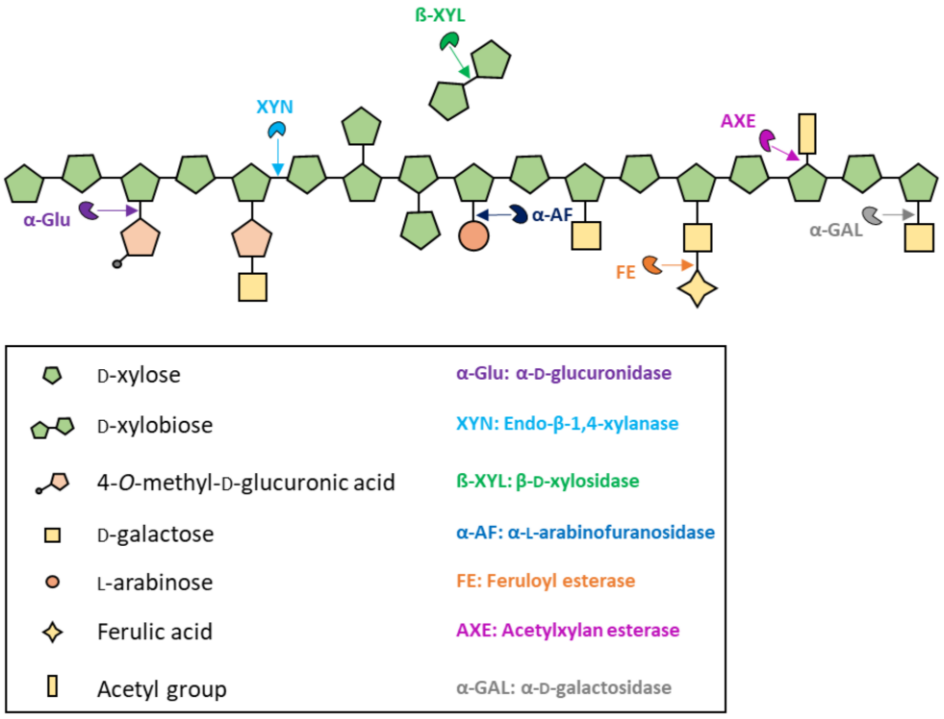


Figure 1.2 Model of enzymatic degradation of hemicellulose. The schematic model of enzymatic hemicellulose degradation is shown, based on Madadi *et al.* (2017).

1.2.3 Pectin and pectinolytic enzymes

As one of the significant components of SBP (19-23%), pectin has the most intricate structures and functions. It is embedded in the junction zones between plant cells, together with hemicellulose

molecules. They play a critical role in integrating the cellulose microfibrils, producing a multi-molecular complex that provides strength to the plant cell wall. Besides contributing to the physiological characteristics of plant cell wall biomass, pectin also exhibits other functions for plant growth, such as maintain the rigidity and integrity of the macromolecular structure of plant tissues. Pectin also plays important roles in plant development, defense, cell expansion, morphogenesis, signaling, binding of ions, wall porosity, fruit development, seed hydration, or leaf abscission (Mohnen, 2008; Ridley *et al.*, 2001; Willats *et al.*, 2001).

Pectic substances are composed of pectin and pectic acid. The former primarily has a backbone of partially methyl-esterified-1,4-D-galacturonan, while pectic acid, also known as polygalacturonic acid, is a demethylated form of pectin (Hoondal *et al.*, 2002). According to structural differences, pectic polysaccharides can be classified into homogalacturonan (HG), rhamnogalacturonan I (RG-I), rhamnogalacturonan II (RG-II) and xylogalacturonan (XGA) (Mohnen, 2008). HG accounts for around 65% of pectin. It has the main chain of α-1,4-linked galacturonic acid, which is esterified with methyl groups at the C-6 carboxyl functions, and may also be acetylated at O-2 or O-3 positions (O' Neill *et al.*, 1990). RG-I is a structurally more variable pectic polysaccharide with a backbone of alternating disaccharide rhamnose and galacturonic acid. Around half of the rhamnosyl residues are modified with acidic and perhaps neutral oligosaccharides at the C-4 position (Figure 1.3). In comparison, RG-II has a complex structure with the same backbone as HG, but the C-2 and/or C-3 positions are decorated with various keto- and aldehyde-sugar oligosaccharide side chains (O' Neill *et al.*, 1990). The complex structure of pectic polysaccharides gives the compounds their characteristic degradation resistance.

So far, pectinolytic enzymes have been found from bacteria (including *Bacillus*, *Pseudomonas*, *Xanthomonas* and actinomycetes such as *Amycolata* and *Streptomyces*), fungi (mainly *Aspergillus*), and plants (Hoondal *et al.*, 2002). Based on the difference in substrate specificity (pectin, pectic, or oligomers of galacturonate) and cleavage mechanism (trans-elimination or hydrolysis, endo- or exo- activity) (Alkorta *et al.*, 1998), pectin-cleaving enzymes are divided into four classes: (i) pectinesterase (PE) can remove the methoxyl group at C6 of GalA which results in pectic acid; (ii) polygalacturonase (PGases) is able to hydrolyze the backbones of the α-1,4-linked polygalacturonic acid chain in non-methylated stretches of homogalacturonan; (iii) pectate lyase (PGL) specifically breaks down the glycosidic linkages on polygalacturonic acid and then generates unsaturated galacturonate oligomers by trans-elimination of

glycosidic bonds. (iv) pectin lyase has a similar catalytic mechanism with PGL but strongly prefers high esterified pectin (Oumer, 2017). Furthermore, since pectin is mostly modified by highly branched side chains that include neutral sugars like arabinose and galactose, eliminating the side chains of the pectin might enhance the accessibility of pectinase to the backbones. Based on the types of side chains, arabinanase, arabinofuranosidase, galactanase, β-galactosidase and feruloylesterases are required in the hydrolysis process (Figure 1.3, Table 1.2). Understanding the pectin degradation mechanism for using pectin-rich plant biomass such as citrus peel and sugar beet pulp is of great importance. The released fermenter sugars can produce high-value products like ethanol and aroma compounds (Edwards *et al.*, 2012).

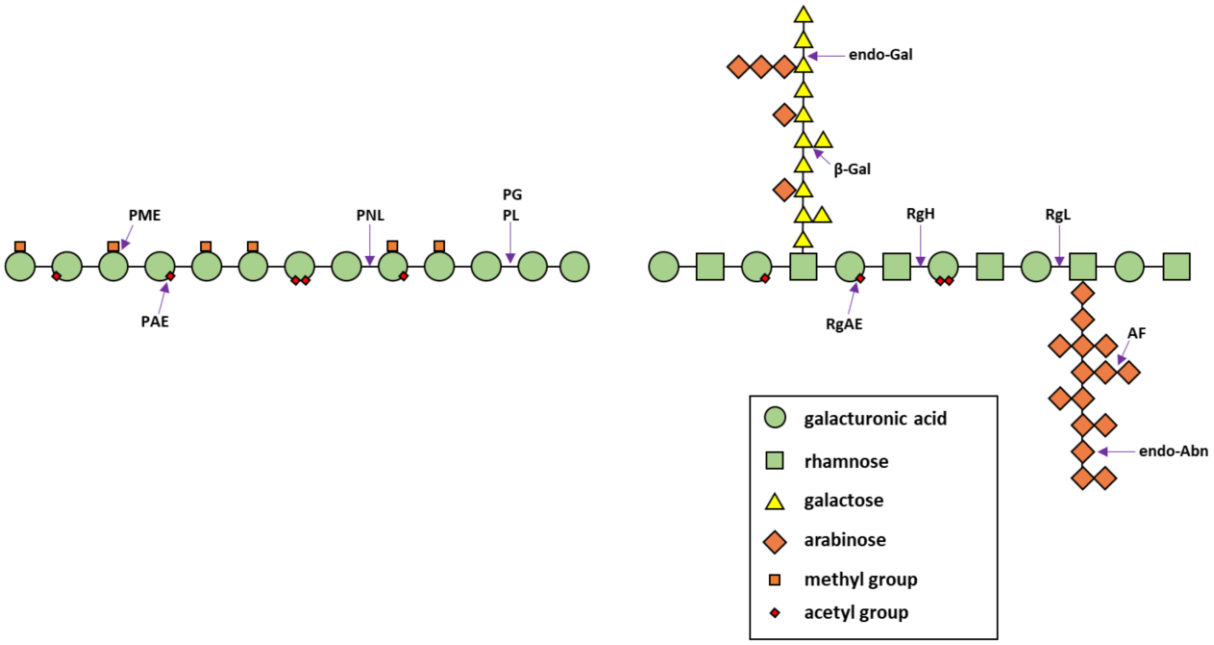


Figure 1.3 Structure of (a) homogalacturonan and (b) rhamnogalacturonan I and their degrading enzymes. ***PG*** polygalacturonase (endo/exo), ***PL*** pectate lyase (endo/exo), ***PNL*** pectin lyase, ***RgH*** RG-hydrolase, ***RgL*** RG-lyase, ***RgAE*** RG-acetyl esterase, ***PME*** pectin methylesterase, ***PAE*** pectin acetylesterase, ***endo-Abn*** endo-arabinanase, ***AF*** arabinofuranosidase, ***endo-Gal*** endo-galactanase, **β-*Gal*** β-galactosidase. Structure information and enzyme activities are based on Bonnin *et al.* (2020).

Table 1.2 Enzymes involved in pectin degradation.

Enzymes	CAZy[a] family	Substrate[b]	Product
Endo-polygalacturonase	GH28	HG-LM Pectin	OligoGalA
Exo-polygalacturonase	GH28	HG-LM Pectin	GalA
Exo-α-D-galacturonosidase	GH28	HG-LM Pectin	GalA2
Rhamnogalacturonan hydrolase	GH28	RG-L-Pectin	OligoRG
Rhamnogalacturonan rhamnohydrolase	GH28	RG-L-Pectin	Rhamnose
Rhamnogalacturonan galacturonohydrolase	GH28	RG-L-Pectin	GalA
Xylogalacturonanhydrolase	GH28	Xylogalacturonan	Xylosylated OligoGalA
Endo-(1,4) galactanase	GH53	1,4-Galactan	1,4 Galactooligosaccharides
Endo-(1,3) galactanase	GH16	1,3-6 Galactan	1,3-6 Galactooligosaccharides
Endo-(1,6) galactanase	GH5-30	1,3-6 Galactan	1,3-6 Galactooligosaccharides
β-galactosidase	GH1-2-35-42	Galactooligosaccharides	Galactose
Endo-arabinanase	GH43	Arabinan	Arabinooligosaccharides
α-L-arabinofuranosidase	GH43-51-54-62	Arabinooligosaccharides	Arabinose
Endo-pectate lyase	PL1-2-9	HG	Unsaturated OligoGalA
Exo-pectate lyase	PL1-2-9	HG	Unsaturated OligoGalA
Endo-pectin lyase	PL1	HM-Pectin	Unsaturated MeOligoGalA
Rhamnogalacturonan endo-lyase	PL4	RG-I	Unsaturated OligoRG
Rhamnogalacturonan exo-lyase	PL11-26	RG-I	Unsaturated OligoGalA
Pectin methyl esterase	CE8	Pectin	Pectic acid + methanol
Pectin acetyl esterase	CE12	Pectin	Pectic acid + acetic acid
Rhamnogalacturonan acetyl esterase	CE12	Pectin	RG-I + acetic acid
Feruloyl esterase	CE1	Feruloylated Oligosaccharides	Oligosaccharides + ferulic acid

[a]*GH* glycosyl hydrolase, *PL* polysaccharide lyase, *CE* carbohydrate esterase,

[b]*HG* homogalacturonan, *RG* rhamnogalacturonan, *LM* low methoxyl, *HM* high methoxyl, GalA galacturonic acid. Data from Bonnin *et al.* (2020).

1.2.4 Other pectic polysaccharides (arabinan and arabinogalactan) and their degradation

Arabinan and galactan/arabinogalactan, as the side-chain substitutions of RG-I, are parts of pectin substances (Broeker *et al.*, 2018). Earlier reports showed that many pectic polysaccharides isolated from plant cell walls had a high arabinan content (O'Neill *et al.*, 1990). The linear backbones of debranched arabinan (DA) consist of α-1,5-linked arabinofuranosyl residues, which can be substituted with arabinofuranosyl residues at O-2 and/or O-3 positions, such as in sugar beet arabinan (SBA). Specifically, for the SBA used in our study, 60% of backbone residues carry 1,3-linked arabinosyl side-chain substitutions while less frequently 1,2-arabinosyl decorations occur (Cartmell *et al.*, 2011).

The hydrolysis of arabinans requires the synergism of different types of arabinosyl hydrolases (Velasco *et al.*, 2020). Alpha-arabinosyl hydrolases include endo-α-L-1,5-arabinanases and exo-α-L-arabinofuranosidases. Endo-arabinanases are only from family GH43 under subfamily 4, 5, 6, which randomly cleave the arabinan backbones (Wefers *et al.*, 2017), while exo-α-L-arabinofuranosidases are present in families GH2, GH3, GH5, GH39, GH43, GH51, GH54, GH62 (http://www.cazy.org/) and are able to cleave 1,2-, 1,3- or terminal 1,5-linked arabinosyl side-chain substitutions in the arabinose-containing substrates such as arabinan and arabinoxylan (http://www.cazy.org/).

Specifically, with arabinoxylan (AX) as a substrate, arabinofuranosidases can be classified into different groups according to their cleavage specificities (Beldman *et al.*, 1997). Arabinofuranosidase enzymes of the type AXHs-m cleave off arabinose moieties at O-2 or O-3 positions of mono-arabinosyl substituted main-chain xylose residues. It is also the most common form of arabinofuranosidase and can be found in most of the mentioned GH families (Biely *et al.*, 2016; Mechelke *et al.*, 2017). In comparison, AXHs-m,d -type enzymes degrade side-chain modifications from mono- and double-substituted xylose residues (Koutaniemi *et al.*, 2016; Mroueh *et al.*, 2019). AXHs-d arabinofuranosidases remove α-1,2- or α-1,3-linked arabinofuranosyl side-chain substitutions from either terminal or internal double arabinosyl substituted xylose backbone residues. For this enzyme type, more specific cleavage actions can be described by introducing an additional number, such as AXH-d3 represents the specific cleavage mode towards 1,3-linked arabinosyl residues from double substituted Xylp motifs (Koutaniemi *et al.*, 2016; Van den Broek *et al.*, 2005). Besides, exo-α-L-1,5-arabinanases, which release arabinose, arabinobiose or arabinotriose from linear arabinan, have been described in the family GH93 and with some representatives in family GH43 (Carapito *et al.*, 2009; Kaji

et al., 2014; Kühnel *et al.*, 2010; Wong *et al.*, 2008). By contrast, β-arabinohydrolases are not commonly studied. So far, few β-arabinohydrolases were described in family GH27 and GH127, family GH27 β-arabinohydrolases were capable of removing β-L-arabinopyranose residues in complexed arabinan side-chain substitutions, while GH127 can cleave β-arabinofuranose residues (Lansky *et al.*, 2014; Lansky *et al.*, 2013).

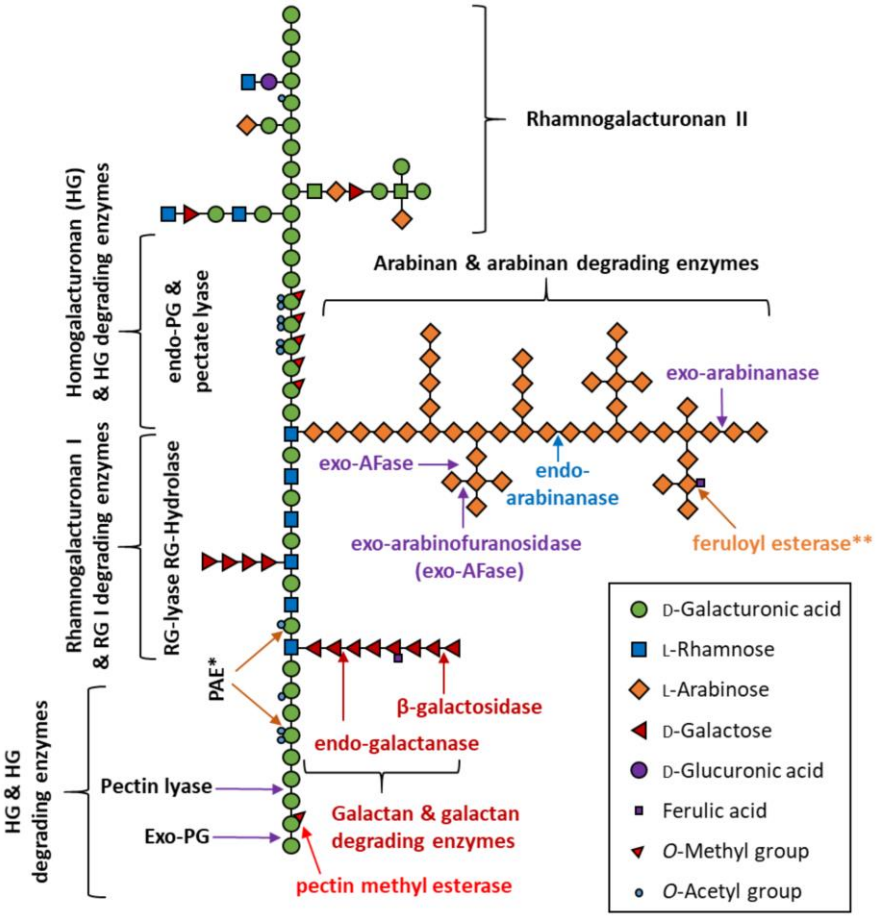


Figure 1.4 General structure of sugar beet pectin and enzymes required for its degradation. Endo-PG: endo-polygalacturonase, exo-PG: exo-polygalacturonase, PME: pectin methylesterase, * active on HG and RG I, **active on arabinan and galactan. Figure is based on Kühnel *et al*. (2011).

As another neutral sugar from side chains of the pectic substrates, galactan contributes to maintaining the correct organization of the different components of the plant cell wall during the termination of growth and the early stages of secondary wall development (Moneo-Sanchez *et al.*, 2020). Galactan side chains in rhamnogalacturonan I are composed of 1,4-linked β-D-Gal*p* units. Other elements, such as feruloyl groups at the O-6 positions and arabinofuranosyl residues, can further modify the Gal*p* units

of galactan (Colquhoun *et al.*, 1994; Ralet *et al.*, 1994; Huisman *et al.*, 2001) (Figure 1.4). Along with the arabinan side chains, galactan increases the interactions between pectin and cellulose in plant cell walls (Zykwinska *et al.*, 2007). Besides serving as the side chains in rhamnogalacturonan-I, another type of galactan with β-1,3-galactan backbones or oligo-β-1,6-galactosyl side chains that substituted with different sugars such as D-glucuronic acid, 4-O-methyl-D-glucuronic acid, L-arabinofuranose, L-arabinopyranose, L-rhamnose, L-fructose also exists. This type of galactan is usually covalently linked to proteins (Knox, 1995; Ponder *et al.*, 1997). Galactosyl linkages can be cleaved by β-galactosidase, endo-galactanase, exo-galactanase (Sakamoto *et al.*, 2013), which are classified into families GH5, GH16, GH30, GH35, GH43 and GH53 (http://www.cazy.org/).

1.3 Microbial strategies for polysaccharide degradation and utilization

1.3.1 Carbohydrate active enzymes (CAZyme)

Degradation of plant biomass is a great challenge due to the complex and partially recalcitrant structures of each component. Various glycosidic bonds can link the building blocks of monosaccharides. Thus, a broad spectrum of cell wall-degrading enzymes is required. Microorganisms, mainly fungi and bacteria, have developed sophisticated strategies for deconstructing lignocellulosic components over time, paralleling the evolution of plant cell walls. A well-studied approach is the secretion of carbohydrate-active enzymes (CAZymes), including glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), plus a range of auxiliary enzymes (AAs). Each group can be further divided into various families based on their catalytic activities, structural and functional amino acid sequences similarities (http://www.cazy.org/) (Lombard *et al.*, 2014). GHs are the largest group of carbohydrate-active enzymes, which can cleave the glycosidic bonds between two monosaccharides or between a monosaccharide and a non-sugar group. So far, GHs in the CAZyme database are further divided into 171 families (http://www.cazy.org/Glycoside-Hydrolases.html). Family GH43 is one of the most prominent GH families, which includes 4,555 family members. The major activities reported in this GH family are β-D-xylosidase, α-L-arabinofuranosidase, endo-α-L-arabinanase, and 1,3-β-galactosidase. They are involved in the degradation of backbones or side chains of hemicellulose (arabinoxylan) and pectic substances (Mewis *et al.*, 2016). The enzymes in GH43 have a 5-bladed β-propeller structure and adopt an inverting mechanism for cleaving the glycosidic bonds in polysaccharides. Besides GH43, family GH51 enzymes

are also important candidates for biotechnological processes due to their stability against harsh industrial production conditions. GH51 enzymes can form hexamers in solution and have a $(\beta/\alpha)_8$-barrel architecture. An additional β-sandwich domain with unknown function is usually co-linked with the $(\beta/\alpha)_8$-barrel structure. Many studies have shown that the $(\beta/\alpha)_8$-barrel fold is very stable (Höcker *et al.*, 2001). So far, many GH51 arabinofuranosidases were reported with the thermo-resistance, as are shown in Table 1.3.

Table 1.3 Top thermostable arabinofuranosidases among various glycoside hydrolase families.

Organisms	GH family	pH_{opt}	T_{opt}	Thermostability	References
Thermotoga thermarum DSM 5069	GH2	5.0	80 °C	Half-life of more than 2 h at 80°C	Shi *et al.* (2014).
Thermotoga thermarum DSM 5069	GH3	6.0	95 °C	Half-life of 2 h at 85°C	Shi *et al.* (2013).
Hungateiclostridium thermocellum ATCC 27405	GH43	5.7	50 °C	Stable for at least 15 min at 50°C	Ahmed *et al.* (2013).
Humicola insolens Y1	GH43	5.0	50 °C	Half-life of approximately 5 min at 50°C	SØrensenet *et al.* (2006).
Hungateiclostridium clariflavum DSM 19732	GH51	6.5	60 °C	Half-life of more than 7 days at 60°C	Geng *et al.* (2019).
Hungateiclostridium thermocellum ATCC 27405	GH51	7.0	82 °C	Stable for at least 1 h at 82°C	Taylor *et al*. (2006).
Thermobacillus xylanilyticus D3	GH51	6.0	75 °C	Half-life of 2 h at 90°C	Debeche *et al.* (2000).
Geobacillus vulcani GS90	GH51	5.0	70 °C	Stable for at least 1 h at 70°C	Ilgü *et al.* (2018).
Geobacillus thermodenitrificans NG80-2	GH51	6.5	60 °C	Half-life of 5 days at 75°C	Huang *et al.* (2017)
Penicillium funiculosum 8/403	GH54	2.6	60 °C	Half-life of approximately 12 min at 60°C	De La Mare *et al.* (2013)
Aspergillus kawachii IFO4308	GH54	4.0	60 °C	Stable for at least 1 h at 60°C	Koseki *et al.* (2006)
Aspergillus fumigatus wmo	GH62	5.0	42 °C	Stable for at least 1 h at 42°C	Pérez and Eyzaguirre *et al.* (2016)
Streptomycesthemoviolaceus subsp. Pingens Henssen	GH62	7.0	55 °C	Half-life of approximately 1 h at 55°C	Wang *et al.* (2014)

Information on each enzyme is derived from Geng *et al.* (2019).

Besides, microorganisms can also employ multifunctional enzymes for the efficient deconstruction of polymeric networks in plant cell walls, this kind of enzymes always contain two or more catalytic modules. Guerriero *et al.* analyzed the modular structures of 16,937 genes putatively encoding enzymes from 34 GH families. Among all these genes, 64 genes encoding putative CAZymes may harbor multi-functions, and up to five catalytic modules have been found in a single ORF (Sarai *et al.,* 2019). According to the polypeptide sequence, multi-function enzymes can be divided into four different classes: hemicellulase-hemicellulase, cellulase-cellulase, cellulase-hemicellulase and hemicellulase-carbohydrate esterase (Xu *et al.*, 2011). The use of multifunctional enzymes as a low-cost method of decomposing lignocellulosic biomass into fermentable sugars might be a valuable option in the future.

1.3.2 Polysaccharide utilization loci (PULs)

Based on known databases, CAZyme repertoires can be easily predicted from the genomes of different bacteria. Besides, they are usually clustered such as in the polysaccharide degrading loci (PULs) of some Gram-negative bacteria. These strictly coregulated, colocalized genes encoding enzymes allow the concerted degradation of polysaccharides to oligosaccharides or monosaccharides (Grondin *et al.*, 2017). The genomes of *Bacteroidetes* can exemplify these elegant systems. PULs in the genomes of *Bacteroidetes* can encode substrate-binding proteins, TonB-dependent transporters, carbohydrate sensors/transcriptional regulators and CAZymes (mostly GHs) (Grondin *et al.*, 2017). All these components constitute the polysaccharide utilization system of *Bacteroidetes* and improve the efficiency of degradation of complex polysaccharides. The released oligosaccharides can be used by both decomposers and the species that surround them. Rogowski and Cuskin *et al*., respectively, set up the synergistic actions of two PUL models. The first model showed the sharing of the xylo-oligosaccharides with *Bifidobacterium adolescentis* released from wheat arabinoxylan by *Bacteroides ovatus* (Rogowski *et al.*, 2015). In contrast, the second model displayed the selfish mechanism adopted by *Bacteroides thetaiotaomicron* with α-mannan degradation. The α-mannan breakdown products may be quickly transported into the periplasm in this model, where the oligosaccharides can be further digested into monosaccharides for growth (Cuskin *et al.*, 2015). The concerted actions of both PUL models benefit the colonization of nutritional niches and the establishment of microbial systems.

1.3.3 Cellulosome complexes

Certain anaerobic microbes, such as the efficient cellulolytic bacterium *Acetivibrio thermocellus* (syn.

Clostridium thermocellum), have developed specifically efficient machineries for the extracellular deconstruction of plant polysaccharides by forming cellulosomes (Bayer *et al.*, 2004). Cellulosomes are multi-enzyme complexes that integrate various cellulases, hemicellulases and/or accessory enzymes by assembling the dockerin domain of each protein to the non-catalytic cohesion domains of scaffolding structures (Hyeon *et al.*, 2013). The known cellulosomes which so far have been studied were shown to have cellulase, xylanase, mannanase, arabinofuranosidase, lichenase and pectin lyase activities (Bayer *et al.*, 2004). This kind of integration can fully utilize the synergistic action of different enzymes and significantly promote polysaccharide (cellulose and hemicellulose) degradation, especially when considering the complexity of various closely interwoven polysaccharides present in plant cell walls (see above). A *xyl-doc* gene cluster (32-kb) encoding 14 putative cellulosome enzymes, which may be related to hemicellulose hydrolysis, was discovered on the genome of *Ruminiclostridium cellulolyticum* (Mroueh *et al.*, 2019). In some cases, the scaffolding may also comprise a unique structure that anchors the cellulosome to the cell wall extracellularly. For example, a cellulosome complex from *A. thermocellus* (*C. thermocellum*) can bind to cellulose fibrils using a cellulose-binding module of the CipA primary scaffold protein. The whole assembled architecture can further tether to the bacterial cell wall by the surface-layer homology (SLH) module of the OlpB anchoring scaffold protein (Ding *et al.*, 2020) (Figure 1.5). Thus, cellulosomes represent a unique strategy and a complex enzyme system with superior activities in degrading recalcitrant cellulosic or hemicellulosic substrates.

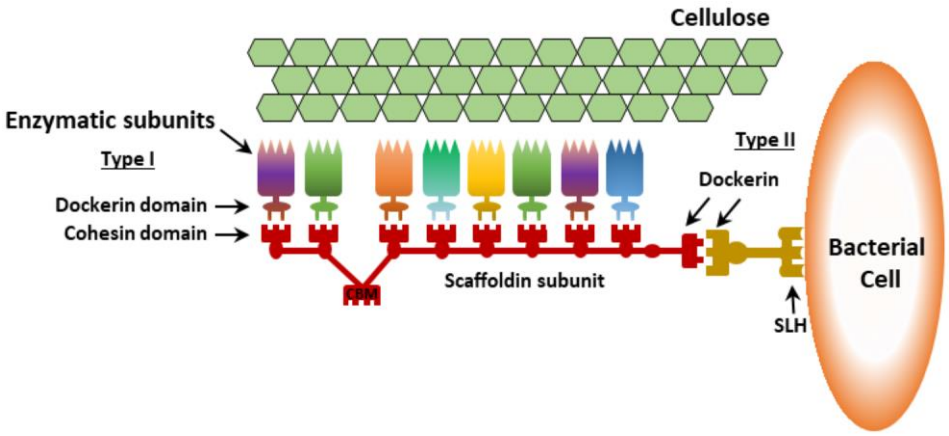


Figure 1.5 Schematic of cellulosome and its functional interaction with the plant cell wall/bacterial cell wall components. The type I dockerin domain interacts with the cohesion domain on the scaffolding protein. In contrast, type II dockerin binds the surface-layer homology (SLH) on the cell membrane, thus tethering the cellulosome complex to the surface of bacteria. The schematic was shown, based on Ding *et al*. (2020).

1.3.4 Sugar transport systems in bacteria

Due to the considerable molecular weight and the hindrance of side-chain substitution, long-chain polysaccharides have limitations in crossing the cell membrane. In comparison, the oligosaccharides can be transported intracellularly, but this process requires specific/non-specific sugar transporters. Transport systems in bacteria allow the uptake of extracellular nutrients such as various sugar oligomers and help the microbes excrete the end products of metabolism (Mitchell, 1967). The most prominent families of membrane transporters include mainly ATP-binding cassette (ABC) superfamily and major facilitator superfamily (MFS), which is also named "uniporter-symporter-antiporter family." ABC transporters are usually multi-component transporters, transporting both small molecules and macromolecules along with ATP hydrolysis (Paulsen *et al.*, 1997). By contrast, MFS transporters are composed of a single polypeptide which can only carry small solutes, such as single sugars or oligosaccharides, across cell membranes in response to chemiosmotic ion gradients (Saier, 1998). Both transporter systems connect the intracellular space with the environment and link the extracellular and intracellular enzymatic hydrolysis of polysaccharides and/or oligosaccharides.

1.4 Strategies for finding novel bacterial enzymes

1.4.1 Searching for novel enzymes by the isolation of novel bacterial strains

A traditional method to find novel bacterial enzymes with strong polysaccharide degrading abilities is to isolate bacterial strains, which can hydrolyze the target polysaccharides specifically, followed by sequencing the whole genome of the single strains. Many bacteria were found from nature or anthropogenic environments with this method, and innumerable enzymes with various activities were annotated by searching against databases of known sequences. For example, Poehlein *et al*. isolated the thermophilic bacterium strain *Thermoclostridium stercorarium* (syn. *Clostridium stercorarium*) DSM 8532, exhibiting strong polysaccharide degrading ability (Poehlein *et al*., 2013). The complete genome sequence of the bacterial strain revealed a vast number of genes encoding putative hydrolytic enzymes. Broeker *et al.* choose 50 genes encoding partially uncharacterized hemicellulolytic enzymes from this organism for further characterization. It showed various enzymatic activities and specificities, which can serve as promising candidates for industrial applications (Broeker *et al.*, 2018). Around 10^6-10^8 bacterial species are estimated to exist. Among them, more than 10^4 microbes were validly described. However, owing to limits in today's microbiology techniques, approximately 98% of bacteria on the

earth have yet to be discovered, indicating a vast wealth of astoyet unknown genes, enzymes, metabolic pathways, and chemicals. Therefore, the major challenge in this field is to make the whole functional biodiversity accessible.

1.4.2 Function-based screenings

Since most bacteria from environmental samples are uncultivable, a cultivation-independent method named metagenomics was developed a decade ago to overcome the limitations of the cultivation-based approach and unveil the microbial diversity, their metabolic potentials and ecological roles (Figure 1.6). With metagenomics, the genomic DNA of the microbial community can be extracted directly from the environment. The enzymes may be identified using the functional metagenomic library screening approach, which involves creating metagenomic libraries and then analyzing them using sequence-based methods.

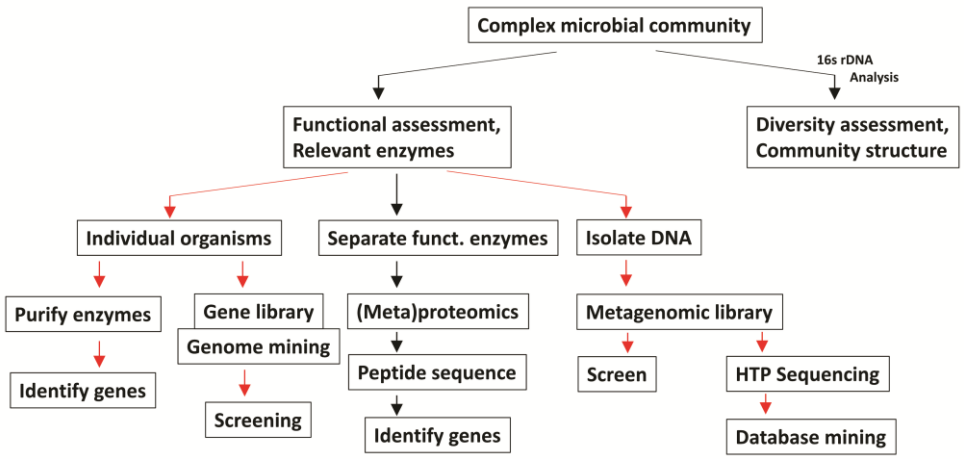


Figure 1.6 Approaches for the isolation of new genes for enzymes. The red arrow represents the methods used in my study.

Function-based metagenomic library screening is based on the successful expression of foreign genes in the host strain. The screening methods are also crucial for the proper isolation of positive clones. According to the principle of screening, the activity-based screening methods can be classified into the following three types. Phenotypic insert detection is commonly used, which is using specific chemical reagents as an indicator for searching for enzymatic activities. By contrast, the modulated detection method only allows the growth of positive library clones under selective conditions due to a separate reporter system or heterologous complementation. The third method is a substrate or product

induction strategy. Where the reporter genes can express in a host only when certain enzymes/molecules are present (Ekkers *et al.*, 2012; Simon *et al.*, 2009a).

For metagenomic library construction, the quality of nucleic acids and a proper host-vector system determines the outcome of the metagenomic library. The quality of genomic material includes the purity of the environment DNA, fragment size, concentration, and degree of bias. Humic substance contamination is a common problem with nucleic acids from soil samples (Wang *et al.*, 2012).

Based on the size of the DNA fragment, different cloning vehicle backbones can be chosen for metagenomic library construction, commonly used vectors or backbones include plasmids, fosmids and BACs, which can carry up to 15 kb, 40 kb and 200 kb DNA fragments, respectively. Each method has its pros and cons. Small-insert libraries can only carry few genes within each small insert, but is suitable for finding the enzymes with low activities due to the high copy number of the plasmids in a host cell and/or a strong promoter within the plasmid. Large-insert libraries such as fosmid or BAC libraries can carry more genes and express genes or even the whole operons mainly by native promoters from the insert. Therefore, such libraries can be beneficial for finding enzymes encoded in gene clusters or the entire metabolic pathways.

E. coli, as a model organism in microbiology, was used as a heterologous host for most function-based metagenomic studies since it is simple to handle microbiologically, grows fast, and is easily transformable with recombinant vectors (Ekkers *et al.*, 2012; Simon *et al.*, 2011; Taupp *et al.*, 2011). But as a host, *E. coli* also has limitations. Many studies showed *E. coli* is unable to functionally express many foreign genes, possibly due to the different transcription signals (Warren *et al.*, 2008), inappropriate ribosomal binding sites (Sorensen *et al.*, 2005), codon usage preference (Villegas *et al.*, 2008) and mRNA stability (Kudla *et al.*, 2009). Therefore, other alternative hosts, such as *Streptomyces* (Courtois *et al.*, 2003), *Pseudomonas* (Martinez *et al.*, 2004), or *Thermus* were also developed in recent years (Leis *et al.*, 2015; Liebl *et al.*, 2014).

1.4.3 Sequence-based analysis

Sequence-based metagenomics is based on sequencing and further analyzing the obtained DNA sequences of the environmental microbial community. Commonly used methods include PCR and hybridization techniques as well as next-generation sequencing (NGS).

With NGS, the sequence of total metagenomic DNA can be indiscriminately determined. NGS methods include all innovative technologies for high throughput sequencing and all of them include fragmentation, preparation of a DNA library and the sequential detection of incorporated nucleotides during *in vitro* synthesis (Metzker, 2010; Shokralla *et al.*, 2012). If the sequencing depth is high enough, even complete genome assembly can be achieved with low microbial community diversity. After obtaining the sequences, phylogenetic diversity of the microbial community, functional genes and metabolic pathways can be identified by using bioinformatic tools or searching against available databases (Wooley *et al.*, 2010; Simon *et al.*, 2011; Thomas *et al.*, 2012).

An alternative approach, either a metagenomic library or isolated metagenomic DNA can serve as the template for amplifying the target genes. Hybridizing DNA probes or primer sequences can be designed using conserved regions of target protein classes, followed by PCR or hybridization techniques to find the desired genes (Leis *et al.*, 2013). Other derived approaches based on PCR amplification were also applied to metagenomic studies, such as primer walking alone the insert to recovery the complete sequence of ORFs or DNA fragments. Quantitative RT-PCR can be used to estimate the specific metabolic activity by transcript analysis (Leis *et al.*, 2013).

1.5 The objectives of the dissertation

Although the degradation of pectin has been studied for many years, it still remains a challenge to understand the diversity of pectin-degrading enzymes. Some pectin structural motifs, particularly in the case of RG-II, are still resistant to enzymatic degradation. Increasing the knowledge about these enzymes is not only beneficial for industrial application, including fruit juices processing, oil extraction, and plant fibers retting, but also opens avenues for other applications related to the processing of plants, either by using bacterial enzymes or by controlling the activity of endogenous enzymes in plants.

Bacterial enzymes that degrade plant cell wall polysaccharides play an essential role in the natural deconstruction of plant-derived biomass. However, the discovery and exploitation of this biological diversity are far from exhausted. To find novel enzymes for plant biomass degradation, it is crucial to extend the portfolio of studied organisms and enzymes and further explore the unknown bacterial biodiversity and their enzymes from natural or manufactured habitats. Therefore, the goals of this project are:

1. Identification of interesting novel cellulolytic, hemicellulolytic, and pectinolytic enzymes from newly isolated polysaccharide-degrading bacteria strains. For this, two approaches were anticipated: First, with a sequence-based method, open reading frames of interesting genes can be identified after elucidating the genome sequence of the bacteria via sequence similarity to known GH enzyme families in the existing database. Second, a sequence-independent functional screening approach was envisaged, where custom-made, dyed polysaccharide substrates are used for screening of small fosmid libraries constructed from the genomic DNA of the bacteria (or from bacterial enrichments) in *Escherichia coli*.

2. Characterization of interesting, novel enzymes identified from the genomic DNA of isolated bacterial strains or the fosmid clones selected by functional screening. To this end, selected open reading frames were to be cloned and expressed as His- or Strep-tagged fusion proteins in *E. coli* for high-level expression. The recombinant, purified enzymes should then be subjected to detailed biochemical analysis.

3. A further goal was to enrich, isolate, and characterize additional new strains of polysaccharide-degrading bacteria from natural or manufactured habitats. Enrichments of bacteria can be used for the sequence-independent functional screening approach for finding enzymes as described above. On the other hand, isolates obtained in pure culture can be characterized microbiologically and genetically, including genome sequencing and description of new species or genera if meaningful and of sufficient novelty.

In conclusion, the work tackled in this study served to uncover new polysaccharide-degrading enzymes and to significantly increase our knowledge in the field of microbial polysaccharide decomposition. In addition, enzymes newly characterized during this study may be of value for biotechnological application purposes.

2 Materials and methods

2.1 Bacterial strains, plasmids, media, chemicals, materials and equipments

2.1.1 Bacterial strains and plasmids

Table 2.1 Bacterial strains used.

Bacterial strains	Genotype	Reference
E. coli BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	BioLabs, Feldkirchen, Germany
E. coli ArticExpress (DE3)	*E. coli B F⁻ompT hsdS($r_B^-m_B^-$) dcm⁺ Tet^r gal λ(DE3) endA Hte [cpn10 cpn60 Gent^r]*	Agilent Technologies, Santa Clara, USA
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*	Stratagene, La Jolla, USA
E. coli EPI300-T1R	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL (StrR) nupG trfA tonA*	Epicentre, Madison, USA
Acetivibrio mesophilus (N2K1)	*Isolated from lab scale biogas plant fed with maize silage (NCBI Reference Sequence ID: 85920)*	Rettenmaier *et al*., (2019)

Table 2.2 Plasmids/Fosmids used.

Plasmids/Fosmids	Features	Reference
pCC1FOS	8,139 bp Fosmid for the construction of a CopyControl fosmid library for the cloning of 40 kbp genomic DNA fragments, Chloramphenicol resistance as an antibiotic selectable marker, *E. coli* F factor-based partitioning and single-copy origin of replication, *ori*V high-copy origin of replication, Bacteriophage lambda cos site for lambda, packaging or lambda-terminase cleavage, Bacteriophage P1 loxP site for Cre-recombinase cleavage, Bacteriophage T7 RNA polymerase promoter flanking the cloning site.	Epicentre, Madison, USA
pCT3FK	11.6 kb large *E. coli* / *T. thermophilus* shuttle fosmid, which is based on the pCC1FOS fosmid Kanamycin-resistance (KmR) and chloramphenicol resistance (CmR), It has flanking regions for integration into *pyr*E locus in *T. thermophiles*,	Angelov *et al*., (2009)
pET24c	5,308 bp, expression vector, It lacks the ribosome binding site and ATG start codon but contains a C-terminal His·Tag® sequence, Kanamycin resistance, IPTG induced expression.	Merck, Darmstadt, Germany

Plasmids/Fosmids	Features	Reference
pMAL-c2X	In the pMAL™ protein fusion and purification system, the cloned gene is inserted into a pMAL vector downstream of the *malE* gene encoding maltose-binding protein (MBP). The outcome is the expression of an MBP fusion protein. The technique uses the robust P_{tac} promoter and translation initiation signals from MBP to express high levels of the fusion protein.	BioLabs, Feldkirchen, Germany

2.1.2 Culture media, buffers and media additives

The media used in the study were prepared according to the following instructions and sterilized at 121°C for 20 minutes in an Autoclaves VX-150 (Systec, Linden, Germany). For solid media, 1.5% (w/v) agar was added to the broth before autoclaving. Other thermolabile additives such as antibiotics, monosaccharides, salts, minerals or vitamins were first filtered sterilely (0.45 µM) and then added to the autoclaved medium, which had been cooled to around 60°C. 25 mL of the agar medium was poured into a Petri dish (Biozym Scientific, Hessisch Oldendorf, Germany) and further used after solidification in the clean bench. Determination of the pH was performed with a 902 pH/Conductivity Meter (Berrytec®, Harthausen, Germany).

Table 2.3 LB (Lysogeny broth) medium.

Component	Amount
Tryptone	10.0 g
Yeast extract	5.0 g
NaCl	5.0 g
Adjust pH to 7.0 with NaOH	
ddH_2O	1000 mL
Solid medium	
Agar	15.0 g L^{-1}

Table 2.4 SOB (Super Optimal Broth) medium.

Component	Amount
Yeast extract	5.0 g
Tryptone	20.0 g
NaCl	10 mM
KCl	2.5 mM
ddH_2O	1000 mL
Adjust pH to 7.0 before autoclaving	
$MgCl_2$ (sterilized by filtration)	10 mM
$MgSO_4$ (sterilized by filtration)	10 mM

Table 2.5 GS2 medium.

Basic medium		**100x GS2 salt solution**	
K_2HPO_4	2.9 g	$MgCl_2 \cdot 6H_2O$	10.0 g
KH_2PO_4	1.5 g	$CaCl_2 \cdot 2H_2O$	1.5 g
MOPS	10.0 g	$FeSO_4$ (1.25%)	1 mL
Yeast extract	6.0 g	ddH_2O	100 mL
Urea	2.1 g		
Na-Citrat $\cdot 2H_2O$	2.9 g		
Cystein	1.0 g		
Resazurin	2.0 mg		
Adjust pH to 7.2 with NaOH			
ddH_2O	990 mL		

The buffers used in the study were prepared according to the following instructions. Buffer filtration was performed with a vacuum filtration unit (Sartorius Stedim, Sartorius Stedim, Goettingen) and a membrane vacuum pump (Vacuubrand Gmbh + Co Kg, Wertheim, Germany). The membrane filter insert was a 0.2 µm with a 47 mm diameter cellulose nitrate paper (Whatman®, Merck, Darmstadt, Germany). For fast protein liquid chromatography (FPLC), buffer B was prepared by adding 1 M NaCl in buffer A. Both buffer A and buffer B were degassed by using a vacuum chamber (max. 2 bar) from GLASWERK WERTHEIM (Mc Labor, Ablstadt, Germany) with a membrane vacuum pump.

Table 2.6 TB solution.

Component	**Amount**
PIPES	10 mM
$CaCl_2$	15 mM
KCl	250 mM
$MnCl_2$	55 mM
Adjust pH to 6.7 with KOH or HCl, sterilized by filtration and store at 4°C.	

Table 2.7 Citrate-phosphate buffer.

pH	0.2 M Na_2HPO_4 (mL)	Concentration [mM]	0.1 M citric acid (mL)	Concentration [mM]
4.0	7.71	77.1	12.29	61.45
4.6	9.35	93.5	10.65	53.25
5.0	10.30	103.0	9.70	48.50
5.6	11.60	116.0	8.40	42.00
6.0	12.63	126.3	7.37	36.85
6.6	14.55	145.5	5.45	27.25
7.0	16. 47	164.7	3.53	17.65
7.6	18.73	187.3	1.27	6.35
8.0	19.45	194.5	0.55	2.75

[Citric acid 1-hydrate (Applichem); $Na_2HPO_4 \cdot 2H_2O$ (Carl Roth)]. The pH was adjusted by mixing both buffers in a different ratio.

pH 2.0 - 3.5: 0.1 M citric acid only, pH adjusted with HCl or NaOH.

pH 4.0 - 8.0: preparation for 20 mL of buffer with specific pH as described by McIlvaine *et al*. (1921).

pH 8.5 - 9.0: 0.2 M $Na_2HPO_4 \cdot 2H_2O$ only, pH adjusted with HCl or NaOH.

All pH values were adjusted under the specific optimal temperature of the enzyme. The buffer was heated in an enclosed water bath (Memmert, Schwabach, Germany) followed by titration of HCl or NaOH if necessary.

Table 2.8 50 mM phosphate buffer, pH 7.0.

Component	Amount	Concentration
$Na_2HPO4 \cdot 7H_2O$ (mw: 268.07 g mol^{-1})	7.75 g	0.0289 M
$NaH_2PO4 \cdot H_2O$ (mw: 137.99 g mol^{-1})	2.91 g	0.0211 M
ddH_2O	1000 mL	

Table 2.9 500 mM Tris-HCl buffer, pH 6.8.

Component	Amount
Tris base	60.6 g
Adjust to pH 6.8 with HCl	
ddH_2O	1000 mL

Table 2.10 1× TE Buffer.

Component	Concentration
Tris-HCl	10 mM
EDTA	1 mM
Adjust pH to 7.5, sterilized by filtration	

Table 2.11 CTAB containing DNA extraction buffer.

Component	Amount/Concentration
Tris-HCl [pH=8.0]	100 mM
Sodium EDTA	100 mM
Sodium Phosphate	100 mM
NaCl	1.5 M
Cetyltrimethylammonium bromide	1% (wt/vol)
Proteinase K (10 mg mL^{-1})	1 mL
RNase	4 mg

Table 2.12 50× TAE buffer.

Component	Concentration
Tris	2 M
EDTA	0.05 M
Acetic acid	1 M

Table 2.13 5× TBE buffer.

Component	Amount/Concentration
Tris	54.0 g
Boric acid	27.5 g
H_2O	900 mL
0.5 M EDTA pH=8.0	20 mL
Adjust total volume to 1000 mL	

Table 2.14 Phage dilution buffer.

Component	Concentration
Tris	10 mM
NaCl	100 mM
$MgCl_2$	10 mM
Adjust pH to 7.5, sterilized by filtration	

Table 2.15 Collection gel buffer for SDS-PAGE.

Component	Concentration
Tris	61 g L^{-1}
SDS	4 g L^{-1}
Adjust pH to 6.6	

Table 2.16 Separation gel buffer for SDS-PAGE.

Component	Concentration
Tris	182.0 g L^{-1}
SDS	4.0 g L^{-1}
Adjust pH to 8.8	

Table 2.17 SDS-PAGE gel preparation (for two gels).

Component	Separation gel	Collection gel
Separation gel buffer	2.5 mL	-
Collection gel buffer	-	1.2 mL
Acrylamide/Bis Solution, 29:1	2.5 mL	0.625 mL
ddH_2O	5.0 mL	3.175 mL
APS	60 µL	40 µL
TEMED	30 µL	20 µL

Table 2.18 4× SDS loading dye.

Component	Amount/Concentration
Tris-HCl	20 mM
4% β-Mercaptoethanol	4.0 g
8% SDS	8.0 g
0.4% Bromophenol blue	0.4 g
40% Glycine	40.0 g
ddH_2O	100 mL

Table 2.19 Tris-Glycine SDS-running buffer.

Component	Amount/Concentration
Tris	5 mM
Glycine	192 mM
0.1% SDS	1.0 g
ddH_2O	1000 mL

Table 2.20 SDS-PAGE staining solution.

Component	Amount/Concentration
Coomassie	0.5 g
Acetic acid	100 mL
Isopropanol	250 mL
ddH_2O	650 mL

Table 2.21 SDS-PAGE de-staining solution.

Component	Amount/Concentration
Acetic acid	100 mL
ddH_2O	900 mL

Table 2.22 DNSA solution.

Component	Amount
DNSA	10.0 g
K-Na-Tartrate	200.0 g
NaOH	10.0 g
Na-Sulfate	0.5 g
Phenol	2.0 g
H_2O	1000 mL

Table 2.23 Concentrations of antibiotics and media additives.

Antibiotic	Initial conc. (mg mL^{-1})	Working conc. (µg mL^{-1})	Company	ID
Kanamycin	50.0	50.0	Carl Roth GmbH&Co.KG, Karlsruhe, Deutschland	T832.2
Ampicillin	100.0	100.0	AppliChem GmbH, Darmstadt, Deutschland	A6352.0025
Gentamycin	20.0	20.0	Sigma-Aldrich, Steinheim, Deutschland	G126
Chloramphenicol	50.0	12.5	Sigma-Aldrich, Steinheim, Deutschland	C0378
Isopropyl-β-D-1-thiogalactopyranoside (IPTG)	500 mM	0.1-1 mM	Glentham LIFE SCIENCES, Corsham SN13 9SW, United Kingdom	CAS 367-93-1

2.1.3 Substrates and enzymes

Commercial substrates

Table 2.24 Used para-nitrophenol glycosides.

Product code	*p*NP-Glycoside	Company	ID
O-PNPAF	4-Nitrophenyl-α-L -arabinofuranoside	Megazyme, Bray, Ireland	6892-58-6
O-PNPX	4-Nitrophenyl-β-D-xylopyranoside	Megazyme, Bray, Ireland	2001-96-9
O-PNPX2	4-Nitrophenyl-β-xylobioside	Megazyme, Bray, Ireland	6819-07-04

Table 2.25 Used linear oligosaccharides.

Product code	Oligosaccharides	Company	ID
O-XBI	Xylobiose	Megazyme, Bray, Ireland	6860-47-5
O-XTR	Xylotriose	Megazyme, Bray, Ireland	47592-59-6
O-XTE	Xylotetraose	Megazyme, Bray, Ireland	22416-58-6
O-XPE	Xylopentaose	Megazyme, Bray, Ireland	49694-20-4
O-XHE	Xylohexaose	Megazyme, Bray, Ireland	49694-21-5
O-ATR	Arabinotriose	Megazyme, Bray, Ireland	89315-59-3

Table 2.26 Used branched oligosaccharides.

Product code	Oligosaccharides	Company	ID
O-A3X	3^3-α-L-Arabinofuranosyl-xylobiose (A^3X)	Megazyme, Bray, Ireland	144938-89-6
O-A2XX	2^3-α-L-Arabinofuranosyl-xylotriose (A^2XX)	Megazyme, Bray, Ireland	152842-73-4
O-XA3XX	3^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX)	Megazyme, Bray, Ireland	84666-93-3
O-A2X3	2^2,3^2-di-α-L-Arabinofuranosyl-xylotriose ($A^{2,3}XX$)	Megazyme, Bray, Ireland	155696-89-2
O-XA23XX	2^2,3^2-di-α-L-Arabinofuranosyl-xylotretraose ($XA^{2,3}XX$)	Megazyme, Bray, Ireland	
O-XAXX MIX	3^3-α-L- plus 2^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX/XA^2XX) mixture	Megazyme, Bray, Ireland	84666-93-3
O-A4B	3^2-α-L-Arabinofuranosyl-(1,5)-α-L-arabinotriose (AA^3A)	Megazyme, Bray, Ireland	1237488-35-5
O-A5BMIX	2^2,3^2-di-α-L-Arabinofuranosyl-(1,5)-α-L-arabinotriose plus 3^2-α-L-arabinofuranosyl-(1,5)-α-L-arabinotetraose ($AA^{2,3}A$+AAA^3A)	Megazyme, Bray, Ireland	1237488-32-2
O-UXX	2^3-(4-O-Methyl-α-D-Glucuronyl)-xylotriose	Megazyme, Bray, Ireland	55196-23-1
O-XUX	2^2-(4-O-Methyl-α-D-Glucuronyl)-xylotriose	Megazyme, Bray, Ireland	103784-24-3

AXOS was named according to the nomenclature of Fauré *et al.* (2009).

Table 2.27 Used polysaccharides.

Product code	Polysaccharides	Company	ID
P-DBAR	Debranched arabinan (Sugar Beet, DA)	Megazyme, Bray, Ireland	9060-75-7
P-ARAB	Arabinan (Sugar Beet, SBA)	Megazyme, Bray, Ireland	11078-27-6
P-LARB	Linear 1,5-α-L-Arabinan (Sugar Beet, LA)	Megazyme, Bray, Ireland	11078-27-6
P-XYLNBE	Xylan (Beechwood)	Megazyme, Bray, Ireland	9014-63-5
P-WAXYI	Arabinoxylan (Wheat flour; Insoluble)	Megazyme, Bray, Ireland	9040-27-1
P-WAXYRS	Arabinoxylan (Wheat flour; for reducing sugar assays) Oat spelt xylan	Megazyme, Bray, Ireland	9040-27-1
P-PGAPT	Pectic galactan (potato)	Megazyme, Bray, Ireland	Not Applicable
P-PGALU	Pectic galactan (lupin)	Megazyme, Bray, Ireland	Not Applicable
P-GALLU	Galactan (ex. lupin)	Megazyme, Bray, Ireland	9037-55-2
P-GALPOT	Galactan (ex. potato)	Megazyme, Bray, Ireland	9037-55-2
LM pectin	Low methoxy pectin	Carbosynth, St. Gallen, Switzerland	9000-69-5
HM pectin	High methoxy pectin	Carbosynth, St. Gallen, Switzerland	9000-69-5
P-ARGAL	Larch wood arabinogalactan	Megazyme, Bray, Ireland	9036-66-2
PGA	Polygalacturonic acid	Carbosynth, St. Gallen, Switzerland	25990-10-7
P-BGBL	β-glucan (Barley)	Megazyme, Bray, Ireland	9041-22-9
P-RHAGN	Rhamnogalacturonan	Megazyme, Bray, Ireland	39280-21-2
CMC	Carboxymethyl cellulose	Sigma-Aldrich,Darmstadt, Germany	9004-32-4
I-AZCEL	AZCL-cellulose	Megazyme, Bray, Ireland	-
I-AZBGL	AZCL-β-glucan	Megazyme, Bray, Ireland	-
-	SBP	Tiense Suikerraffinaderij n.V., Tienen, Belgium	

CPH/ICB substrates and enzymes

The chromogenic polysaccharide hydrogel (CPH) and insoluble chromogenic biomass (ICB) substrates were synthesized according to a protocol described by Kračun *et al*. (2015). The substrate polymers were mixed with 0.5 M NaOH in a concentration of 10% (w/v) by stirring (110 rpm) at 50°C for 30 min. Throughly mixed substrate solutions were then dyed by adding a chlorotriazine dye in a ratio of 4:1 (w/w, reactive blue 4 for xylan, arabinoxylan, arabinan and SBP, reactive green 19 for pectin and arabinan or reactive red 120 for CMC) followed by incubating at 50°C for 1-2 h with stirring. After being cooled down to room temperature, cross-linking of the dyed polysaccharides were obtained by adding 1,4-butanediol diglycidyl ether as cross-linker (1.2% for xylan and SBP, 1% for arabinoxylan, 10% for CMC and arabinan). The reaction mixtures were mixed thoroughly for 1 min and then left to stand for 48 h at room temperature without agitation for hydrogel formation. The resulting hydrogels were homogenized to a paste-like material by using a laboratory mortar. To purify the substrates, the ground

hydrogels were transferred onto a Whatmann® filter paper placed over a Büchner funnel. Then they were washed with boiling Milli-Q-water until the unbound dye was removed and filtrates become colorless. After that, the samples were collected and stored at 4°C for enzyme assays. Before application, the chromogenic substrates were tested by using commercial enzymes as listed in the table below:

Table 2.28 Enzymes used in the study.

Name of the substrates	Positive Control			
	Color	**Name**	**Source**	**Amount of enzyme**
CPH-arabinoxylan	Blue	Viscozyme®L	Megazyme	10 µL (1:5dilluted)
CPH-arabinan	Blue, Green	Arabinanase	*Pp*Abn43	1 U mL^{-1}
AZCL-cellulose	Blue	Endo-cellulase	Megazyme	1 U mL^{-1}
AZCL-β-glucan	Blue	Viscozyme®L	Sigma-Aldrich	10 µL (1:5 dilluted)
AZCL-avicel	Blue	Viscozyme®L	Sigma-Aldrich	10 µL (1:5 dilluted)
CPH-xylan	Blue	Endo-xylanase	Megazyme	1 U ml^{-1}
ICB-sugar beet pulp	Blue	Viscozyme®L	Sigma-Aldrich	10 µL (1:5 dilluted)
ICB-HM Pectin	Green	Viscozyme®L	Sigma-Aldrich	10 µL (1:5 dilluted)
CPH-CMC	Red	Endo-cellulase	Megazyme	1 U mL^{-1}
CPH-Rhamnogalacturonan	Red	Viscozyme®L	Sigma-Aldrich	10 µL (1:5 dilluted)

The enzymes were incubated in 200 µL reactions with 50 µL CPH or ICB substrate at pH 7.0, 37°C, overnight under 600 rpm shaking.

Table 2.29 Other enzymes used in the study.

Product code	Enzymes	Company	ID
V2010-50ML	Viscozyme®L	Sigma-Aldrich, Darmstadt, Germany	-
E-XYAN4	Eodo-1,4-β-Xylanase M4 (*Aspergillus niger*)	Megazyme, Bray, Ireland	9025-57-4
E-CELAN	Cellulase (Endo-1,4- β-D-glucanase) (*Aspergillus niger*)	Megazyme, Bray, Ireland	9012-54-8
M0491	PCR Using Q5® High-Fidelity DNA Polymerase	BioLabs, Feldkirchen, Germany	-
R6000	RNase	Merck, Darmstadt, Germany	9001-99-4
P2308	Proteinase K	Merck, Darmstadt, Germany	39450-01-6
ER0581	NdeI	BioLabs, Feldkirchen, Germany	-
ER0692	XhoI	Thermo Scientific, Waltham, MA, USA	-
ER0051	BamHI	Thermo Scientific, Waltham, MA, USA	-
ER0662	SmaI	Thermo Scientific, Waltham, MA, USA	-
ER0271	EcoRI	Thermo Scientific, Waltham, MA, USA	-

2.1.4 Chemicals, materials and kits

The chemicals, materials and kits used in my study were described in the table below.

Table 2.30 Chemicals used.

Substance	Manufacture
Acetone (for UV, IR, HPLC, GPC, ACS) (99.9%)	AppliChem,Darmstadt, Germany
Agar	Merck, Darmstadt, Germany
Agarose	Merck, Darmstadt, Germany
Acetic acid	Applichem, Darmstadt, Germany
Acetone	Merck, Darmstadt, Germany
Aniline	Merck, Darmstadt, Germany
Bradford reagent	SERVA, Heidelberg, Germany
Chloroform	Carl Roth, Karlsruhe, Germany
Citric acid 1-hydrate	Applichem, Darmstadt, Germany
Diphenylamine	Merck, Darmstadt, Germany
EDTA	Merck, Darmstadt, Germany
Gold agarose	Biozym Scientific, Hessisch Oldendorf, Germany
Glycerol	Merck, Darmstadt, Germany
Na_2HPO_4	Roth, Karlsruhe, Germany
NaH_2PO_4	Roth, Karlsruhe, Germany
NaCl	Applichem, Darmstadt, Germany
NaOH (50% solution)	VWR International, Radnor, USA
Q5 enhancer	BioLabs, Feldkirchen, Germany
Reactive blue 4, sc-215794	Santa Cruz Biotechnology, Heidelberg, Germany
Reactive green 19, sc-215796	ChemCruz, Texax, USA
Reactive red 120, sc-215799	ChemCruz, Texax, USA
SDS	Merck, Darmstadt, Germany
Sodium acetate (≥99%)	Sigma-Aldrich, St. Louis, USA
1,4-Butanediol diglycidyl ether, sc-488306	ChemCruz, Texax, USA

Table 2.31 Material used.

Substance	Manufacture
Aluminum crimp	Sigma-Aldrich, St. Louis, USA
Aluminum crimp cap	Diagonal, Münster, Deutschland
Butyl rubber stopper	Glasgerätebau Ochs, Bovenden, Deutschland
Fisherbrand™ Electroporation Cuvettes Plus™ 2mm gap	Fisher Scientific GmbH, Schwerte, Germany
Filtropur S 0.45 filter	SARSTEDT, Nümbrecht, Germany
GLASWERK WERTHEIM	Mc Labor, Ablstadt, Germany
HiTrap™ 1 ml QFF column	Cytiva, Marlborough, USA
LIGHTrun NXP – lab record barcodes	Eurofins Genomics Germany GmbH, Ebersberg, Germany
Petri dish	Greiner Bio-One International AG, Kremsmünster, Austria
PCR adhesive foil	Biozym Scientific, Hessisch Oldendorf, Germany
TLC Silica gel 60 20 cm x 20 cm	Merck, Darmstadt, Germany
VIVASPIN TURBO 4 (5,000 MWCO membrane)	Sartorius Stedim, Goettingen, Germany
VIVACELL 250	Sartorius Stedim, Goettingen, Germany

Substance	Manufacture
Whatman® cellulose nitrate paper (0.2 µM, 47 mm)	Merck, Darmstadt, Germany
Whatman™ filter paper grade 1	GE Healthcare, Munich, Germany
96-well plate	Eppendorf, Hamburg, Germany
96-well PCR plate	Biozym Scientific, Hessisch Oldendorf, Germany

Table 2.32 Kits and standards used.

Chemical	Manufacture
Bradford reagent	SERVA, Heidelberg, Germany
Bovine serum albumin (BSA)	Thermo Scientific, Waltham, MA, USA
DNA Clean&Concentrator™-5	ZYMO RESEARCH, Freiburg, Germany
dNTPs	Thermo Scientific, Waltham, MA, USA
GeneRuler® 1kB DNA Ladder, 11823963	Thermo Scientific, Waltham, MA, USA
Monarch® Miniprep Kit	BioLabs, Feldkirchen, Germany
NEBuilder HifiDNA Assembly Master Mix	BioLabs, Feldkirchen, Germany
Protino® Ni-TED 2000 Packed Columns Kit	MARCHERY-NAGEL GmbH & Co.KG, Düren, Germany
PD-10 Desalting Columns contain Sephadex™ G-25 Medium	GE Healthcare, Munich, Germany
PageRuler™ Prestained Protein Ladder, 26616	Thermo Scientific, Waltham, MA, USA
Zymoclean Gel DNA Recovery Kit	ZYMO RESEARCH, Freiburg, Germany
6X TriTrack DNA Loading Dye, R1161	Thermo Scientific, Waltham, MA, USA

2.1.5 Equipments

Table 2.33 Equipment used in this study.

Equipment	Manufacture
A27-8 × 50 rotor	Thermo scientific, Waltham, MA, USA
Absorbance microplate reader Sunrise™	Tecan GmbH, Crailsheim, Germany
Alphalmager MINI	Biozym Scientific, Hessisch Oldendorf, Germany
Anaerobierzelt Type A Anaerobic Chamber	COY Laboratory Product, Grass Lake, MI, USA
Autoclaves VX-150	Systec, Linden, Germany
ÄKTApure 25L 1 FPLC system	Cytiva, Marlborough, USA
Mili-Q® Direct Water Purification System (0.055 µS)	MilliporeSigma Life Science Center, Burlington, Massachusetts
Butyl rubber stopper	Glasgerätebau Ochs, Bovenden, Germany
Bio-Rad PCR Thermo-Cycler	Bio-Rad Laboratories GmbH, Feldkirchen, Germany
Chromeleon 7.2 software	Thermo Fischer Scientific, Waltham, MA, USA
ChromaJet DS 20 reagent sprayer	Desaga Sarstedt, Nümbrecht, Germany
Dionex™ ICS-3000 SP system	Thermo Fischer Scientific, Waltham, MA, USA
Electrophoresis apparatus EPS 600	Pharmacia Biotech, Uppsala, Schweden
GLASWERK WERTHEIM: Glass desiccator	Mc Labor, Ablstadt, Germany
MIKRO 22R	Hettich, Tuttlingen, Germany
Membrane vacuum pump	Vacuubrand Gmbh + Co Kg, Wertheim, Germany
Magnetic stirrer Stuart® US152	Cole-Parmer, Wertheim, Germany
Microwave privileg 8018E	Bauknecht Hausgeräte, Stuttgart, Germany
NanoDrop	Thermo Fischer Scientific, Waltham, MA, USA
PCR Thermo-Cycler Biometra TADVANCED	Analytik Jena, Göttingen, Germany
Photometer SmartSpec™ 3000	Bio-Rad Laboratories, Feldkirchen, Germany
Pulse Controller Plus	Biorad, Santa Clara, United Kingdom
PowerPRO 300 Power Supply	Cleaver Scientific Ltd, Warwickshire, United Kingdom
ScanMarker 1000XL	MICROTEK, Udyog Nagar, India

Equipment	Manufacture
Sorvall LYNX 6000 Centrifuge	Thermo scientific, Waltham, MA, USA
Thermomixer	Eppendorf, Hamburg, Germany
TLC development chamber	MACHEREY-NAGEL GmbH & Co.KG, Düren, Germany
UNIVERAL 320 R centrifuge	Hettich, Tuttlingen, Germany
UP200S	Hielscher Ultrasound-Technology, Teltow, Germany
Ultrospec 2100 pro	Amersham Biosciences, Buckinghamshire, United Kingdom
Vacuum Filtration Unit (Product Nr. 61391)	Sartorius Stedim, Goettingen
Water bath	Memmert, Schwabach, Germany
902 pH/Conductivity Meter	Berrytec®, Harthausen, Germany

2.2 Microbiological methods

2.2.1 Cultivation of *E. coli*

E. coli was cultured aerobically at 37°C in liquid LB or SOB medium (for competent cell preparation) overnight with 180 rpm shaking. For a culture volume of 30 mL or more, Erlenmeyer flasks with baffles were used, and test tubes were used for smaller culture volumes (usually 5 mL). For solid medium, *E. coli* glycerol stocks and liquid cultures were streaked onto an LB agar plate, incubated overnight at 37°C and stored at 4°C. Media for culturing *E. coli* with the exogenetic vector were added antibiotic at a certain concentration for selection. Specific *E. coli* strains such as ArcticExpress, the media contain gentamycin at 20 µg mL^{-1} to select the expression plasmid. For EPI300, the media were added 12.5 µg mL^{-1} chloramphenicol to select the CopyControl Fosmid clones and 0.1 % arabinose for induction.

2.2.2 Cultivation of N2K1 strain (*Acetivibrio mesophilus*)

Strain N2K1 was grown in oxygen-free GS2-based media with additional carbon source in an N_2/H_2 atmosphere (98: 2%, v/v) in crimp neck bottles made of glass closed with a butyl rubber stopper (Glasgerätebau Ochs, Bovenden, Deutschland) and an aluminum crimp (Sigma-Aldrich, St. Louis, USA). GS2 medium, according to Johnson *et al.* (1981) and modified by Koeck *et al.* (2015), was used in our study (Table 2.5). L-cysteine and resazurin served as an oxygen-reducing agent and a color indicator in the medium, respectively, to ensure the anaerobic conditions. 0.5% polysaccharides such as filter paper, arabinan, xylan or arabinoxylan substrate were added as carbon sources. The sterile saline solution (Filter sterilized, Table. 2.5) was added after autoclaving. Other soluble, thermolabile additives, such as mono- (xylose, arabinose) or disaccharides (cellobiose), were sterile-filtered separately and added after autoclaving if required. To cultivate N2K1, 100 µL N2K1 glycerol stock was inoculated into a 10 mL GS2-based medium with specific substrates, followed by stationary incubation at 45°C. Degradation of the

filter paper and cellobiose served as a positive control.

2.2.3 Cultivation of bacterial consortia

To select a bacterial community with strong cellulolytic, hemicellulolytic or SBP-digesting ability, the fresh fermenter content, called fermentation residue, was collected from a biogas digester at the Bayerische Landesanstalt für Landwirtschaft (LfL) that was fed with sugar beet silage and used for enrichment. The fermentation residue was diluted 1: 50 in a GS2-based medium anaerobically and used as the seed sample. In order to minimize the interference of the nutrients in the yeast component on the selection, 100 µL samples were initially inoculated (biological triplicates) into 10 mL of GS2 medium with four different concentrations of yeast extract (0.6 %, 0.3 %, 0.1 %, and 0.05 %), using three different substrates (SBP, Oat spelt xylan, filter paper). The incubation took place at 37°C, in agreement with the operating temperature of the fermenter. Once the systems had reached a high bacterial cell density or the substrate had been visually degraded, aliquots (100 µL) of the enrichment culture were continuously transferred into fresh GS2 medium with the lowest yeast concentration microbial consortia can grow. The same procedure was repeated nine times. Along with the whole enrichment process, a negative control without a carbon source was also included in the study.

Other eleven different fermentation residues, which were kindly provided by Dr. Zverlov *et al.* (Table 2.34), were tested for the degradation ability of SBP and further chosen for metagenomic library construction. For this purpose, biological triplicates of each sample were inoculated anaerobically with a spatula tip or 1 mL syringe in 10 mL of medium. Serial 1:10 dilutions were performed in the GS2-based medium until the dilution level 10^{-6} was reached. The incubation took place either mesophilically (37°C) or thermophilically (50-55°C), depending on the operating temperature of the source biogas fermenter. The bacterial growth was determined by following the turbidity of the media and the formation of gas. In addition, the degradation of the filter paper indicated the activity of cellulolytic organisms, which served as a positive control.

Table 2.34 Information about fermenter samples for metagenomic library construction.

Samples	Sample descriptions
T1T2	Thermophilic digestate, 55°C, the sample was collected on 23.03.2018
MOD17	Mesophilic digestate from M1M2 fermenter fed with maize silage at the LfL, 38°C, sample from 29.08.17
S4MN4	Mixed isolates, 4 dominant organisms from a mesophilic colony, 38°C. The optimal temperature was at 45°C, pasteurized
MOD18	Thermophilic digestate, 50°C, the sample was collected on 18.07.2018

Samples	Sample descriptions
MBF-VZ-239	*Ruminococcus* sp. 55°C
Pas-Past3	*Ruminococcus* sp. 55°C
MA18P.1	Pure isolate from mesophilic fermenter, microaerophilic, 37°C, 98% similar to *Clostridium sufflavum*
N2K1	Pure isolate from S4MN4, similar to *Clostridium straminisolvens*
045	Mixed isolates, 55°C
MN3	Mixed isolates, 55°C
AN6	*Paenibacillus fonticola*, 30°C

2.2.4 Optical density determination of bacterial cultures

The optical density of bacterial cultures was measured with an Ultrospec 2100 pro (Amersham Biosciences, Buckinghamshire, United Kingdom) at a wavelength of 600 nm (OD_{600}). Bacterial cultures with an OD_{600} above 0.5 were first diluted in the corresponding medium in a ratio of 1:2 and 1:10 (1:10 usually used for overnight culture). For each measurement, the OD_{600} of the corresponding medium was read as a reference.

2.2.5 Bacterial glycerol stock

2.2.5.1 Glycerol stock of *E. coli* cultures

Glycerol stocks of the corresponding *E. coli* strains were prepared to preserve expression strains. 500 µL of an overnight *E. coli* culture was added to 500 µL of 50% sterilized glycerol, mixed well, and then quick-freezing using liquid nitrogen. The glycerol stocks were stored at -80°C.

2.2.5.2 Glycerol stock of anaerobic bacterial (consortia) enrichment

Glycerol bottles were prepared by adding 1 mL of pure glycerol in a 5 mL crimp neck bottle and degassing in an anaerobic tent (COY Laboratory Product, Grass Lake, MI, USA) for 30 min. Then the bottle was closed with a butyl rubber stopper and sealed with aluminum crimp anaerobically before autoclaving. After the glycerol bottles were autoclaved and cool down to room temperature, the 1 mL of anaerobic bacterial (consortia) enrichment were transferred into the bottles and frozen at -80°C.

2.2.6 Chemically competent cell preparation

In order to prepare competent *E. coli* cells, 250 mL of SOB medium in a 1 L flask was inoculated with 10-12 large *E. coli* colonies, which grow on LB agar plates. The inoculum was set up at 19°C with 150 rpm shaking until OD_{600} = 0.5 (24-36 h). After that, the culture was put on ice for 10 min before being centrifuged for 10 min at 4,000 rpm at 4°C. The supernatant was then removed from the EP tube, while

the resulting cell pellets were resuspended in 80 mL of ice-cold TB and store on ice for 10 min. Afterward, the cell suspension was centrifuged at 4,000 rpm at 4°C for 10 min, and the cell pellets were then gently resuspended in the 20 mL ice-cold TB with 1.4 mL DMSO, finally aliquot the cell to 50 µL for transformation or store at -80°C.

2.3 Genetic methods

2.3.1 Primers

Oligonucleotide primers were synthesized by Eurofins Genomics (Ebersberg, Germany), which were diluted in sterile ddH_2O to a concentration of 10 pmol μL^{-1}, and stored at -20°C. All primers were designed so that their melting temperature was approximately 60°C.

Table 2.35 Cloning primers for Gibson assembly.

Protein ID	Protein	Forward Primer
RXE59351.1	*Am*Araf43-1	F-CTTTAAGAAGGAGATATACATATGATAAAGAGGGCTCTAAA
(Without dockerin)	(01572)	R-TCAGTGGTGGTGGTGGTGGTGCGCCGGTGTACCGTCAGCATTCC
RXE58498.1	*Am*Araf51	F1-ATGCATCATCACCATCACCATAAAAAAGCCAGAATGACC
	(02400)	F2-ACTTTAAGAAGGAGATATACAATGCATCATCACCATCACCAT
		R-CAGTGGTGGTGGTGGTGGTGCGTTATTTCCCCAGTCGAATTAC
RXE58509.1	*Am*Araf43-2	F-CAACCTCGGGATCGAGGGAAGGATGAAAAAAATTTTTTTAATGCTG
	02412	R-GTAAAACGACGGCCAGTGCCATTAACCTTCAGGAAAAACAGTT
RXE58512.1	*Am*Araf43	F-CTTTAAGAAGGAGATATACATATGTTATTTACAAAAAAAGCC
(Without dockerin)	(02415)	R-TCAGTGGTGGTGGTGGTGGTGCGCTTCAATAAAAGTAAACCAATTC
RXE58512.1	*Am*Araf43$_{truncated}$	F-CTTTAAGAAGGAGATATACATATGTTATTTACAAAAAAAGCC
(Only GH43 domain)		R-TCAGTGGTGGTGGTGGTGGTGCGCAGGCCAGCCATCCTGCCATG
PROKKA_02440	*Am*Pgl28	F-CTTTAAGAAGGAGATATACATATGTCTACAATATTTGATATTAC
		R-CAGTGGTGGTGGTGGTGGTGCGCCTCACTAATGTCAAAATTGC
05004(*Paenibacillus*	*Pp*Abn43	F-CTTTAAGAAGGAGATATACATATGAAGCAATCATTTGTTCG
polymyxa)		R-GTGGTGGTGGTGGTGGTGCGCCAATTCTGCCAGTGCTTTAATTTC
MC57GH51	MC57GH51	F-CTTTAAGAAGGAGATATACAATGATAAATGTAACTATCAATGCAG
		R-TCAGTGGTGGTGGTGGTGGTGCGCGCAAACCTCTAAAACAAC
MC60GH51	MC60GH51	F-CTTTAAGAAGGAGATATACAATGTCAAATACTCAAAAAGCAATGATG
		R-TCAGTGGTGGTGGTGGTGGTGCGCCTTCGCTAGCCTTATAACATTCC
MC68GH43-1	MC68GH43-1	F-CTTTAAGAAGGAGATATACAATGAGGAAAAAAGTCATTTGGCTGG
		R-TCAGTGGTGGTGGTGGTGGTGCGCTTCAATAGCCCAAATACCGC
MC60GH43	MC60GH43	F-CTTTAAGAAGGAGATATACAATGAGCAAATCAGTAAATTTTG
		R-TCAGTGGTGGTGGTGGTGGTGCGCATATGTAAATACAGGGTTACCAG
MC68GH43-2	MC68GH43-2	F-CTTTAAGAAGGAGATATACAATGATGGAACGAAACTTGGTAAAAAC
		R-TCAGTGGTGGTGGTGGTGGTGCGCACCATATACCTGTTTCCGG
MC72GH43-1	MC72GH43-1	F-CTTTAAGAAGGAGATATACAATGAATTGGGAGTTAATAGCGTCAG
		R-TCAGTGGTGGTGGTGGTGGTGCGCTTCTATAGGTAGGCCTAAAG
MC72GH43-2	MC72GH43-2	F-CTTTAAGAAGGAGATATACAATGTTGACAATTACATTTGTATC
		R-TCAGTGGTGGTGGTGGTGGTGCGCCACTTCAGGAATTTCACCTG
MC40GH10	M_Xyn10	F-AACTTTAAGAAGGAGATATACAATGCTGAAAGATATTTACGCCGG
		R-TCAGTGGTGGTGGTGGTGGTGCGCGCGATCCTGTGCAGTGCTG
Primers for pET24c	pETUPstream	F-ATGCGTCCGGCGTAGA
	T7 Terminator	R-GCTAGTTATTGCTCAGCGG

Protein ID	Protein	Forward Primer
pCC1Fosmid	-	F-GGATGTGCTGCAAGGCGATTAAGTTGG R-CTCGTATGTTGTGTGGAATTGTGAGC
pCT3FK	-	F-GTAATACGACTCACTATAGGGCG R-GTAAAACGACGGCCAGTGCCAAGC
N2K1-specific primers	Prokaryotic 16S rRNA/16S N2K1^{T}	F- AGAGTTTGATYMTGGCTC R-CAKAAAGGAGGTGATCC

2.3.2 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was used for the in vitro amplification of genes. The reaction included denaturation, annealing and elongation. These three phases were repeated in cycles, which amplifies the target DNA fragment specifically. PCR reactions were conducted in Bio-Rad's thermal cycler (Bio-Rad Laboratories GmbH, Feldkirchen, Germany). Q5 High-Fidelity DNA Polymerase (from New England Biolab) and the manufacturer's corresponding protocol were used for this purpose. Reactions were performed in 100 µL mixtures (10 µL mixtures usually used for colony PCR) and Milli-Q water instead of the template DNA served as negative control.

The PCR setup and the corresponding program are given in the table below (Table 2.37). The hybridization temperature X was based on the CG-content of the primer pair, the elongation time Y and cycles were determined by the length of the amplicon (the 30 s per 1,000 bps, 30 cycles per 1,000 bps). PCR products were purified using the DNA Clean&Concentrator^{TM}-5 (ZYMO RESEARCH, Freiburg, Germany) following the protocol provided.

Table 2.36 PCR approach for Q5 High-Fidelity Polymerase.

Components	Volume (µL)
5× Q5® High GC Enhancer (BioLabs, Feldkirchen, Germany)	20
5× Q5® Reaction buffer (BioLabs, Feldkirchen, Germany)	20
0.10 mM dNTPs (Thermo Scientific, Waltham, MA, USA)	2
Q5® High-Fidelity DNA Polymerase (BioLabs, Feldkirchen, Germany)	1
10 pmol Forward Primers	5
10 pmol Reverse Primers	5
Template (200-500 ng for genomic DNA and 40 ng for plasmid DNA per 100 µL reaction.)	
Add ddH_2O until the final volume reach to 100 µL	

Table 2.37 Q5 PCR program.

Step		Temperature	Time
Initial denaturation		98.0°C	30 s (Normal PCR) 5 min (Colony PCR)
30 x	Denaturation	98.0°C	10 s
	Annealing	X°C	20 s
	Elongation	72.0°C	30 s/1 kbp
Final elongation		72.0°C	5 min
Cool		4°C	

2.3.3 Plasmid and genomic DNA isolation

2.3.3.1 Plasmid DNA isolation

Plasmid DNA was isolated from a 5 mL *E. coli* XL1 or EPI300 overnight culture (LB medium) using the Monarch® Plasmid Miniprep Kit (BioLabs, Feldkirchen, Germany) as described in the manufacturer's protocol. All recommended washing steps were performed, except the plasmid binding column was centrifuged empty for 1 min and air-dried for 5 min after the washing steps to increase the yield of plasmid DNA. The plasmid DNA was eluted into 30 µL pre-warmed ddH_2O (60°C) and frozen at -20°C.

2.3.3.2 Genomic DNA preparation from microbial consortia

For metagenomic DNA preparation, a CTAB-chloroform: isoamyl alcohol-based method was applied to our study as Lebuhn *et al.* (2016) described since it performed well in reducing humic substance contamination. Specifically, microbial consortia and fiber material from 500 mL cell culture were harvested by centrifugation at 9,000 g for 5 min at 4°C. The pellets were further resuspended in 50 mL 50 mM ice-cold PBS and centrifuged at 200 g for 5 min. This step was repeated three times to remove the fiber material and other organic substances. The supernatant from each washing step was collected and centrifuged at 9,000 g for 15 min. The resulting cell pellets were further used for metagenomic DNA preparation. For cell lysis, 10 g cell pellets were resuspended in 25 mL CTAB-containing DNA extraction buffer containing 2 mg proteinase K and 0.8 mg RNase (Table 2.11). The mixture was horizontally shaken at 180 rpm and 37°C for one hour, followed by adding 3 mL of 20% (w/V) sodium dodecyl sulfate. Then the sample was incubated at 65°C for 2 h with gentle end-over-end inversion every 15 min. The cell debris was pelleted by centrifugation at room temperature for 10 minutes at 6,000 rpm. Then, the resulting supernatants were transferred into a new falcon, while the rest of the pellets were extracted two more times by resuspending the pellets in 10 mL of DNA extraction buffer and 1.2 mL of 20% sodium dodecyl sulfate with the methods described above. The combined supernatants were mixed with chloroform-isoamyl alcohol (24: 1, v/v) in a ratio of 1:1 (v/v). After

centrifugation at 6,000 rpm for 10 min, the aqueous phase was transferred into a new falcon tube. The genomic DNA was precipitated by adding isopropanol (1:0.6, v/v) and further fished out. Then it was washed with 70% (v/v) ethanol and resuspended in 2 mL prewarmed deionized water after drying for few minutes to remove all the ethanol.

2.3.4 DNA fragmentation (Restriction reaction/DNA shearing)

2.3.4.1 Restriction reaction

In a cloning process, the DNA molecules/recombinant plasmids may contain specific nucleotide sequences (restriction sites) that can be recognized and cleaved by endonucleases. This process may result in two or more DNA fragments. Therefore, it is a fast method to check the cloning results. Buffers recommended by the manufacturer for the respective enzyme or enzyme combinations were used. The incubation time was 0.5 - 1 hour at 37°C/30°C. For higher amounts of DNA, the volumes of the other components were increased accordingly.

Table 2.38 Restriction reaction.

Components	Volume (µL)
DNA	1-5
Restriction enzyme	0.5
Buffer	1
H_2O	x
Total volume	10

2.3.4.2 Shearing of genomic DNA for metagenomic library construction

In order to clone high-molecular-weight environmental DNA (eDNA) into pCC1 fosmid, mechanical shearing of the eDNA into around 40 kb fragments was required. To do this, 80-100 µg eDNA was diluted to 50 µL with TE buffer in an EP tube. Place the needle tip in the tube's bottom and pull the plunger up until all the tube contents are drawn into the needle. Then press the plunger down to expel the sample back into the tube, repeat different times until the DNA was sheared into the desired size. The degree of shearing related to the times of shearing can be correlated and checked by pulsed-field gel electrophoresis (PFGE) to find the best parameters for obtaining the 40 kb of sheared DNA.

2.3.5 Agarose gel electrophoresis

2.3.5.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the analytical or preparative separation of DNA samples. 0.8%

(w/v) agarose gels were used for separation. Agarose was microwaved with 1x TAE buffer and poured to form a gel. DNA samples were mixed with DNA loading buffer (6× TriTrack® DNA Loading Dye; Thermo Scientific; Waltham, MA, USA) and applied to the gel. In addition, 5 µL of a DNA marker (GeneRuler® 1 kB DNA Ladder, Thermo Scientific; Waltham, MA, USA) was also applied. At 120 V, the DNA fragments were separated for 40 min. Then, the agarose gel was submerged in a 0.01% (w/v) ethidium bromide solution for 10 min to stain the DNA and then briefly rinsed in a water bath. Under UV light, the DNA bands were visualized. The desired DNA band can be cut out of the agarose gel with a razor if needed.

2.3.5.2 Pulsed-field gel electrophoresis (PFGE)

200 mL 1% agarose (Gold Agarose, Biozym Scientific, Hessisch Oldendorf, Germany) was prepared in 1x TBE buffer. The solution was cooled down to at least 55-60°C before pouring. The agarose chamber was sealed by dropping agarose solution into the corners and waiting until agarose was gelled, followed by pouring the complete agarose into the chamber and inserting the comb. After removing the comb, the marker (MidRange PFG Marker, BioLabs, Feldkirchen, Germany) was loaded onto the gel, and the gel's pocket with the MidRange PFG Marker was sealed with agarose. Upon transfering the gel into an electrophoresis tank filled with 2 L 1x TBE buffer cooled down to 14°C, the rest of the samples premixed with DNA Loading Dye were loaded into the pockets. Then, the gel was run under the following conditions:

Table 2.39 Setup for pulsed-field gel electrophoresis.

Step	Time
Initial time	1 s
Final Time	25 s
Run Time	15 h
Volt	6 V/cm

2.3.6 Purification of DNA (PCR purification/gel recovery)

The removal of DNA polymerase and free dNTPs from the PCR reaction was achieved using DNA Clean & Concentrator™-5 (ZYMO RESEARCH, Freiburg, Germany) according to the manufacturer's instruction. In order to obtain the single DNA band, the PCR products have to be purified from an agarose gel. For this purpose, gel electrophoresis was first carried out. Since mutations caused by ethidium bromide staining or UV radiation were unwanted, the staining step was avoided for the samples to be purified.

The marker was loaded into the first and last lanes, and 2-5 µL of the sample was loaded into pockets next between the two makers. The rest of the samples were loaded in the remaining lanes. Before staining, the gel part with markers and 2-5 µL samples were cut out with a scalpel and placed in the ethidium bromide bath. The bands containing the desired fragment size were then incised under UV light for marking. The whole gel was reassembled when the UV light was turned off, and the required fragments were cut out according to the positions of the marks. The purification was carried out with the Zymoclean Gel DNA Recovery Kit (ZYMO RESEARCH, Freiburg, Germany) using the protocol supplied. Milli-Q water was used instead of the elution buffer to dissolve the DNA.

2.3.7 Evaluation of DNA concentration

The concentration of DNA molecules was determined at a wavelength of 260 nm in the UV/VIS spectrometer (Thermo Fischer Scientific, Waltham, MA, USA). First, a blank adjustment was conducted with 1 µL of ddH_2O. Afterward, 1 µL DNA sample was pipetted onto the measuring holes, the DNA concentration was determined in ng μL^{-1}, the value of A260/A280 and A260/A230 can be used as an indicator of the DNA purity.

2.3.8 Cloning by Gibson assembly

Cloning of PCR products into the pET-24c (+) vector was achieved by the Gibson Assembly procedure using Gibson Assembly Master Mix (BioLabs, Feldkirchen, Germany). Gibson Assembly Master Mix, linear plasmid DNA and PCR product were mixed and incubated at 50°C for 60 min. In the reaction, 50-100 ng of plasmid DNA was used with a 2-3-fold molar excess of PCR product (pmol) to optimize the cloning efficiency. Following this, the resulting recombinant vector was directly transformed into *E. coli* XL1 by heat shock transformation. The DNA concentration of the PCR product was estimated from the intensity of the corresponding DNA band on an agarose gel. The molar DNA concentration was calculated using the following formula.

$$\text{pmol}\ (DNA) = \mu\text{g}\ DNA \cdot \frac{\text{pmol}}{660\ \text{pg}} \cdot \frac{10^6\ \text{pg}}{\mu g} \cdot \frac{1}{N}$$

600 pg/pmol: The average molecular weight of a nucleotide pair,
N: The number of nucleotides.

Table 2.40 Gibson Assembly mix.

Components	Amount
2x Gibson Assembly Master Mix	5.0 µL
Plasmid DNA	0.02-0.03 pmol
PCR product	0.04-0.06 pmol
Add ddH_2O until the final volume reach to 10 µL	

2.3.9 Metagenomic library construction

For metagenomic library construction, the high-molecular-weight environmental metagenomic DNA (eDNA) from the microbial consortia was prepared using the standard method ("see chapter 2.3.3.2"). The isolated DNA was then cloned into pCC1 Fosmid using the CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre Technologies, Maharashtra, India) as recommended by the manufacturer's instruction. Precisely, the eDNA extracted from SBP enriched biogas fermenter samples was sheared into approximately 40 kb fragments using Hamilton syringe treatment ("see chapter 2.3.4.2"). Then the DNA fragments were size-selected by Pulsed-Field Gel Electrophoresis (PFGE) ("see chapter 2.3.5.2"). After that, the ~40 kb DNA fragments were end-repaired using End-Repair kit followed by ligated into the Cloning-Ready pCC1FOS vector (vector and insert were in a molar ratio of 10:1), linearized at the unique Eco72 I site and dephosphorylated. Ligations were packaged into lambda phage heads using MaxPlax Lambda Packaging Extracts according to the manufacturer's instructions. The packaged CopyControl Fosmid clones (titer of the phage particles) were determined according to the manufacturer's protocol, further plated on the LB plate containing 12.5 µg mL^{-1} chloramphenicol and incubate at 37°C overnight to select for the CopyControl Fosmid clones. The recombinant colonies were inoculated to 96-deep well plates containing 1 mL LB medium with 12.5 µg mL^{-1} chloramphenicol and incubate at 37°C, 180 rpm overnight. Then 100 µL of growth were stored in 96-well microtiter plates at -70°C with sterile glycerol to a final concentration of 25%. In addition to the newly constructed libraries, existing metagenomic libraries (Table 6.2), constructed in either the pCC1 Fosmid or pCT3FK Fosmid, were also used in our study.

2.3.10 Transformation

2.3.10.1 Heat shock transformation

For transformation of plasmid DNA into *E. coli* XL1/ EPI300, chemically competent cells of the *E. coli* XL1/EPI300 ("see chapter 2.2.6") were thawed on ice. 1 µL of plasmid or 5 µL of Gibson Assembly reaction was added to 50 µL of the competent cells, mixed gently and incubated on ice for 30 min. After heat shock for 1 min at 42°C, the cells were incubated on ice for 5 min, followed by adding 1 mL of pre-warmed LB medium (37°C) and incubating at 37°C, 180 rpm for 1 h. Subsequently, 100 µL and 900 µL of the bacterial culture were plated out on LB agar plates containing 50 µg mL^{-1} kanamycin, respectively. The LB agar plates were further incubated at 37°C overnight.

2.3.10.2 Electroporation transformation

For electroporation transformation, 5 mL LB was inoculated with 30 µL of fresh overnight culture. Following this, the inoculum was shaken at 37°C and 180 rpm for 3 h. The cell pellets were harvested by centrifuging at 11,000 rpm and 4°C for 30 s with a MIKRO 22R (Hettich, Tuttlingen, Germany). The supernatant was discarded, and the cell pellets were resuspended with 1 mL of chilled, sterilized ddH_2O by pipetting up and down three times to resuspend the pellets fully. The washing step was repeated three times. For the last washing step, leave 20 to 30 µL of water in the tube. Resuspend the cell pellets and place the EP tube on ice. Then, 1-5 µL of DNA was added to the cell suspension. Mix briefly and keep the tube on ice. The cell suspension was transferred from the tube to the cooled electroporation cuvette. Electroporation was done at 2,500 V, 25 µF, 200 ohms. This setting is for a 2 mm cuvette. After electroshock, 1 mL of prewarmed LB medium was added to the tube, and the cell suspension was incubated at 37°C for one hour with shaking before the transformation mixture was plated onto the selection agar plates.

2.3.11 DNA-sequencing

2.3.11.1 Illumina Shotgun sequencing

Based on a functional metagenomic library screening, the positive fosmid clones, showing activity against one or two chromogenic substrates, were selected and subjected to a hybrid Sanger/Illumina sequencing as described by Lam *et al.* (2014). Fosmid DNA was isolated from *E. coli* EPI300 using a Monarch® Plasmid Miniprep kit ("see chapter 2.3.3.1"). For Illumina Shotgun sequencing, two fosmid library pools were created based on the vector backbone, including 40 pCC1- and 15 pCT3FK-based

recombinant fosmids, respectively. Each library pool included the same concentrations of each fosmid and had a final total DNA concentration of 20 µg mL^{-1}. Illumina shotgun sequencing was performed at ZIEL-Core Facility Microbiome/NGS, TUM. The Illumina DNA Prep Kit (Illumina Inc, San Diego, USA) was used to prepare amplicon libraries according to the manufacturer's instructions. The Illumina NovaSeq System was applied for sequencing. Fosmids were sequenced to a read depth of approximately 50,000-fold coverage. After sequencing, the obtained reads were assembled by using Unicycler v0.4.8.0. By manually BLAST analysis of the acquired contigs against the NCBI database, the vector backbone and contaminating *E. coli* genomic DNA sequences were removed.

2.3.11.2 Sanger end sequencing

Sanger sequencing, also known as the dideoxy method or chain termination method, is a well-established, low-cost method for sequencing dsDNA and ssDNA created in 1977 by Frederick Sanger and coworkers (Sanger *et al.*, 1977). For this purpose, the sequenced DNA is usually cloned into a vector or a purified PCR product ("see chapter 2.3.6"). DNA Sanger sequencing was carried out either at Eurofins Genomics (Ebersberg, Germany) from a total of 15 µL PCR product with a concentration of 10 ng $µL^{-1}$, or at GeneWiz (Leipzig, Germany) from 5 µL PCR product with a concentration of 10-50 ng $µL^{-1}$. For sequencing, one of the primers, which served to amplify the target DNA, was added to the purified PCR product or recombinant vector as follows: 2 µL primer with 10 pmol $µL^{-1}$ (Eurofins Genomics) or 5 µL primer with 5 pmol $µL^{-1}$ (GeneWiz). The sequencing primers used were standard primers for specific vectors or the primers used for gene cloning (Table 2.35). The sequencing chromatograms were evaluated and assembled using BioEdit Sequence Alignment Editor (BIOEDIT LTD, Manchester, United Kingdom) or Clone manager (Sci Ed Software LLC, Westminister, CO, USA). Specifically, for pooled sequencing strategy, the generated terminal 1,000 bp of each fosmid insert was used for assigning the contigs to the individual fosmid clones.

2.3.11.3 16S rRNA amplicon sequencing

All samples for 16S rRNA amplicon sequencing were generated by centrifuging 2 mL cultures at 13,400 rpm for 5 minutes, then resuspending the pellets in 600 µL RNAlater and storing them at -20°C for sequencing. Using an Illumina MiSeq technology, high-throughput sequencing of the V3-V4 gene region of the bacterial 16S rRNA was performed at ZIEL, Core Facility Microbiome/NGS, TUM, Freising, Germany. Amplicon library preparation and the following sequencing process were performed following the description from Reitmeier *et al*. (2020b). The sequencing was performed with 24 ng

extracted DNA as a template and a two-step PCR amplification technique (15 cycles for each PCR program; Primers 431F and 785R in Lagkouvardos's study) (Berry *et al.*, 2011; Lagkouvardos *et al.*, 2015). Sequencing was carried out in paired-end mode with purified DNA spiked-in using 20% (v/v) PhiX standard DNA in an Illumina MiSeq system prepared according to the manufacturer's instructions after purification using AGENCOURT AMPure XP beads (Beckman Coulter, Krefeld, Germany) (Bazanella *et al.*, 2017).

2.3.12 Sequencing results analysis

2.3.12.1 ''Pool Sequencing'' results analysis

The ORF searching function in Clone Manager (Sci Ed Software LLC, Westminister, CO, USA) was used to identify the ORFs encoded on the contigs (from Shotgun sequencing). Searches for carbohydrate-active enzymes and CAZyme gene cluster (CGC) were performed by searching against the dbCAN database with default parameters. The predicted amino acid sequences were submitted to the KEGG database for further functional analysis (E-value 1.0E-5). Signal peptides were predicted by SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP/). The Pfam database (http://pfam.xfam.org/) was used to estimate the enzymes' conserved domains. Clone manager was used to predict molecular weights and isoelectric points (pI) of the proteins.

2.3.12.2 16S rRNA amplicon sequencing results analysis

The 16S rRNA amplicon sequencing raw data analysis was carried out with the IMNGS (Integrated Microbial Next Generation Sequencing), which was provided by ZIEL, Core Facility Microbiome, TUM (Reitmeier *et al.*, 2020a; Reitmeier *et al.*, 2020b). This evaluation pipeline was developed at the Technical University of Munich by Lagkouvardos *et al*., Specifically developed as an easily accessible pipeline to analyze prokaryotic 16S rRNA amplicon sequencing (Lagkouvardos *et al.*, 2016). The pipeline comprises an assembly of paired reads and quality filtering and length, chimeras & redundancy filtering, also known as demultiplexing. It is followed by taxonomic classification. In addition to the standardized pipeline, the following settings were made: The number of allowed mismatches of the barcodes, 2; Minimum Fastq quality score for trimming unpaired reads, 3; Length of single reads or amplicons for paired, overlapping sequences, 350-600 bp; Trim (forward and reverse), 10 bp; Minimum relative abundance of OTUs, 0.0001-0.0025. The evaluation of the raw data via IMNGS.org enabled a statistical evaluation via RHEA in R (https://www.r-project.org/) (Lagkouvardos *et al.*, 2017).

2.4 Protein biochemical methods

2.4.1 Protein production in *E. coli*

2.4.1.1 ArcticExpress as the expression host

An ArcticExpress transformant bearing the recombinant plasmid was inoculated into 5 mL of lysogenic broth (LB) medium supplemented with 20 μg mL^{-1} gentamycin (Selection for the *cpn10*/*cpn60* chaperonin expression plasmid) and 50 μg mL^{-1} kanamycin and was grown aerobically at 37°C overnight. Then, the 5 mL overnight culture was inoculated into 1 L of LB and incubated with shaking (180 rpm) at 30°C for 6 h before inducing the recombinant protein expression with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 13°C for 24 h.

2.4.1.2 BL21 as the expression host

For pure protein isolation, 30 mL LB medium with the appropriate antibiotic (50 μg mL^{-1} kanamycin for pET24c, 0.2 % glucose & 100 μg mL^{-1} ampicillin for pMAL vector) were inoculated with a single colony, followed by shaking at 37°C overnight to obtain a preculture. The OD_{600} of the preculture was first measured. Then it was inoculated into a 1 L fresh LB medium to obtain a starting OD_{600} of 0.1-0.2. The inoculated medium was incubated at 37 ° C, 180 rpm until the OD_{600} of the culture was around 0.5. Induction was carried out by adding IPTG (concentration depending on the protein, usually 0.1 mM), followed by further incubation at 37°C/30°C, 180 rpm for 6 h/overnight (lower concentration of IPTG, temperature and shaking speed can reduce the formation of inclusion body). When the induction process was finished, the cell pellets were harvested by centrifugation at 13,400 rpm for 30 min, and resuspended in 25 mL of 50 mM phosphate buffer (pH 8). Then the cells were disrupted by sonication (settings: 50 %, 5 minutes, interval 0.4) with a Hielscher UP200S apparatus (Hielscher Ultrasonics GmbH, Teltow, Germany). The suspension, also called crude extract (CE), was centrifuged with a Sorvall LYNX 6000 centrifuge (Thermo Scientific, Waltham, MA, USA) at 13,400 rpm and 4°C for 30 min, and the cell-free supernatant was filtered through a Filtropur S 0.45 filter (SARSTEDT, Nümbrecht, Germany) using a syringe, which was then called filtered supernatant.

2.4.2 Protein purification by immobilized metal affinity chromatography

A recombinant clone, constructed using the pET24c vector, can express a recombinant protein with a His-tag under the IPTG induction. The purification of such proteins was done via immobilized metal

affinity chromatography (IMAC). For this purpose, a Protino® Ni-TED 2000 Packed Columns kit (MARCHERY-NAGEL GmbH & Co.KG, Düren, Germany) was used according to the manufacturer's instruction. Specifically, the column was equilibrated with 4 mL of 1x LEW buffer. The filtered SN then flowed through the column. After washing twice with 4 mL 1x LEW buffer, the target protein was eluted three times with 3 mL 1x elution buffer.

2.4.3 Protein purification by anion exchange chromatography

After IMAC, the proteins were further purified by anion exchange chromatography using a linear gradient from 0 to 1 M NaCl on an ÄKTA pure 25L1 FPLC system (GE Healthcare Life Sciences, UK) equipped with a 1 mL HiTrap™ QFF column. SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was used to determine the size and purity of the proteins, which was followed by pooling the fractions containing pure target protein. The protein concentration was determined using the Bradford assay (“see chapter 2.4.5”), which used bovine serum albumin as a reference.

2.4.4 Concentrating and desalting of the protein

In order to concentrate the protein, a protein cut-off column (Sartorius, Göttingen, Germany) was applied, which was allowed to concentrate the protein based on its size. Specifically, the VIVASPIN TURBO 4 (5,000 MWCO membrane) was chosen for most proteins in our study. After transferring the protein solution into the column, it was centrifuged in a UNIVERSAL 320 R at 4,500 rpm for 30 min-1 h until it reached the desired concentration. Besides, for the large molecular size of proteins, 30,000 and 50,000 MWCO membrane inserts can also be chosen.

Sephadex™ G-25 M PD-10 Desalting Column (GE Healthcare, Munich, Germany) was prepacked and designed for desalting, buffer exchange of a protein solution, which was adopted in our study to desalt the protein solution according to the manufacturer’s introduction. The buffer was chosen based on a protein’s isoelectric point. When the buffer pH is equal to the pI (isoelectric point), the protein has no charge. Consequently, the protein is not able to bind to an ion-exchange column (both anion and cation). When buffer pH < pI, a positive charge was generated, cation exchange can be used for protein purification. By contrast, a negative charge can form when buffer pH > pI, in this case, anion exchange is a good option.

2.4.5 Protein concentration measurement

The Bradford assay was used to determine the protein concentration (in mg mL^{-1}) with a photometer. The Coomassie Brilliant Blue dye forms complexes with proteins' nonpolar and cationic side chains, shifting the absorption maximum from 479 nm to 595 nm. The color of the reaction shifts from reddish to bluish based on the amount of protein, so calibration with a specific protein is necessary to determine the concentration. In this work, a calibration curve was determined with the standard protein Bovine Serum Albumin (BSA) and Bradford reagent (SERVA, Heidelberg, Germany). For this purpose, 10-100 µL of the sample were added to a 1 mL cuvette which is filled with 1 mL 1× Bradford reagent. After mixing, the sample was incubated for 5 min, shielded from light, and then the OD_{595} was measured with 1 mL of 1× pure Bradford reagent as a reference. Similarly, BSA was diluted to a concentration of 10 µg $µL^{-1}$ and pipetted into the cuvettes to get a desired amount of BSA, and then the OD_{595} was measured in the same way. The protein concentration in the cuvette could subsequently be determined based on a standard curve created with BSA. The standard curve was plotted with BSA concentrations of 0-150 µg/cuvette. The OD_{595} was then plotted against the quantity of BSA, and formula was created using Excel's linear range, which could then be used to compute the protein content of the samples. The obtained value was then divided by the sample volume used to calculate the protein concentration.

$$\text{Protein concentration } [\mu g\ \mu L^{-1}] = \frac{\text{Protein concentration } [\mu g/\text{Kuvette}]}{\text{Protein volume } [\mu L/\text{Kuvette}]}$$

2.4.6 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate proteins according to their molecular weight and analyze the purity of the recombinant proteins. To improve separation, discontinuous SDS gels with a 5 % collection gel and a 10 % separation gel were used. For gel preparation (Table 2.17), the separation gel (approx. 4 mL) was first prepared and overlaid with 1 mL isopropanol. After it had polymerized, the isopropanol was poured off, and the collecting gel was poured over the separating gel. The sample comb was then stuck into the collecting gel. For sample preparation, 20 µL protein sample was mixed with 4 µL 5x SDS loading buffer (Table 2.18) and incubated for 10 min at 95°C. 10 µL of the protein samples were then added to the gel pockets. As a reference, PageRuler™ Prestained Protein Ladder (Thermo Scientific, Waltham, MA, USA) was used to determine the protein size. The proteins were separated in 1× SDS

running buffer (Table 2.19) with a PowerPRO 300 Power Supply at a voltage of 80 V for 20 min and 120 V for around 1 h until the blue band left the gel. The gel was stained overnight in staining solution (Table 2.20) and then de-stained for 4 h in de-staining solution (Table 2.21). A ScanMaker 1000XL (MICROTEK, Udyog Nagar, India) was used to scan the gel.

2.4.7 Enzyme assay

The ability of the enzymes to perform catalysis was determined by assaying the enzymes with artificial substrates (Table 2.24), natural substrates (Table 2.27), and synthetic substrates (Table 2.25 and Table 2.26), respectively, followed by measuring products released both spectrophotometrically (*p*NP assay or DNS assay) and using thin-layer chromatography and/or high- performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Before initiating the enzyme reaction, all insoluble substrates (including WAX-I and DA) were washed with Milli-Q water by centrifugation to eliminate the soluble reducing sugars.

2.4.7.1 Enzymatic hydrolysis of polysaccharides

To test the hydrolytic activity of the purified enzymes against different polysaccharides. The enzyme reactions were performed in triplicates by incubating enzymes with different substrates, while the reaction either without enzymes or without substrate was set up as a negative control. The pH of the enzyme reaction was maintained by using 25 mM buffer. The buffer used in the enzyme reaction include 250 mM citrate phosphate buffer (pH 2.0-9.0), 250 mM sodium acetate buffer (pH 4.0-5.5), 250 mM phosphate buffer (pH 6.0-7.5) and Tris-HCl buffer (pH 8.0-9.0). The reactions were performed at the optimal temperature with 600 rpm shaking on an Eppendorf thermomixer. After incubation, the enzyme reactions were stopped by incubation on ice for 5 min, and the hydrolysates were centrifuged at 4°C, 13,400 rpm for 5 min to precipitate insoluble substrate components. The supernatant can be used for DNS assay. For time-course analysis of the hydrolysis products, the reactions were stopped by boiling the sample for 10 min. For DNS Assay, we preferably prepared samples in a 1.5 mL EP tube with 200 µL total reaction volume (more than three times the amount of each required assay volume to reduce the standard deviation). For detection of synergism between enzymes, polysaccharides were hydrolyzed with up to six enzymes simultaneously. For chromatographic separation and analysis of the hydrolysates, the hydrolysis products were undiluted/diluted to a final volume of 500 µL with water depending on the concentration of reducing sugar released from a polysaccharides.

Table 2.41 Enzyme reactions for the hydrolysis of polysaccharides.

Component	Volume		
1% polysaccharide solution	100 µL	250 µL	500 µL
250 mM buffer	20 µL	50 µL	100 µL
Enzyme solution	X µL	X µL	X µL
ddH_2O	(80 - X) µL	(200 - X) µL	(400 - X) µL
Total volume	**200 µL**	**500 µL**	**1000 µL**

2.4.7.2 Determination of the enzyme activity with DNS assay

The enzyme activity was measured photometrically with the DNS assay after the enzymatic hydrolysis of polysaccharides. In the reaction, 3,5-dinitrosalicylic acid (DNS) is reduced by the free carbonyl group (C=O) of reducing sugars to 3-amino-5-nitrosalicylic acid. The color changes from orange to brown during this reaction, and the concentration of reducing sugars can be estimated correspondingly by measuring the change in absorbance at 540 nm. As a result, increasing the concentration of reducing sugars in the hydrolysate can be used to assess the enzymatic hydrolysis of the polysaccharides used. The DNS assay was carried out in 96-well PCR plates. 50 µL of hydrolysates were mixed with 75 µL of DNSA solution (Table 2.22). A calibration curve for arabinose concentration determination was generated by nine different concentrations of arabinose reference (250 µg mL^{-1}, 500 µg mL^{-1}, 750 µg mL^{-1}, 800 µg mL^{-1}, 900 µg mL^{-1}, 1000 µg mL^{-1}). The PCR plate was sealed with an PCR adhesive foil and incubated at 95° C for 10 min in a thermal cycler and cooled on ice for 5 min. After centrifugation at 1,000 g for 30 s, 75 µL of the reaction were transferred to a 96-well microtiter plate. The absorption at 540 nm was then measured in the UV/VIS spectrometer, and the amount of reducing sugars was determined in arabinose equivalents using the arabinose standard curve. The amount of arabinose in the negative control (only substrate in the reaction) was subtracted from the amount of arabinose in the hydrolysate, and the enzyme activity was calculated using the following formula.

$$\text{Specific activity (µmol min}^{-1}\text{ mg}^{-1}) = \frac{[\mathrm{m1(\mu g)} - \mathrm{m2(\mu g)}] \cdot 1000}{\mathrm{M_{Arabinose}\ (\mu g\ \mu mol^{-1}) \cdot V_E(\mu L) \cdot C_E(\mu g\ \mu L^{-1}) \cdot t(min)}}$$

$M_{arabinose}$: Molecular mass of arabinose (150.13 g mol^{-1}),

V_E: Volume of the enzymes in the reaction (µL),

C_E: Concentration of the enzymes in the reaction (µg $µL^{-1}$),

m1: The amount of arabinose released in the reaction (calculated from standard curve),

m2: The amount of arabinose released in the negative control (calculated from standard curve).

t: Incubation time (min).

2.4.7.3 Determination of the enzyme activity with *p*NP assay

In addition to the hydrolysis of polysaccharides, the enzyme activities of GHs were determined by the hydrolysis of para-nitrophenyl substrates (Table 2.24). For *p*NP substrate, when the glycosidic bond is hydrolyzed, the yellow para-nitrophenol (*p*NP) can be released, which absorbs light with a wavelength of 405 nm. The *p*NP assay was carried out in a total volume of 50 µL reaction in triplicates by incubating enzymes with 25 mM buffer and 1 mM *p*NP substrate. In addition, negative control without enzyme was carried out. The specific activity was determined with incubation times of 10 min to 2 h at the optimal condition of the enzyme. After the incubation, the reaction was stopped by adding twice the volume of 1 M Na_2CO_3 and centrifuging at 2,000 g for 5 min. 100 µL reaction was then transferred to a microtiter plate, and the absorption at 405 nm was read in the UV/VIS spectrometer. The concentration of released *p*NP was determined using a reference line made up of six different concentrations of *p*NP (0 µmol, 0.022 µmol, 0.043 µmol, 0.065 µmol, 0.086 µmol, 0.108 µmol). The released amount of *p*NP substance in the negative control was subtracted from the total amount of *p*NP in the hydrolysates and the enzyme activity was calculated using the following formula.

$$\text{Specific activity } (\mu\text{mol}\cdot\text{min}^{-1}\cdot\mu\text{g}^{-1}) = = \frac{(\mathrm{m1}_{p\mathrm{NP}} - \mathrm{m2}_{p\mathrm{NP}})\ (\mu\mathrm{g})}{\mathrm{M}_{p\mathrm{NP}}\ (\mu\mathrm{g}\ \mu\mathrm{mol}^{-1}) * \mathrm{t(min)} * \mathrm{V_E}(\mu\mathrm{L}) * \mathrm{C_E}(\mu\mathrm{g}\ \mu\mathrm{L}^{-1})}$$

$m1_{pNP}$: The amount of *p*NP (µg) released in the reaction,
$m2_{pNP}$: The amount of *p*NP (µg) released in the negative control,
M_{pNP}: Molecular weight of 4-Nitrophenol (139.11 µg µmol^{-1}),
t: Incubation time (min),
V_E: Enzyme volume (µL),
C_E: Enzyme concentration (µg µL^{-1}).

Table 2.42 Reaction of *p*NP assay.

Enzyme mixture	250 mM Buffer	5 µL	20 µL
	Enzyme solution	X µL	X
	H_2O	(25-5-X) µL	(100-20-X) µL
Substrate mixture	0.1 M *p*NP	0.5 µL	2 µL
	H_2O	24.5 µL	98 µL
	Total volume	50 µL	200 µL

2.4.7.4 Determination of optimal temperature and pH

The samples were prepared as described above and pipetted into a 96-well PCR plate, sealed with PCR adhesive foil to avoid evaporation. To determine the optimal temperature, a buffer (pH=6.0) was applied to the reaction since this pH is very close to the cellular environment. In a TAdvanced PCR

thermo cycler (Analytik Jena, Göttingen, Germany) with a 40-degree gradient ranging over 12 well-rows, we incubated the reaction with a temperature gradient ranging from 25°C to 90°C for 10 min-60 min (The incubation duration depends on the activity of the enzyme and the substrate employed). Two runs are sufficient for the full spectrum, including running 1: 25°C-65°C; running 2: 55°C-90°C. Afterward, the amount of reducing sugar released was optical measured via DNS assay at 540 nm or colorimetric with the *p*NP-AF substrate at 405 nm. Then, the pH dependence at the temperature optimum of the enzyme was determined with the same method. The pH of the reactions was maintained by adding 25 mM buffer at a pH range between 2.0-9.0 (Buffer pH was adjusted at the temperature applied). The pH and temperature optimum for the activity of the enzyme were combined using a pH-vs-temperature profile. The measurement was carried out in three independent, experimental replicates.

2.4.7.5 Thermostability of the Enzyme

To determine the thermo-resistance of the enzymes, the loss of activity at a different time interval of a long incubation time was determined in the absence of substrate. This assay was performed with polysaccharides or *p*NP-arabinofuranoside (*p*NP-AF). In the enzyme assay, a master mix with over 3 times the concentration of the enzyme used in a standard assay was incubated in 25 mM buffer solution for 24 h/48 h at different temperatures (Center around their optimal temperature). 200 µL master mix was removed and store at 4°C at different time points. When the last reaction was finished, all the samples were incubated with a 0.5% (w/v) polysaccharide or 1 mM *p*NP-AF for 10 min-60 min at their optimal temperature, 600 rpm for a particular period incubation. In addition, a reaction that has not been heat-treated was chosen as zero-timepoint. The enzyme activity was determined with the DNS assay or *p*NP assay. The enzymes' resistance to thermo-inactivation was determined by measuring their activity against substrate before and after heat treatment for different periods.

2.4.7.6 Determination of kinetic parameter

Kinetic parameters, including Michaelis Menten constant (K_m) and maximal velocity of the substrate digestion (V_{max}), were determined mainly on SBA, washed DA and WAX-RS for enzymes in our study. The enzyme reactions were set up by incubating at the enzyme's optimal temperature and pH, under shaking at 600 rpm using an Eppendorf thermomixer. Specifically, 200 µL of enzyme reaction contains 25 mM buffer, the various concentration of substrates, which is usually ranging between 0-50 g L^{-1} (50 g L^{-1}, 45 g L^{-1}, 40 g L^{-1}, 35 g L^{-1}, 30 g L^{-1}, 25 g L^{-1}, 15 g L^{-1}, 10 g L^{-1}, 7.5 g L^{-1}, 5 g L^{-1}, 2.5 g L^{-1}, 1 g L^{-1}), and

two or three different concentrations of enzymes, after 30 min to 2 h incubation at their optimal temperature and pH (Table 2.44). The reactions were stopped by cooling down on the ice, followed by centrifuging at 4°C, 13,400 rpm for 5 min. Then, as described above, 50 µL of supernatant was mixed with 75 µL of DNS reagent in 96-well plates for the DNS assay and 540 nm absorbance was readout. All assays were carried out in triplicate. The kinetic parameters were calculated based on the Michaelis-Menten formula (shown below) using Microsoft Solver non-linear regression programs.

$$r = \frac{V_{max} \cdot [s]}{K_m + [s]}$$

K_m: Michaelis Menten constant,

V_{max}: the maximal velocity of the substrate digestion,

[s]: The concentration of a substrate,

r: Rate of formation of product.

Table 2.43 Enzyme reaction for kinetic parameter.

Final substrate conc. (w/v)	10% Substrate solution (µL)	0.25 M Buffer (µL)	ddH_2O (µL)	Enzyme(µL)
5.00%	100	20	80-X	X
4.50%	90	20	90-X	X
4.00%	80	20	100-X	X
3.50%	70	20	110-X	X
3.00%	60	20	120-X	X
2.50%	50	20	130-X	X
2.25%	45	20	135-X	X
2.00%	40	20	140-X	X
1.75%	35	20	145-X	X
1.50%	30	20	150-X	X
1.25%	25	20	155-X	X
1.00%	20	20	160-X	X
0.75%	15	20	165-X	X
0.50%	10	20	170-X	X
0.25%	5	20	175-X	X
0.10%	2	20	178-X	X
0%	0	20	180-X	X

2.4.7.7 Enzymatic cleavage mode against oligosaccharides

The oligosaccharides A3, O-A4B, O-A5B, O-XBI, O-XTR, O-XTE, O-XPE, O-XHE, O-A3X, O-AX3, O-XAXX MIX, O-XA3XX, O-A2X3, O-XA23XX were used in the assay to investigate the enzymatic action mode and preference towards certain side chains of arabino-/xylo-oligosaccharides. Each 10 µL or 20 µL mixture contains 0.25 g L^{-1} oligosaccharide, 50 mM buffer and enzyme (Table 2.45). The reaction was incubated at their optimal temperature and pH in a PCR thermocycler for 12 h or 24 h. Specifically, for

the time-course analysis of the hydrolysis product, 10 µL of the reaction mixtures were collected at specific intervals (5 min, 10 min, 30 min, 1 h, 5 h and 20 h), then the reactions were stopped by boiling at 95°C for 10 min.

Table 2.44 Enzyme reaction with oligosaccharides as substrate.

Component	Volume	
1% Substrates	2.5 µL	5 µL
0.25 M Buffer	2 µL	4 µL
Enzyme	X µL	X µL
ddH_2O	(3-X) µL	(11-X) µL
Total volume	10 µL	20 µL

2.4.7.8 Influence of metal Ions, denaturants and hydrolysis products on enzymatic activity

For examining the influence of metal ions ($CuCl_2$, $CoCl_2$, $CaCl_2.2H_2O$, $FeCl_2.4H_2O$, KCl, $MgCl_2.6H_2O$, $MgSO_4$, $MnCl_2.4H_2O$, $NiSO_4$, NaCl, $ZnCl_2$), denaturants (SDS, Urea), and chelator (EDTA), these substances at three different final concentrations (1 mM, 5 mM, 10 mM), were incubated with the enzyme in 25 mM citrate phosphate buffer at room temperature for 2 h, negative controls which contained only individual metal compounds or only enzyme were also included in the study (Table 2.45). Upon incubation, 1 mM *p*NP-AF was added to the enzyme reaction and residual activities were determined with the colorimetric *p*NP assay described above.

The effect of arabinose on arabinosyl hydrolases was investigated by incubating the enzymes with four different concentrations of arabinose (67 mM, 133 mM, 266 mM or 533 mM) for two hours. In the negative control, using water instead of arabinose. 1 mM *p*NP-AF was added to the reaction mixtures after incubation, and residual activities were evaluated using the colorimetric *p*NP assay described above.

Table 2.45 List of chemicals used for the influence of metal ions, denaturants and chelator on enzymatic activity.

Metal ion	Compound	Company
Ca^{2+}	$CaCl_2 \cdot 2H_2O$	VWR International, Radnor, USA
Co^{2+}	$CoCl_2$	Merck, Darmstadt, Germany
Zn^{2+}	$ZnCl_2$	Sigma-Aldrich, St. Louis, USA
Na^{+}	NaCl	Roth, Karlsruhe, Germany
K^{+}	KCl	Thermo scientific, Waltham, MA, USA
Cu^{2+}	$CuCl_2$	Thermo scientific, Waltham, MA, USA
Mg^{2+}	$MgCl_2 \cdot 6H_2O$	Applichem, Darmstadt, Germany
Mg^{2+}	$MgSO_4$	Sigma-Aldrich, St. Louis, USA
Mn^{2+}	$MnCl_2 \cdot 4H_2O$	Thermo scientific, Waltham, MA, USA
Fe^{2+}	$FeCl_2 \cdot 4H_2O$	Thermo scientific, Waltham, MA, USA
EDTA	$C_{10}H_{16}N_2O_8$	Applichem, Darmstadt, Germany
Urea	CH_4N_2O	Roth, Karlsruhe, Germany
SDS	$NaC_{12}H_{25}SO_4$	Applichem, Darmstadt, Germany
Ni^{2+}	$NiSO_4$	Sigma-Aldrich, St. Louis, USA

2.4.7.9 Determination of synergistic effects of enzymes

The synergistic effects of the enzymes were tested by using arabinose-containing substrates, including sugar beet arabinan (SBA), debranched arabinan (DA) and/or wheat arabinoxylan (for reducing sugar assay, WAX-RS). Specifically, for *Am*Araf51 and *Am*Araf43 from *A. mesophilus*, four different reaction types were set up (Table 2.46). The first one was the negative control which only contains a 0.5% substrate buffer solution (without enzyme), including 0.5% SBA or 0.5% WAX-RS. The second was the single enzyme reacted with the respective substrate (*Am*Araf51, *Am*Araf43 were incubated with SBA and WAX-RS, respectively; *Pp*Abn53 and M_Xyn10 were specifically incubated with SBA or WAX-RS, respectively), which was used to compare the synergistic effect with the single enzymatic treatment. For the test group, we performed two different types of enzyme assays. The first one is the so-called "one batch treatment," where the endo- and exo-acting enzymes (*Pp*Abn43 were incubated with *Am*Araf51 or *Am*Araf43 with SBA as substrate, M_Xyn10 was incubated with *Am*Araf51 or *Am*Araf43 with WAX-RS as substrate) were added in the reaction simultaneously. The second one is the stepwise treatment. In this treatment, the substrates were cleaved by the endo-active enzyme (*Pp*Abn43 for SBA or M_Xyn10 for WAX-RS) for 12 hours. Then, the samples were boiled to denature the endo-active enzyme, thus preventing any further activity. After cooling down, the exo-acting enzyme (*Am*Araf51 and *Am*Araf43) was added to the mixture and further incubated to a maximum time of 12 h.

Table 2.46 All the reactions applied in our study.

Reaction types	SBA	WAX-RS
Negative control	No enzyme	No enzyme
Single enzyme treatment	*Pp*Abn43	M_Xyn10
	*Am*Araf51	*Am*Araf51
	*Am*Araf43	*Am*Araf43
One batch treatment	*Pp*Abn43+*Am*Araf51	M_Xyn10+*Am*Araf51
	*Pp*Abn43+*Am*Araf43	M_Xyn10+*Am*Araf43
Stepwise treatment	*Pp*Abn43→*Am*Araf51	M_Xyn10→*Am*Araf51
	*Pp*Abn43→*Am*Araf43	M_Xyn10→*Am*Araf43

In the metagenomic study, we used SBA and DA to explore the synergistic effects of the characterized enzymes. Up to six enzymes were added to the same enzyme reaction for exploring the enzyme synergism towards arabinan degradation. The reactions were performed at 45°C, pH 5.5 for 24 h.

For both synergistic effects exploration, the reactions were terminated by boiling for 10 min. The supernatant was obtained after centrifugation at 4°C for 5 min, which was further used for DNS assay or HPAEC-PAD.

2.5 Bioinformatical methods

2.5.1 Software used

In this work, raw data were analyzed with the following software and database, unless otherwise stated, with standard settings (Table 2.47).

Table 2.47 Software used in the study.

Software	Hersteller
BioEdit Sequence Alignment Editor	BIOEDIT LTD, Manchester, United Kingdom
Blast	https://blast.ncbi.nlm.nih.gov/Blast.cgi;
Clone manager	Sci Ed Software LLC, Westminister, CO, USA
SignalP-5.0 Server	http://www.cbs.dtu.dk/services/SignalP/; (Almagro *et al.*, 2019)
Pfam	http://pfam.xfam.org/; (Mistry *et al.*, 2020)
Microsoft Excel 2016	Microsoft Corporation, Albuquerque, NM, USA
Graph Pad Prism 7	https://www.graphpad.com/scientific-software/prism/;
IMNGS	https://www.imngs.org/; (Lagkouvardos *et al.*, 2016)
MEGA-X	https://www.megasoftware.net/; (Kumar *et al.,* 2016)
R	https://www.r-project.org/;
RHEA	https://github.com/Lagkouvardos/Rhea; (Lagkouvardos *et al.*, 2017)
dbCAN2	http://bcb.unl.edu/dbCAN2/; (Zhang *et al.*, 2018)
CAZy (Carbohydrate-Active Enzymes database)	http://www.cazy.org/; (Lombard *et al.*, 2014)
Kyoto Encyclopedia of Genes and Genomes (KEGG)	https://www.genome.jp/kegg/; (Ogata *et al.*, 1999)

2.5.2 In silico cloning

The genome sequence of N2K1 (*Acetivibrio mesophilus*) was kindly provided by Dr. Vladimir Zverlov (Lehrstuhl für Mikrobiologie, TUM). Cloning the genes into the pET-24c (+) vector was initially performed in silico using the Clone Manager 9.0 Professional program. The cloning and primers were also designed using Clone Manager 9.0. Open Reading Frames (ORFs) were searched by the Clone Manager ORF search function. The endonuclease used for restriction reaction was chosen based on restriction site analysis in the clone manager.

2.5.3 Signal peptide prediction and protein parameter determination

The SignalP 5.0 server (http://www.cbs.dtu.dk/services/SignalP/) was used to predict N-terminal signal peptides. Clone Manager 9 was used to determine the molecular mass and isoelectric point of the recombinant proteins.

2.5.4 Protein module prediction

Putative CAZymes were identified using the dbCAN database (http://bcb.unl.edu/dbCAN2/blast.php); Further modules of the proteins were determined with the Pfam database (https://pfam.xfam.org/); Specific functions of the putative CAZymes was predicted by KEGG database (https://www.genome.jp/tools/kofamkoala/).

2.6 Analytical methods

2.6.1 Thin-layer chromatography (TLC)

Enzymatic hydrolysis products, such as mono- and oligosaccharides, were chromatographically separated and identified using Thin-layer Chromatography (TLC). To do that, the hydrolysate with insoluble sugar as substrate was centrifuged at 13,400 g for 5 min to precipitate insoluble substrate components. The resulting supernatant was then thoroughly mixed and applied to TLC silica gel 60 20 cm × 20 cm aluminum foils (Merck, Merck, Darmstadt, Germany) in steps. A maximum of 5 µL of hydrolysate was loaded in one step and dried in the air before the next step. 1-15 µL of the hydrolysate was used for an unknown concentration (1.5 cm from the edge, 0.5 cm between samples). In addition, 1 µg (1 µL 1g L^{-1} in H_2O) of each known mono- or oligosaccharides were applied as a standard. The hydrolysis products were separated using a 6:7:1 (v/v/v) mobile phase consisting of chloroform, acetic

acid and Milli-Q water (filling height: 1 cm). Chromatography run until the buffer was close to the top of the TLC plate. The running process was repeated three times to separate the different sugars better, and the plate was completely dry between each run. Subsequently, the dried TLC plate was sprayed twice with TLC staining solution (Table 2.48) to detect the sugars in the DESAGA ChromaJet DS 20 (Desaga Sarstedt, Nümbrecht, Germany) and incubated at 120°C for 10 min. The stained TLC plates were scanned. If it is necessary, the brightness and contrast were edited with the program Affinity Photo. Qualitative identification of the hydrolysis products was based on the run length or the retention factor of the products compared with known mono- and oligosaccharide standards.

Table 2.48 TLC staining solution.

Component	Amount
Acetone	100 mL
Aniline	1 mL
Diphenylamine	1.0 g
10 % (v/v) H_3PO_4 (85 %) is added before application.	

2.6.2 HPAEC-PAD

High-Performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was used to chromatographically separate enzymatic hydrolysates of mono- and oligosaccharides and determine the products. For this purpose, a Thermo Fisher Scientific Dionex ICS-3000 system equipped with a CarboPac PA1 column (4 x 250 mm) and a PA1 pre-column (4 x 50 mm) was used for HPAEC-PAD. The mono- and oligosaccharides were deprotonated in the basic running medium and were analyzed by different interactions with the stationary phase of the used Dionex™ CarboPac™ PA1 chromatography column. Before operating, the pre-column and main column was tempered to 30°C. The saccharides were identified by oxidizing the hydroxyl groups on a gold electrode's surface and measuring the electrical current that resulted. For sample preparation, hydrolysates of polysaccharides were first centrifuged at 13,400 rpm for 5 min. Samples were then diluted 1:5 to 1:100 in ddH_2O in a total volume of 500 µL. Separation was achieved in 100 mM sodium hydroxide using an increasing sodium acetate gradient within 45 min (Table 2.49). The program was set as follows: 0-10 min: a linear gradient from 0 to 100 mM sodium acetate, 10-30 min: a linear gradient from 100 mM to 800 mM sodium acetate, 30-45 min: 0 mM sodium acetate. To identify the hydrolysis products, oligosaccharide standards were separated before the first sample and after every ten samples. To check the stability of

the detected signal and the constancy of the retention times, check standards with 50 mg L^{-1} arabinose and 50 mg L^{-1} xylose were measured at regular intervals. The quantitative determination of monosaccharides was carried out with a standard linear curve obtained by serious dilution of the monosaccharide, including arabinose and xylose (200 mg L^{-1}, 100 mg L^{-1}, 50 mg L^{-1}, 25 mg L^{-1}, 12.5 mg L^{-1}). The chromatography data were evaluated with the program Chromeleon 7.2 and GraphPad Prism 6.

Table 2.49 Running buffer for HPAEC-PAD.

Running buffer	A	C
50% (w/v) NaOH (VWR International, Radnor, USA)	100.0 mM	100.0 mM
≥99% Sodium acetate (NaOAc, Sigma-Aldrich, St. Louis, USA)	0.0 mM	1000.0 mM
50% (w/v) NaOH	20.80 mL	10.40 mL
≥99% Sodium acetate	-	164.06 g
Fill water to	4000.0 mL	2000.0 mL

2.6.3 MALDI-TOF mass spectrometry

TLC was combined with the MALDI-TOF to achieve a high-resolution analysis to assign the unidentified spots to defined compounds. For this purpose, the standard was loaded onto the first and last lane of the TLC plate. 3 µL of the sample was loaded next to the two lanes of the standards. The rest of the lanes were repeatedly loaded with the sample mixture to ensure the sufficient unidentified sugar can be isolated from the TLC plate. Then, sugar molecules on the TLC plates were separated in the solvent as previously described. Followed that, the silica gel with standards and 3 µL of the sample was cut off with a scissor for staining. The rest part with the samples for further MALDI-TOF avoided the staining step. After that, the whole plate was reassembled, then the spot of an analyte of interest was scraped from the TLC plate corresponding to the spot position on the stained TLC plate. The isolated powder materials were dissolved in 200 µL of water and centrifuged at 13,500 rpm, 4°C for 5 min. The supernatant was transferred to a new EP tube and stored at -20°C for further MALDI-TOF. For the MALDI sample preparation, 0.5 µL of the matrix solution (DHB and HCCA) was deposited onto each position of the MALDI target plate that will be applied. After it was dried, depositing 0.5 µL of the sample solution, calibration buffer and an identified oligosaccharide as positive control onto each matrix spots. Meanwhile, let one position with the empty matrix as a negative control. The plate was sent for analysis after all the samples were dried.

3 Results

Parts of the Results and the Discussion have been published in a similar form in: Liu Y, Vanderhaeghen S, Feiler W, Angelov A, Baudrexl M, Zverlov V, Liebl W (2021) Characterization of two α-L-arabinofuranosidases from *Acetivibrio mesophilus* and their synergistic effect in degradation of arabinose-containing substrates. *Microorganisms* 9(7): 1467. Permission to use material from the publication was granted by the journal *Microorganisms*.

Further results have been included in a manuscript entitled ''Characterization of metagenome-derived arabinosyl hydrolases from biogas fermenter microbial communities.'' by Liu Y, Angelov A, Feiler W, Baudrexl M, Zverlov V, Liebl W, Vanderhaeghen S, which has been submitted for publication.

3.1. Sequence-based analysis of an isolated strain: N2K1

3.1.1 Biochemical activity of N2K1 and its secreted enzymes

Acetivibrio mesophilus (basonym *Hungateiclostridium mesophilum*) strain N2K1^T is a novel anaerobic, mesophilic and cellulolytic bacterium isolated from a biogas fermenter, as recently described by Rettenmaier *et al*. (2019). According to the physiological characterization data from Rettenmaier *et al*., N2K1 can grow on filter paper and cellobiose. However, most of the tested monosaccharides (0.5% w/v), including arabinose, fructose, glucose, lactose, mannitol, mannose, ribose, sucrose, trehalose, xylose, did not support growth (Rettenmaier *et al.*, 2019). To further explore the hemicellulose-/ pectic substrates degrading ability of N2K1, the arabinose/xylose-configured nutrients, including different concentrations of pentose (Xylose and Arabinose), different types of xylan substrates (OSX, WAX-I, BWX), and pectic substrates (SBA and SBP), were also tested for growth. The results showed that N2K1 can grow on a high xylose concentration (> 1%) but not arabinose. The strain also grew on OSX and WAX-I but not on BWX. Further, it slowly degraded SBP, whereas SBA did not support growth (Table 3.1).

Table 3.1 Growth and substrate utilization characteristics of N2K1.

Carbon sources	N2k1
Filter paper[a]	+
Cellobiose[a]	+
Fructose[a]	-
Glucose[a]	-
Lactose[a]	-
Mannitol[a]	-
Mannose[a]	-
Ribose[a]	-
Starch[a]	-
Sucrose[a]	-
Trehalose[a]	-
0.5% Xylose	-
1% Xylose	+ (slow)
2% Xylose	+ (rapid)
0.5% Arabinose	-
1% Arabinose	-
2% Arabinose	-
0.5% OSX (84% Xylp + Araf, Glu and Gala)	+
0.5% WAX-I (Araf: Xylp= 36:51)	+
0.5% BWX	-
0.5% SBP	+
0.5% SBA	-

The tests were done in GS2-based medium supplemented with mono-/oligo-/polysaccharides. Unless otherwise stated, the concentration of substrates used in the study is 0.5% (w/v). OSX: Oat spelt xylan; WAX-I: Insoluble wheat arabinoxylan; BWX: Beechwood xylan; SBP: sugar beet pulp; SBA: sugar beet arabinan. "+": bacterium can grow, "-": bacterium did not grow.

[a] The data was obtained from Rettenmaier *et al.*, (2019).

The activities of the secreted enzymes in the culture supernatant towards WAX-RS and SBA were also tested. TLC analysis of the hydrolysis products released from WAX-RS by incubating the culture supernatant with this substrate showed that the hydrolysis products mainly consisted of a minimal amount of xylose, various xylo-oligosaccharides, which are typical products formed by endo-xylanases, and arabinose, which was released by arabinofuranosidases (Figure 3.1). Since N2K1 has a weak ability to utilize pentoses, the small amount of xylose released from xylan substrates was presumably insufficient to support the growth. Hence, the hydrolysis products of xylooligosaccharides can serve as the major "carbon source" for N2K1, which can be efficiently internalized by some yet unknown sugar transporters.

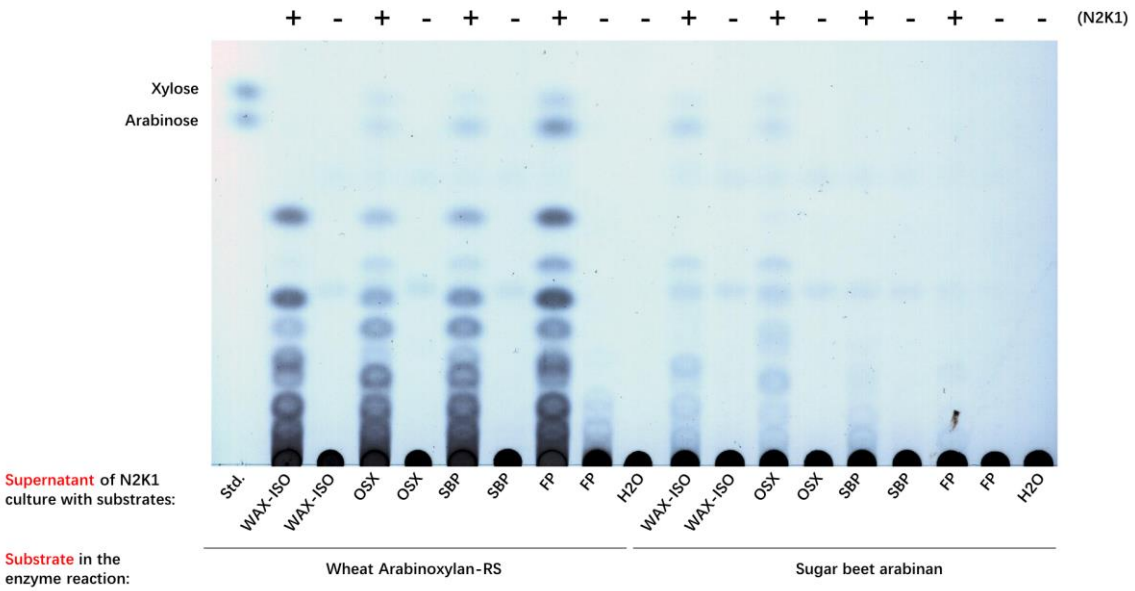


Figure 3.1 TLC analysis of the hydrolysis products of secreted proteins of N2K1 with WAX-RS and SBA as substrates. Secreted proteins were obtained by culturing N2K1 on WAX-I, OSX, SBP, FP, 0. 5 mg L^{-1} secreted proteins were incubated with 0.5% (w/v) substrate at pH 6.5 and 37°C in 25 mM citrate phosphate buffer in a total of 200 µL reaction for 12 h with 600 rpm shaking. The protein concentration was evaluated by comparing it to a standard curve of known bovine serum albumin (BSA) values.

3.1.2 Heterologous production of glycoside hydrolases from N2K1

3.1.2.1 Pectinolytic/arabinolytic enzyme-encoding genes from N2K1

This work aims to characterize the enzymes that hydrolyse the pectic substrates (including arabinan, pectin, sugar beet arabinan) from the newly isolated strain *A. mesophilus*. Among all the putative CAZymes, glycoside hydrolases (GHs), one of the most significant enzyme classed for polysaccharide cleavage, were singled out. Based on Prokka automatic annotation, in total 5 ORFs were predicted to encode arabinofuranosidases, including PROKKA_00452, PROKKA_00933, PROKKA_01572 (RXE59351.1), PROKKA_02400 (RXE58498.1), PROKKA_02412 (RXE58509.1), but no genes were predicted to encode endo-arabinanase. We choose three of them for characterization, specifically the putative GH51 arabinofuranosidase (GenBank: RXE58498.1, designated *Am*Araf51), two putative GH43 arabinofuranosidase (GenBank: RXE59351.1 and GenBank: RXE58509.1, designated *Am*Araf43-1 and *Am*Araf43-2, respectively) and a putative GH43 endo-xylanase (Gene bank: RXE58512.1, designated *Am*Araf43) for characterization. Three of four genes (RXE58498.1, RXE58509.1 and RXE58512.1) are located on a large co-linearly arranged genes cluster (around 26 kbp) comprising of various predicted GHs, including a β-galactosidase (GenBank: RXE58508.1), two arabinofuranosidases

(GenBank: RXE58498.1 and RXE58509.1), four endo-1,4-β-xylanases (consecutive genes with the GenBank number RXE58510.1, RXE58511.1, RXE58512.1, RXE58513.1), which are related to hemicellulose and pectin substrates degradation (Figure 3.2). Besides, the only putative GH28 polygalacturonase encoding gene (locus_tag: EFD62_RS12165, designated *Am*Pgl28) in the genome of N2K1 was also chosen for characterization.

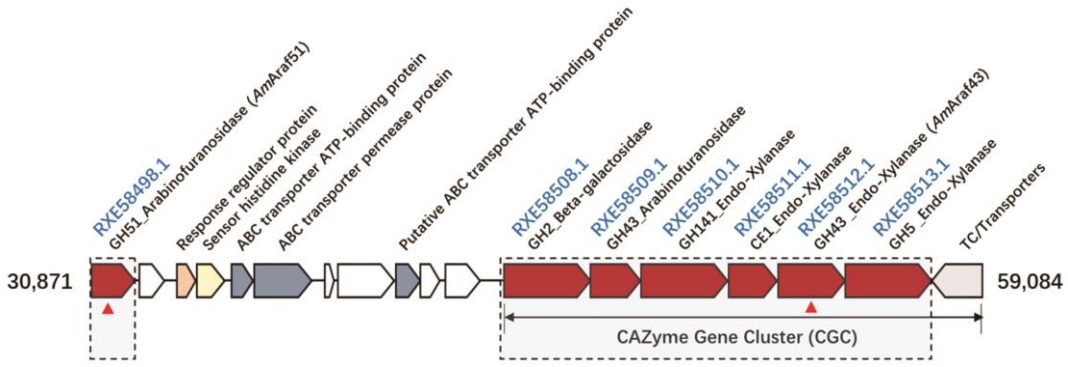
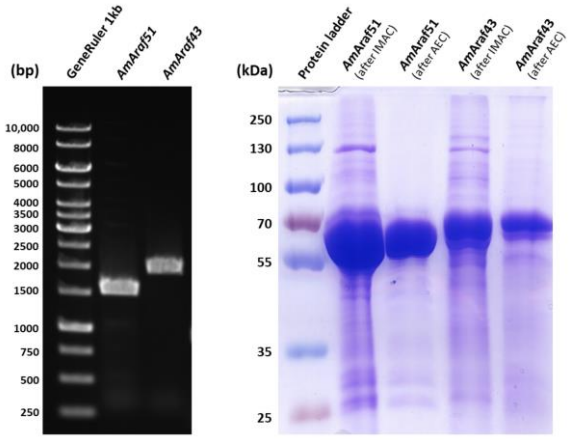


Figure 3.2 Organization of the putative CAZyme gene cluster on the genome of *Acetivibrio mesophilus*. Red triangles indicate the genes that were characterized in this study. (Functional annotation was performed by Prokka genome annotation anddbCAN database). Permission to use this figure from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.2.2 Heterologous production of the recombinant enzymes

To investigate the biochemical characteristics of the enzymes mentioned above, the recombinant proteins *Am*Araf51, *Am*Araf43-1, *Am*Araf43-2, *Am*Araf43, *Am*Araf43$_{truncated}$ (only GH43 domain of *Am*Araf43) and *Am*Pgl28 were heterologously expressed in *E. coli*. Specifically, the ORF encoding *Am*Pgl28 was cloned into pET24c and protein expression was induced in *E. coli* BL21, and the ORFs for *Am*Araf51, *Am*Araf43-1, *Am*Araf43, *Am*Araf43$_{truncated}$ were cloned into the same vector, but protein expression was induced in *E. coli* ArcticExpress, while the ORF for *Am*Araf43-2 was cloned into the pMAL vector and protein expression was induced in *E. coli* ArcticExpress. DNA fragments containing the complete ORFs encoding *Am*Araf51, *Am*Araf43-2 and *Am*Pgl28, respectively or the ORF regions for modules *Am*GH43A, *Am*GH43B and CBM6 of RXE58512.1 (designated *Am*Araf43) or only *Am*GH43A of the same gene (designated *Am*Araf43$_{truncated}$) were amplified with PCR with the primer-based addition of hexa-histidine tags at the N-terminus of *Am*Araf51 and the C-termini of *Am*Pgl28, *Am*Araf43-1, *Am*Araf43 and *Am*Araf43$_{truncated}$. *Am*Araf43-1 on the other hand was expressed as an MBP-fusion protein. The C-terminal dockerin domain of *Am*Araf43 was excluded since such domains usually do not

influence the enzymatic activity and in addition initial experiments had suggested that its presence interferes with the affinity of *Am*Araf43 to bind to the nickel column during immobilized metal ion affinity chromatography (IMAC). The resulting PCR fragments of 1,530 bps for *Am*Araf51, 1,776 bps for *Am*Araf43-1 and 1,654 bps for *Am*Araf43-2, 1,951 bps for *Am*Araf43, 897bps for *Am*Araf43$_{truncated}$, 1,320 bps for *Am*Pgl28 were consistent with the theoretical values (partially shown in Figure 3.3a). The recombinant proteins were purified by IMAC with nickel columns or with amylose resin (for the MBP fusion protein), followed by anion exchange chromatography. SDS-PAGE was used to determine the purity of the proteins, which also demonstrated that their migration was consistent with the predicted molecular masses (57.58 kDa for *Am*Araf51, 101.22 kDa for *Am*Araf43-2, 72.92 kDa for *Am*Araf43, 32.92 kDa for *Am*Araf43$_{truncated}$, 47.84 kDa for *Am*Pgl28) (Figure 3.3b, Figure 6.3a, Figure 6.4a,b), however, *Am*Araf43-1 failed expression. *Am*Araf43-2, *Am*Pgl28, *Am*Araf43$_{truncted}$ were not further studied since the recombinant protein did not show the corresponding enzymatic activity (*Am*Araf43-2, *Am*Araf43$_{truncted}$) or showed very weak activity (*Am*Pgl28). Therefore, further characterization focused on *Am*Araf51 and *Am*Araf43.



(a) (b)

Figure 3.3 Agarose gel electrophoresis and SDS-PAGE analysis of PCR products and purified proteins of *Am*Araf51 and *Am*Araf43, respectively. (a) Q5 PCR amplification of the target genes *Am*Araf51 and *Am*Araf43 (without dockerin domain) using the genomic DNA of N2K1 as template. Lane 1: 3 µL GeneRuler® 1kB DNA Ladder, Lane2: *Am*Araf51 (1,530 bps), Lane 3: *Am*Araf43 (1,951 bps). (b) SDS-PAGE analysis of overexpression and purification of *Am*Araf51 and *Am*Araf43. Lane 1: 3 µL Prestained protein standards; Lane 2: 8.88 µg of *Am*Araf51 (57.6 kDa) was purified by immobilized metal affinity chromatography (IMAC) with a Ni-TED packed column; Lane 3: 2.69 µg of *Am*Araf51 (72.9 kDa) was further purified by anion exchange chromatography (AEC); Lane 4: 10.9 µg of *Am*Araf43 (72.9 kDa) was purified by IMAC with a Ni-TED packed column; Lane 5: 2.74 µg of *Am*Araf43 was further purified by anion exchange chromatography (AEC). Permission to use these figures from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.3 Bioinformatic analysis of *Am*Araf51 and *Am*Araf43

The enzymes *Am*Araf51 and *Am*Araf43 have different molecular architectures (http://pfam.xfam.org). *Am*Araf51 comprises 509 amino acids with only one GH51 catalytic module (residues 297 to 501) (Figure 3.5c). The presence of a signal peptide was not predicted (SignalP 5.0), leading to the presumption of intracellular localization. Phylogenetic analysis of both enzymes showed that *Am*Araf51 shares 94.04% amino acid sequence identity with a putative GH51 arabinofuranosidase from *Acetivibrio thermocellus* DSM 1313 (syn. *Clostridium thermocellum* DSM 1313) (GenBank accession: ADU73229.1) and 93.84% identity with a characterized arabinofuranosidase (*Ct*Araf51A, PDB accession: 2C7F_A) from *A. thermocellus* (syn. *Clostridium thermocellum*) (Figure 3.4), but *Ct*Araf51A and *Am*Araf51 have significantly different activities. *Clocl*_2445 (GenBank accession: AEV69020), another recently identified GH51 arabinofuranosidase from an *Acetivibrio clariflavus* strain (syn. *Clostridium clariflavum, Hungateiclostridium clariflavum*), shared 28.66% identity with the GH51 from the N2K1 strain (Figure 3.4). *Am*Araf43, on the other hand, is a 657-residue enzyme containing an N-terminal signal peptide, a GH43 domain (*Am*GH43A, residues 36 to 291), a module associated with some GH43 modules, GH43_C2 (Pfam designation PF17851, beta-xylosidase C-terminal concanavalin A-like domain) (*Am*GH43B, residues 321 to 504), a CBM6 module (residues 529 to 650), and a C-terminal dockerin domain connected to the rest of the enzyme by a short linker sequence (Figure 3.5d). Two GH43 GHs from *Clostridium* sp. with GenBank accession numbers ABN53398.1 and ABN52503.1 had 52.67% and 52.45% identity with *Am*Araf43, respectively. *Am*Araf43 is presumed to be extracellularly localized, similar to all other putative enzymes encoded in the same gene cluster (Figure 3.4).

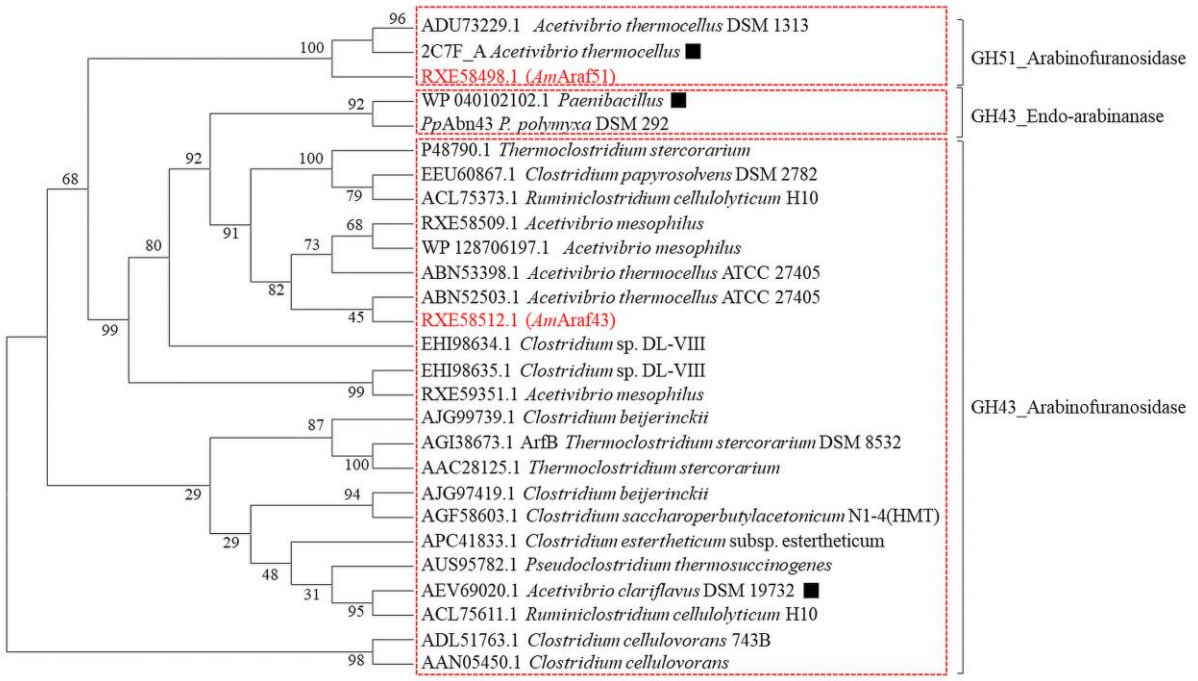


Figure 3.4 Phylogenetic tree of *Am*Araf51, *Am*Araf43 and representative arabinosyl hydrolases from families GH43 and GH51, including an endo-arabinanase *Pp*Abn43 from *Paenibacillus polymyxa* used in this study as well as two other putative GH43 arabinofuranosidases (encoded by RXE59509.1 and RXE59351.1) from *A. mesophilus* strain N2K1. The amino acid sequences of each enzyme are aligned using the MUSCLE algorithm. The tree was then built by the Neighbor-Joining (NJ) method in MEGAX and tested with 500 replicate bootstraps. Bootstrap values were listed near each branching point. The biochemical characterized arabinosyl hydrolases were labeled with a black square symbol (■). Permission to use this figure from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.4 Enzymatic properties of *Am*Araf51 and *Am*Araf43

Using *p*NP-AF as a substrate, the temperature and pH profiles of *Am*Ara51 and *Am*Ara43 activity were studied to optimize the reaction conditions. Temperatures of 70°C for *Am*Araf51 and 57°C for *Am*Araf43 were found to lead to the highest activity under the specified assay conditions (20 min enzyme reactions, pH 6.0 for both *Am*Araf51 and *Am*Araf43) (Figure 3.5a,b). *Am*Ara51 was more than 60% active at temperatures between 44°C and 75°C (Figure 3.5a), whereas *Am*Ara43 displayed over 60% activity in the temperature range from 38°C to 61°C (Figure 3.5b). *Am*Araf51 preferred a pH range of 5.0-6.5, while the optimal pH for *Am*Araf43 was between 4.0-5.0 (Figure 3.5a, b).

The resistance of enzymes to thermoinactivation is a significant factor in their application in biotechnological processes. *Am*Araf51 and *Am*Araf43 were incubated at their optimum pH and different temperatures centered around the optimal temperature for up to 48 hours to assess this characteristic, and residual activity was measured using the *p*NP assay under standard reaction

conditions. When the enzymes were compared after a total of 48 hours of incubation, *Am*Araf43 activity was decreased to 82.9 % of its original activity at 40°C and dropped to 15.8 % at 50°C. Above 60°C, the residual activity after the same incubation period was less than 5% of the original activity (Figure 3.5d). *Am*Araf51, on the other hand, was more stable at higher temperatures. It remained active after two days of incubation at 50°C. After two days at 60°C or 65°C, *Am*Araf51 kept around 80% or 60% of its original activity, respectively (Figure 3.5c).

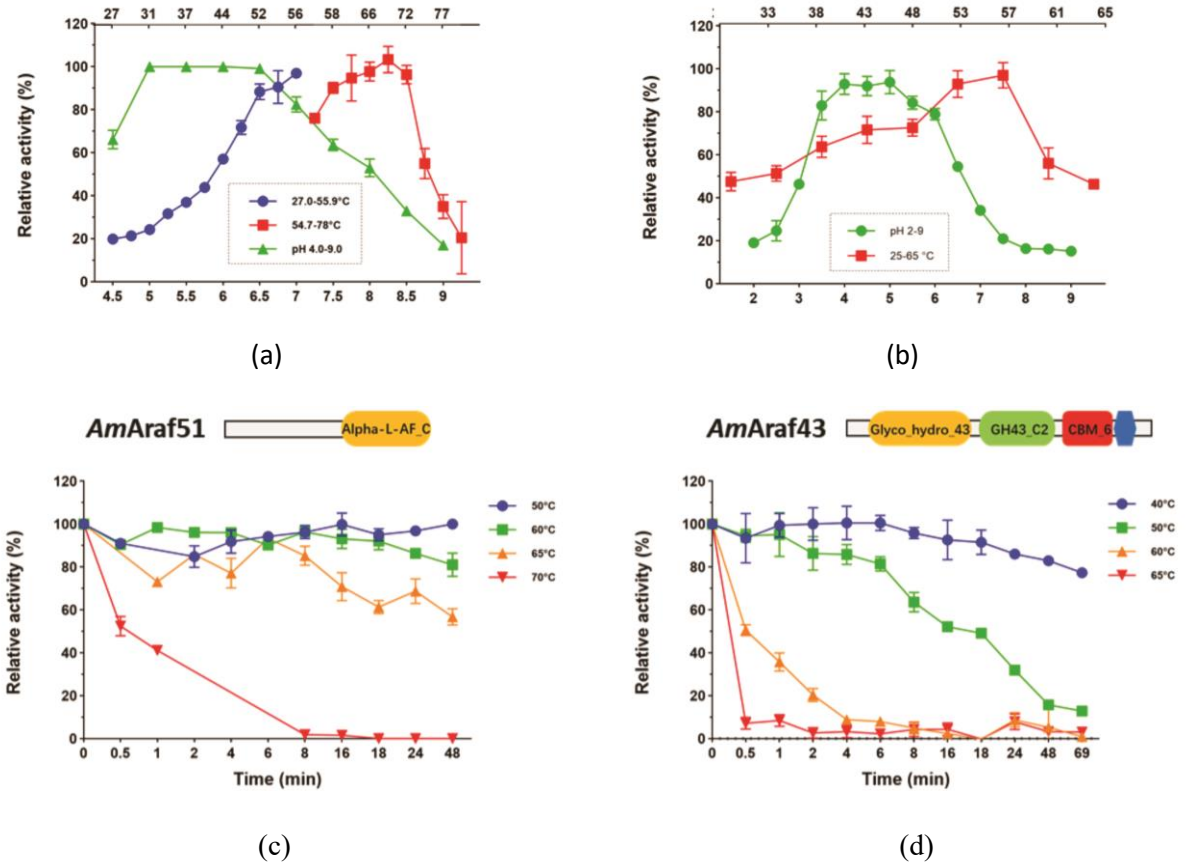


Figure 3.5 Determination of the dependence of activity on temperature and pH optimum, and of the resistance to thermoinactivation of *Am*Araf51 and *Am*Araf43. (a-b). Dependence of activity of *Am*Araf51 and *Am*Araf43 on temperature and pH; (c-d). Thermostability of *Am*Araf51 and *Am*Araf43, respectively. Relative activities were determined by *p*NP standard assay (2 mM *p*NP-AF was incubated with 6.2 nM *Am*Araf51 or 2.4 µM *Am*Araf43 in a 25 mM pH 5.5 citrate phosphate buffer for 20 min, with 600 rpm shaking). The maximal activity was set as 100%. Error bars represent standard deviations (n=3). Permission to use these figure from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.5 Specific activities of *Am*Araf51 and *Am*Araf43 on polysaccharides, and kinetic parameters

Cleavage specificity of *Am*Araf51 and *Am*Araf43 was assessed using various arabinose-containing

substrates with different compositions and structures. Both arabinofuranosidases demonstrated activity against *p*NP-AF, an artificial substrate for exo-α-L-arabinofuranosidase activity, indicating that they are exo-α-L-arabinofuranosidases. *Am*Araf51 displayed the highest specific activity against *p*NP-AF (1,835.79 U mg^{-1}). In comparison, *Am*Araf43 can also degrade *p*NP-AF, with a specific activity of 0.41 U mg^{-1}, which is over 4,400 times lower than *Am*Araf51. *Am*Araf51 also showed enzymatic activity against arabinan, with 4.3 U mg^{-1} activity on linear arabinan (LA), 1.97 U mg^{-1} activity on debranched arabinan (DA) and 0.50 U mg^{-1} activity on sugar beet arabinan (SBA) (Table 3.2). Only arabinose was liberated as a result of the hydrolysis of arabinan substrates (Figure 6.16a). These findings suggest that *Am*Araf51 is an exo-α-L-arabinofuranosidase. Additionally, even though the fact that both soluble wheat arabinoxylan (WAX-RS) and insoluble wheat arabinoxylan (WAX-I) used in our study are substituted with arabinofuranosyl residues (Araf/Xylp=36:51 in WAX-I and Araf/Xylp=38:62 in WAX-RS), *Am*Araf51 activity on both xylan substrates was hardly detectable. By contrast, *Am*Araf43 activity on arabinan substrates was scarcely detectable, while the enzyme had significant activity on arabinoxylan, particularly WAX-RS (0.56 U mg^{-1}) (Table 3.2). Within 24 hours of incubation of WAX-RS with *Am*Araf43 at the enzyme's optimum reaction conditions, the product released during hydrolysis of these substrates was solely arabinose (Figure 6.16b). Neither xylose nor xylo-oligosaccharides were found. As a result, *Am*Araf43 is likewise an exo-α-L-arabinofuranosidase rather than a xylanase, as the initial gene annotation suggested.

Based on the rate of arabinose equivalents released from the substrate polymers, the kinetic constants of *Am*Araf51 for SBA and DA hydrolysis and *Am*Ara43 for WAX-RS hydrolysis were determined. The Michaelis-Menten model was fit to all data using nonlinear regression. Our results showed *Am*Araf51 has a higher K_m and V_{max} on DA (K_m=38.54 g L^{-1}, V_{max}=7.11 nmol min^{-1} μg^{-1}) than on SBA (K_m=25.69 g L^{-1}, V_{max}=2.84 nmol min^{-1} μg^{-1}), indicating a slightly higher binding affinity of *Am*Araf51 towards SBA than DA. *Am*Araf43 only showed an affinity with WAX, with a similar K_m value (26.56 g L^{-1}) for WAX-RS as *Am*Araf51 for SBA (K_m=25.69 g L^{-1}) (Table 3.2).

Table 3.2 Specific activities and kinetic parameters of *Am*Araf51 and *Am*Araf43 on different substrates.

	Specific Activity: U/mg							
	LA	**DA**	**SBA**	**WAX-I**	**WAX-RS**	**BWX**	**SBP**	***p*NP-AF**
*Am*Araf51	4.3 ± 0.24	1.97 ± 0.15 K_m = 38.54 V_{max} = 7.11	0.50 ± 0.05 K_m = 25.69 V_{max} = 2.84	NA	NA	-	NA	1,835.79 ± 15.02
*Am*Araf43	-	-	NA	0.03 ± 0.01	0.56 ± 0.04 K_m = 26.56 V_{max} = 1.64	-	-	0.41 ± 0.03

Enzyme activity was determined by DNS assay on 0.5% of different substrates. One unit of activity was defined as the amount of enzyme to release 1 µmol of L-arabinose equivalent per minute under the optimal temperature and pH. SBA: Sugar beet arabinan; LA: Linear arabinan; DA: Debranched arabinan; WAX-RS: Wheat arabinoxylan for reducing sugar assay; WAX-I: Insoluble wheat arabinoxylan; BWX: Beechwood xylan; SBP: Sugar beet pulp; *p*NP-AF: 4-Nitrophenyl α-L-arabinofuranoside; NA: not available. ''-'': no activity detected; U mg^{-1}=µmol mg^{-1} min^{-1}; K_m=g L^{-1}; V_{max}=nmol min^{-1} μg^{-1}. Permission to use this table from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.6 Hydrolysis patterns of *Am*Araf51 and *Am*Araf43 on AOS

*Am*Araf51 and *Am*Araf43 released arabinose from SBA and DA, or WAX, all including arabinofuranosyl side-chain substitutions linked to O-5, O-3 and/or O-2 positions of arabinose residues or O-3 and/or O-2 positions of xylose residues, respectively. The activity of both enzymes on arabino-, arabinoxylo-, and xylo-oligosaccharides (AOS, AXOS, and XOS, respectively) with varied branching patterns was further investigated by TLC or HPAEC-PAD to obtain additional insight into their cleavage specificities. The results showed that within 1 to 5 hours incubation with three different AOS, *Am*Araf51 could almost completely hydrolyze these oligosaccharides to arabinose (Figure 3.6a,b). Comparing the structures of these two substrates, O-A4B has a trisaccharide backbone with 1,3-linked arabinofuranosyl side chains connected to the second arabinose of the 1,5-linked arabinan backbone (counting from the non-reducing end). By contrast, the backbone structures of the O-A5B mixture include arabinotriose and arabinotetraose, with the former having two arabinofuranosyl substitutions at O-3 and O-2 positions of the second main chain arabinose and the latter having an arabinofuranosyl branch at position O-3 of the third main chain arabinose (Figure 3.6d). As a consequence of the studies using AOS as substrates, *Am*Araf51 has a broad cleavage specificity for arabinofuranosyl side-chain types. This enzyme can cleave 1,5-, 1,3-, and 1,2-linkages. The time required for total hydrolysis of the three AOS (O-A5BMIX > O-A4B > A3) also revealed that the complexity and the degree of polymerization (DP) of side-chain substitutions are negatively correlated with this enzyme's cleavage efficiency, implying that terminal α-1,5-linked main-chain backbones are preferred over glycosidic bonds linking arabinosyl side groups at

the O-3 and O-2 positions (Figure 3.6a,b). Incubation of *Am*Araf43 with α-1,5-linked arabinotriose, on the other hand, did not result in the release of arabinose. In contrast, the minor amount of arabinose liberated from O-A4B and O-A5BMIX suggests that *Am*Araf43 has weak activities either toward 1,3-/1,2-linkages or activities toward both side-chain substitutions. However, *Am*Araf43 can hardly cleave α-1,5-linkages in A3 (Figure 3.6c).

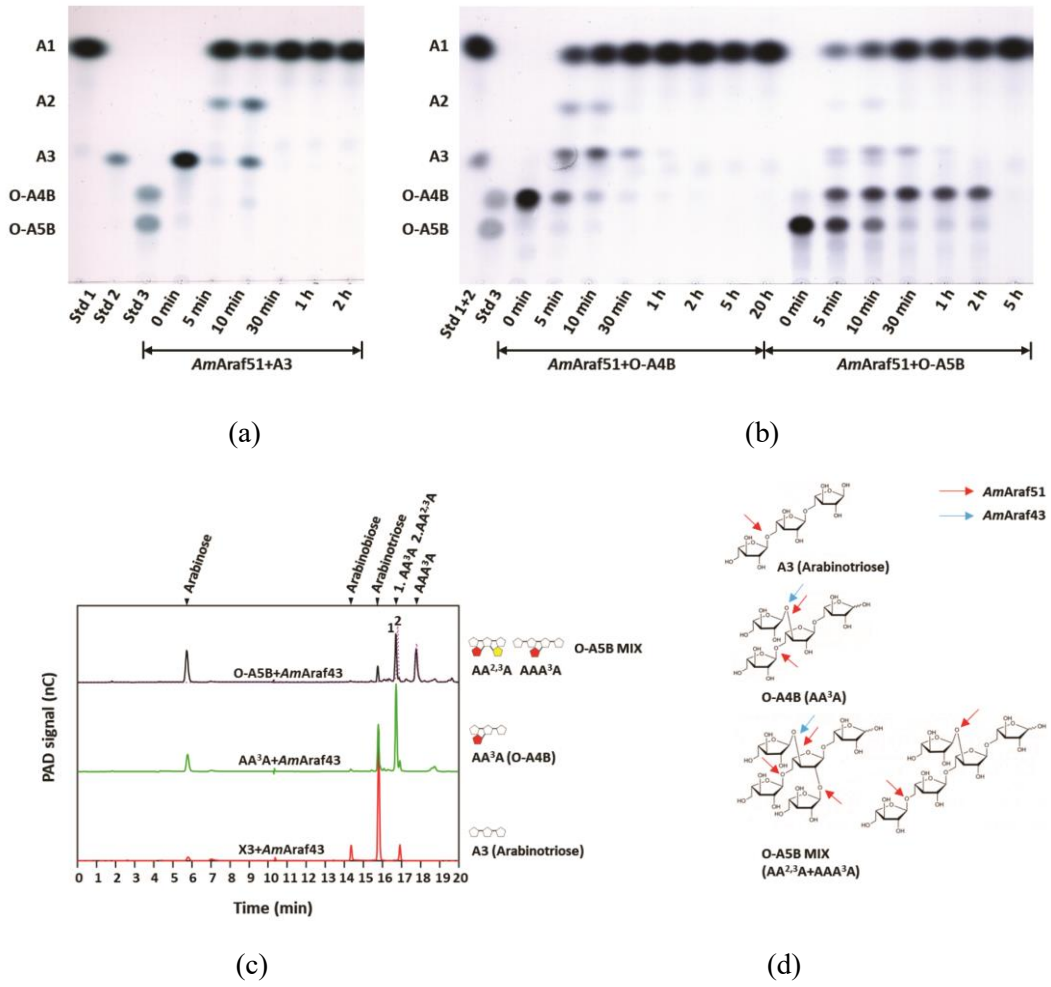


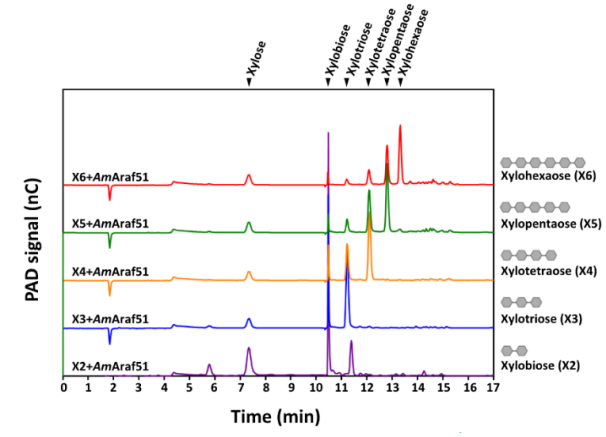
Figure 3.6 TLC/HPAEC-PAD analysis of the hydrolysis products released from AXOs. TLC time-course profile of *Am*Araf51 hydrolysis products using (a) arabinotriose, (b) O-A4B or O-A5B as substrates. (c) HPAEC-PAD identification of *Am*Araf43 hydrolysis products with AOS as substrates after 22 hours of incubation (AA^3A and $AA^{2,3}A$ are quite close, AA^3A matches the left of the two peaks at a retention time of around 17 min). (d) The structure of A3 (arabinotriose), AA^3A (O-A4B), $AA^{2,3}A$ and AAA^3A (O-A5B MIX). Standard reactions were carried out by incubating 85 nM *Am*Araf51 and 82 nM *Am*Araf43 with various AXOs at pH 6.0 and 60°C for *Am*Araf51 and at pH 5.0 and 40°C for *Am*Araf43 for varying time duration. 2.5 µL of the hydrolysis products were separated three times on TLC plates using a solvent mixture of chloroform, acetate and water in a ratio of 6:7:1 (v/v/v). Permission to use these figures from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.7 Action mode of *Am*Araf51 and *Am*Araf43 on XOS and AXOS

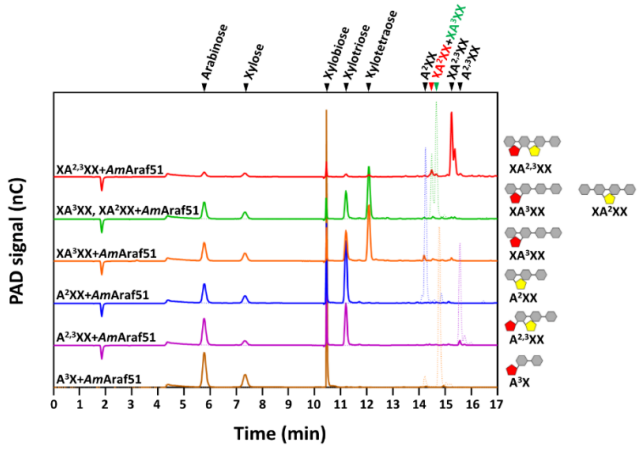
The activity of *Am*Araf51 and *Am*Araf43 towards linear XOS and arabinosyl decorated XOS (AXOS) were also studied. The hydrolysis products were determined by HPAEC-PAD by comparing the hydrolysates' retention times to those of the oligosaccharide standards. For *Am*Araf51, beside the arabinofuranosidase activity, the xylosidase side-activity was also confirmed using linear XOS, which resulted in cleavage of linear XOS from DP2 (degree of polymerization of 2) to DP6 and the release of XOS with a lower DP and of xylose (Figure 3.7a). *Am*Araf51 was also able to liberate arabinose from O-2 and O-3 positions of arabinose when incubated with AXOS, and a β-xylosidase side-activity was evident from the liberation of xylose following a 24 h incubation with AXOS (Figure 3.7b). Presumably, the arabinofuranosidase activity of *Am*Araf51 can debranch the AXOS first, followed by the xylosidase activity of *Am*Araf51, which cleaves the debranched XOS intermediates further but not completely (Figure 3.7b). For *Am*Araf43, by contrast, no xylosidase side-activity towards either the linear or branched AXOS was observed.

Specifically, *Am*Araf51 was able to cleave the arabinosylations off of singly modified AXOS, including XA^3XX, XA^2XX, A^2XX and A^3X, resulting in the linear XOS xylotretraose, xylotretraose, xylotriose, and xylobiose, respectively. As previously noted, its xylosidase activity can cleave off terminal xylose residues. After 24 hours of incubation, *Am*Araf51 was able to fully cleave both arabinose moieties (1,2- and 1,3-linked) off the terminally twofold-arabinosylated $A^{2,3}XX$ without producing mono-substituted A^3XX or A^2XX. However, in the hydrolysis product of $XA^{2,3}XX$, which includes an internal double arabinose-substituted xylose, only a tiny amount of arabinose and xylose can be identified in the hydrolysis product (Figure 3.7b).

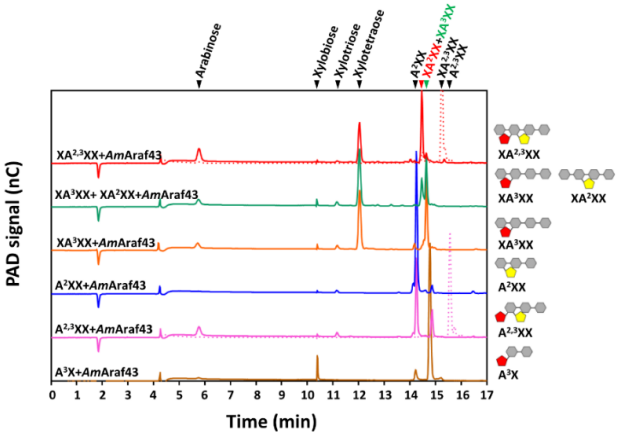
*Am*Araf43, on the other hand, was able to cleave off the 1,3-linked arabinofuranosyl residues from either terminally or internally di-arabinosylated $A^{2,3}XX$ and $XA^{2,3}XX$, resulting in the hydrolysis products A^2XX and XA^2XX, but it is unable to cleave the 1,2-linked arabinose side chain in these products. Furthermore, *Am*Araf43 can slightly remove the single 1,3-linked arabinosyl residues from terminally or internally substituted A^3X and XA^3XX, resulting in just a small portion of these oligosaccharides being degraded to X2 and X4, respectively, in the corresponding assays (Figure 3.7c).



(a)



(b)



(c)

Figure 3.7 HPAEC-PAD analysis of the hydrolysis products released from XOS and AXOS by *Am*Araf51 and *Am*Araf43. (a) Linear XOS degraded by *Am*Araf51, (b) AXOS degraded by *Am*Araf51, and (c) AXOS degraded by *Am*Araf43. Standard reactions were carried out by incubating 85 nM *Am*Araf51 or 82 nM *Am*Araf43 with 0.25 % of various XOS at 60°C and pH 6.0 for *Am*Araf51 and 40°C and pH 5.0 for *Am*Araf43 for 24 h. Permission to use these figures from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.8 Influence of metal ions, denaturants and metal chelator on the activity of *Am*Araf51 and *Am*Araf43

Under standard *p*NP assay conditions, the impact of different metal ions such as Cu^{2+}, Ni^{2+}, Fe^{2+}, Zn^{2+}, Co^{2+}, Mn^{2+}, Ca^{2+}, Na^{+}, K^{+}, Mg^{2+}, the protein denaturants SDS and urea, and the metal chelator EDTA was determined at concentrations of 10 mM, 5 mM and 1 mM. With 10 mM Ca^{2+}, 5 mM/10 mM Na^{+}, 1 mM Mn^{2+}, and 5 mM K^{+}, as well as 5-10 mM urea or EDTA, slight activity increases of around 30% to 60% were found for *Am*Araf51 (Figure 3.8a). Cu^{2+}, Ni^{2+}, and Fe^{2+} at all doses tested, Zn^{2+} at 5 mM or higher, and Co^{2+} or Mn^{2+} at 10 mM all substantially reduced *Am*Araf51 activity (Figure 3.8a). In comparison, Mn^{2+} at 5 and 10 mM, in particular, exhibited a 40-60% stimulating impact on *Am*Araf43 (Figure 3.8b). Furthermore, both *Am*Araf51 and *Am*Araf43 showed a 530 mM arabinose tolerance (Figure 3.8c).

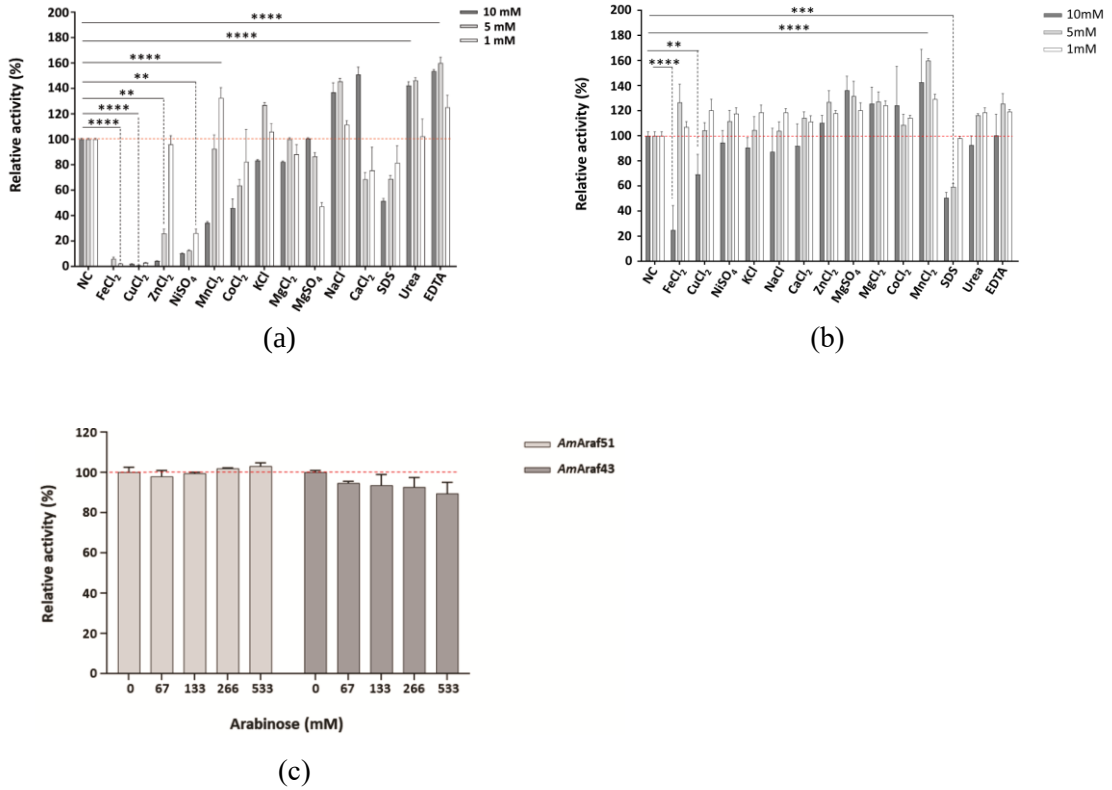


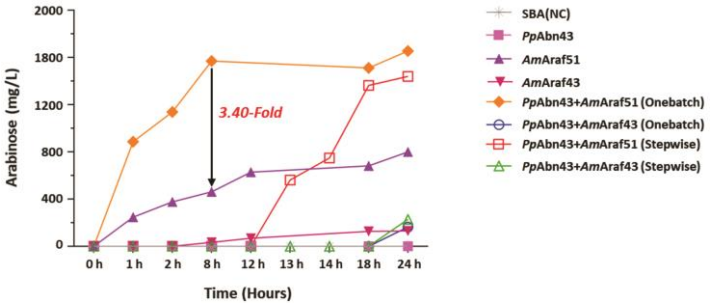
Figure 3.8 Influence of metal ions, protein denaturants, metal chelator and arabinose on the activity of *Am*Araf43 and *Am*Araf51. Standard *p*NP-Assays were carried out by incubating 0.56 µM *Am*Araf51 and 0.16 µM *Am*Araf43 (after 10 mM, 5mM, 1 mM metal ions/protein denaturants/metal chelator, or 67 mM, 133 mM, 266 mM, 533 mM arabinose treatment at RT for 2 h) with 0.2 mM *p*NP for 20 min (at pH 6.0, 60°C and pH 4.5, 40°C in 25 mM citrate phosphate buffer for *Am*Araf51 and *Am*Araf43, respectively). Relative activity was calculated based on the ratio between the activity of the chemical reagent treated sample and the activity of the negative control (without treatment). The standard deviation of triplicates is represented by error bars. Permission to use these figures from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.1.9 Synergistic action between *Am*Araf51, *Am*Araf43 and *Pp*Abn43, M_Xyn10

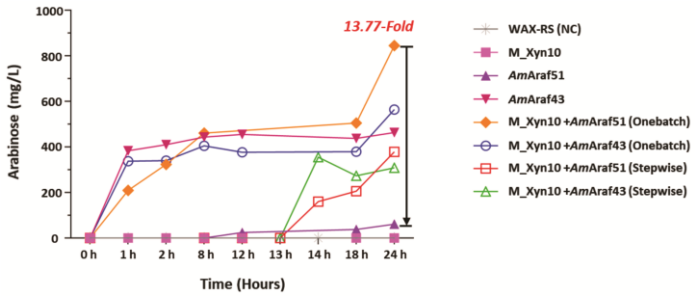
*Am*Araf51 was unable to cleave the arabinofuranosyl side chains from WAX, while it was able to hydrolyze the arabinose homopolymers SBA and DA efficiently. In addition, *Am*Araf51 exhibited β -xylosidase side-activity in addition to its main activity as an exo-arabinofuranosidase. *Am*Araf43, on the other hand, exhibited merely weak activity on arabinan substrates but high activity on arabinoxylan, particularly WAX-RS. It was speculated that these two arabinofuranosidases (*Am*Araf51 and *Am*Araf43) could be combined with endo-cleaving GHs, i.e., endo-arabinanase and endo-xylanase, for the synergistic deconstruction of arabinan (SBA) and xylan (WAX), respectively, and thus could be of use as components of defined enzyme cocktails for polysaccharide monomerization. This was investigated using two enzymes from our laboratory: the endo-arabinanase *Pp*Abn43 from *Paenibacillus polymyxa* and the endo-xylanase M_Xyn10 from a metagenomic library screening. *Pp*Abn43 hydrolyzed SBA to release mostly AOS but little arabinose (Figure 3.9a, Figure 6.17a), whereas M_Xyn10 degraded WAX-RS to produce XOS and AXOS while releasing some xylose (Figure 3.9c, Figure 6.17b). When a cocktail of *Pp*Abn43 and *Am*Araf51 was added to SBA, arabinose production rose significantly compared to each enzyme acting alone, resulting in a 2.1-fold higher arabinose concentration compared to single enzyme *Am*Araf51 treatment (Figure 3.9a). Stepwise treatment, first incubating SBA with *Pp*Abn43 for 12 h (which released no detectable monomeric arabinose), followed by addition of *Am*Araf51 for another 12 h incubation, which yielded slightly less arabinose than the simultaneous incubation of SBA with both enzymes for 24 h (Figure 3.9a). Since *Am*Araf43 did not display strong activity against SBA or AOS (Table 3.2, Figure 3.6a), only a tiny amount of arabinose was released from SBA with its assistance in combination with *Pp*Abn43, neither during one batch treatment (incubation with *Pp*Abn43 and *Am*Araf43) nor via stepwise treatment (addition of *Am*Araf43 after pretreatment with *Pp*Abn43) (Figure 3.9a).

Using wheat arabinoxylan (WAX-RS) as the substrate, the synergistic effects of the arabinofuranosidases *Am*Araf51 and *Am*Araf43 in concert with endo-xylanase M_Xyn10 were investigated. We measured both arabinose and xylose's yields in this case. When WAX-RS was incubated with M_Xyn10 and *Am*Araf51 for 24 h, the production of arabinose was 13.77 times higher (845.1 mg L^{-1}) than when *Am*Araf51 was set alone (61.4 mg L^{-1}) (Figure 3.9b). Stepwise treatment with the same enzyme combination, i.e., incubation of WAX-RS with M_Xyn10 first, then *Am*Araf51, yielded 378.6 mg L^{-1} arabinose (Figure 3.9b). By contrast, *Am*Araf43, without the addition of the endo-xylanase

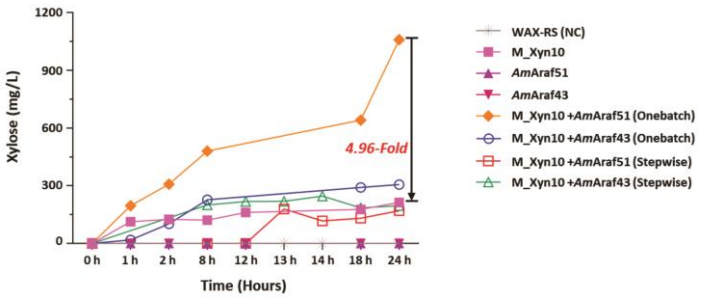
M_Xyn10, produced 463.6 mg L^{-1} arabinose after 24 h, which was much higher than *Am*Araf51 on WAX-RS (Figure 3.9b). When incubating *Am*Araf43 with M_Xyn10 simultaneously with WAX-RS as substrate, a kind of higher arabinose (564.5 mg L^{-1}) was detected (Figure 3.9b). The stepwise treatment of WAX-RS by first adding M_Xyn10, followed by *Am*Araf43, the yield of arabinose (308.5.0 mg L^{-1}) was significantly lower than simultaneous treatment (Figure 3.9b). In terms of xylose production, the combined and simultaneous activity of M_Xyn10 and *Am*Araf51 produced 1,059.0 mg L^{-1} of xylose, whereas M_Xyn10 alone produced nearly five times less xylose (213.5 mg L^{-1}) (Figure 3.9c).



(a)



(b)



(c)

Figure 3.9 HPAEC-PAD quantification of arabinose and xylose released from SBA or WAX-RS. (a) Yield of arabinose with SBA as substrate, (b) yield of arabinose with WAX-RS as substrate, (c) yield of xylose with WAX-RS as substrate. Standard reactions were performed by incubating 0.5% (w/v) SBA or WAX-RS with the different combinations of endo- and exo-active enzymes (endo-active enzymes include 0.90 µM *Pp*Abn43 or 2.9 µM M_Xyn10, exo-active enzymes include 0.3 µM *Am*Araf51 or *Am*Araf43) at 50°C and pH 5.5 over time. Permission to use these figures from the publication was granted by the journal *Microorganisms* (Liu *et al.*, 2021).

3.2. Function based screenings

3.2.1 Production of CPH/ICB substrates

Due to the limited efficiency of conventional metagenomic library screening methods, a high-throughput assay system was developed, which was meant to be faster and more reliable than traditional screening methods. The polysaccharide chains in chromogenic substrates (including CPH and ICB substrates) are covalently linked with chromophores (chlorotriazine dyes), and they are also covalently crosslinked between polymer chains with a bi-functional epoxyl-based cross-linker (1,4-butanediol diglycidyl ether), which can make the substrate insoluble (Figure 3.10). After enzymatic digestion, the insoluble matrix disintegrates into smaller fragments that diffuse into the solution changing the color of the supernatant (Figure 3.10). The CPH substrates were produced in various colors based on the color of chlorotriazine dyes, such as green, red and blue, allowing for the development of a high-throughput assay method by mixing the several color-labeled substrates in a single well of the reaction plate. Multiple enzymes from the same fosmid clone and a single enzyme with side activity can be identified in this way. Usually, combining red with green or blue produces an instantly recognizable hue. The applicability of labeled substrates was verified with commercial enzymes acquired from Megazyme or enzymes available in our laboratory before they were applied to metagenomic library screening. The results indicated that labeled xylan, arabinoxylan, arabinan and CMC functioned effectively, but that labeled pectin, sugar beet pulp and avicel released only negligible quantities of color into the supernatant, even using commercial enzymes (Figure 3.11a,b).

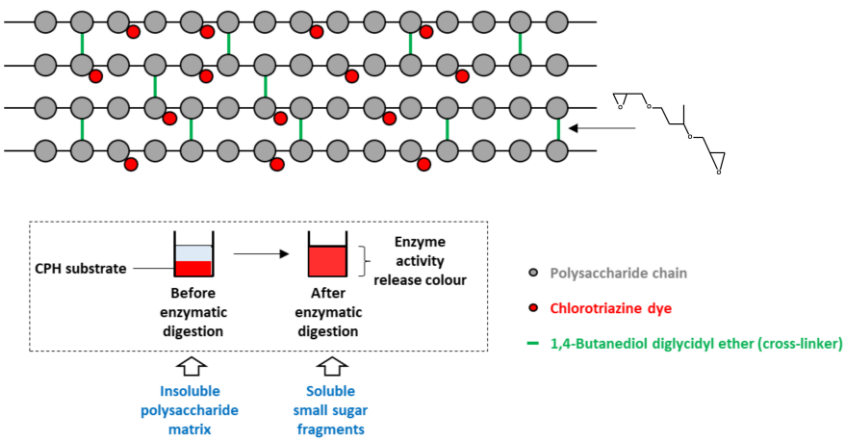


Figure 3.10 The scheme of the CPH and ICB substrate insoluble matrix and its enzymatic digestion (CPH: Chromogenic polysaccharide hydrogel, ICB: Insoluble chromogenic biomass). Figure was based on Schückel *et al*. (2018).

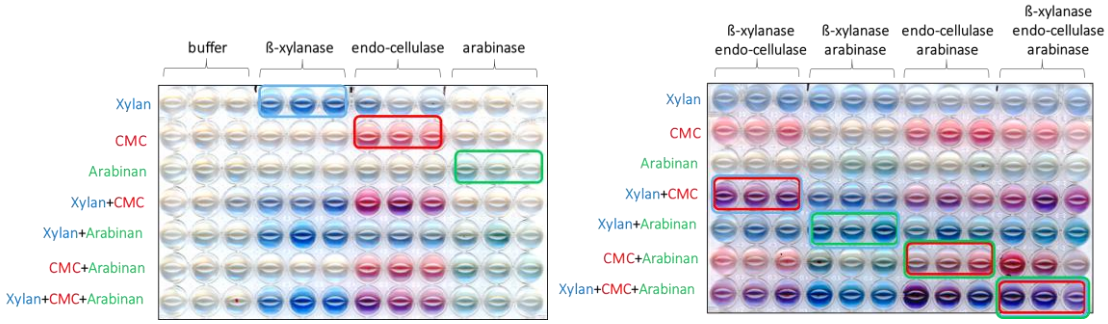


Figure 3.11 Enzyme reaction for checking the applicability of the CPH substrates. Part of reaction plates containing a chromogenic polysaccharide hydrogel (CPH) substrate or simultaneous use of two/three different color labeled CPH substrates incubated with commercial enzymes from Megazyme or own enzymes available in the laboratory. CPH substrates were distributed into each well by syringe. The enzyme mixture used in the reaction included: 1 µL β-endo-xylanase (Megazyme), 1 µL endo-cellulase (Megazyme), 9.25 pmol arabinanase *Pp*Abn43. Negative control: CPH substrates were incubated with buffer instead of enzymes.

3.2.2 Metagenomic library construction and screening

To identify microbial consortia capable of degrading complex lignocellulosic biomass, such as SBP, which includes a high pectin content (24-32%). Eleven different fermenter samples (Table 2.34) were obtained as kind gifts from Dr. Vladimir Zverlov and were further inoculated into a GS2-based medium with SBP as a carbon source. Only four of these, designated "T1T2," "MOD18," "MOD17" and "S4MN4", shown a significant capacity to entirely and swiftly deconstruct SBP. According to 16S rRNA amplicon sequencing, the major phyla in the SBP-enriched fermenter samples "T1T2" and "MOD18" included *Firmicutes* (82.12% in T1T2 and 88.66% in MOD18), *Thermotogae* (17.22% in T1T2 and 6.28% in MOD18) and *Bacteroidetes* (0.03% in T1T2 and 4.96% in MOD18) (Table 3.3).

Table 3.3 16S rRNA amplicon sequencing showed the composition of microbial communities from biogas fermenter "MOD18" and "T1T2".

	MOD18 library		T1T2 library	
Phylum	**OUT of V3**	**Percentage**	**OUT of V3**	**Percentage**
Firmicutes	16,756	88.66	20,185	82.12
Thermotogae	1,187	6.28	4,233	17.22
Bacteroidetes	938	4.96	8	0.03
Proteobacteria	0	0	1	0
Others	18	0.1	152	0.62
Sum	18,899	1	24,579	1

This result has been included in the manuscript by Liu *et al.*, (2022).

Regarding the strong SBP saccharification ability, these bacterial communities were likely to contain interesting genes involved in pectin-enriched plant biomass degradation. Thus, the microbial community from SBP enriched thermophilic biogas fermenter sample "T1T2" and "MOD18" were chosen for additional metagenomic library construction, using the CopyControl™ Fosmid Library Production Kit and pCC1 fosmid as the backbone, as described in Materials and methods. Finally, 20,000 clones in the "MOD18" library and 5,000 clones in the "T1T2" library were generated and subjected to function-based activity screening to identify enzyme activities able to cleave xylan, arabinoxylan, arabinan, CMC and pectin. In addition to the newly constructed metagenomic libraries, 7,776 existing fosmid clones, produced from metagenomic DNA from various environment samples in either pCC1 Fosmid or pCT3FK Fosmid, were also applied to the same screening process (Table 3.4, Table 6.2). This effort yielded 62 positive fosmid clones, including 39 positive fosmid clones with CMCase activity, 6 with xylanase activity, 6 with arabinanase activity, 3 with arabinoxylanase activity, and 8 with dual enzyme activity (Figure 3.12). To eliminate the presence of potentially identical clones from the functional screens, the fosmid clones were digested with BamH I, and the digested recombinant fosmids revealing the same patterns on an agarose gel were assumed to be identical clones, of which only one was retained (Figure 6.11). After deleting seven identical positive clones with CMCase activities, 55 positive fosmid clones were chosen for further sequencing, including 40 positive fosmid clones from the pCC1 Fosmid library and 15 positive fosmid clones from the pCT3FK Fosmid library (Table 3.4).

Table 3.4 Construction and functional screening of metagenomic libraries to find lignocellulose-degrading enzymes.

Name of Libraries	Fosmid Backbone	Origin	Number of Clones	Size of Inserts	Positive Clones for Sequencing
Existing library	pCT3FK	"ED library" (Elephant feces; gift from W. Streit, Hamburg)	1,372	~40 Kb	3 (2 CMCases [a], 1 xylanase)
Existing library	pCT3FK	"BM1T library" from A. Pechtl, TUM)	980	~40 Kb	15 (3 arabinoxylanases, 9 dual activities [b], 3 xylanases)
MOD18 library	pCC1FOS	Thermophilic digestate, fermenter, 50°C	20,000	~40 Kb	18 (4 arabinanases, 14 CMCases)
T1T2 library	pCC1FOS	Thermophilic digestate, fermenter, 55°C	5,000	~40 Kb	19 (2 arabinanases, 2 xylanases, 15 CMCases)
Others	pCC1/pCT 3FK	various	5,424	~40 Kb	0
Total			32,776		55

[a] CMCases: carboxymethyl cellulases.

[b] Dual activities: both xylanase and arabinoxylanase activities.

Metagenomic libraries were screened by using chromogenic polysaccharide hydrogel (CPH) substrates and insoluble chromogenic biomass (ICB) substrates.

Detailed information about the fermenter samples used for metagenomic library construction as well as existing metagenomic libraries can be found in Table 2.34 and Table 6.2, respectively.

This result has been included in the manuscript by Liu *et al.*, (2022).

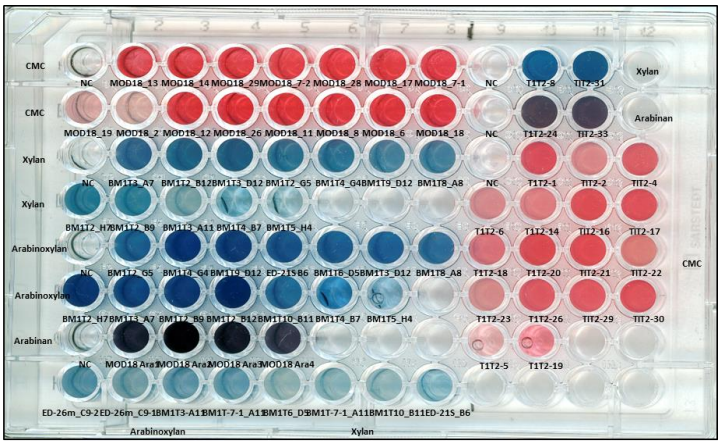


Figure 3.12 Functional screening of fosmid libraries for the four major hydrolases activities. Including CMCase (Red), xylanase (Blue), arabinoxylanase (Blue) and arabinanase (Blue). Positive clones were screened by using CPH (chromogenic polysaccharide hydrogel) substrates. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.3 Positive fosmid clone sequencing

To unveil the genes encoding functions of interests, a "pooled sequencing" strategy, which combines Illumina and Sanger sequencing, was used to gather sequence information from the fosmids giving rise to GH activity.

With this method, Sanger end sequencing was used to sequence the termini of the fosmid inserts, resulting in around 1,000 bps sequences. Illumina shotgun sequencing after assembly generated contigs which were assigned to different fosmid clones via the Sanger "end-tags." In our study, four of the 55 fosmid clones repeatedly failed Sanger end sequencing, probably due to secondary structure effects of the insert DNA. Table 6.3 lists the remaining 102 Sanger end sequencing results (including upstream and downstream sequencing).

Shotgun sequencing resulted in 258 contigs, 128 from the pCT3FK library pool and 130 from the pCC1 fosmid pool. Of the 258 contigs, 54 contigs containing putative CAZyme-encoding genes (dbCAN annotation) originated from microbial communities from biogas fermenters enriched on SBP, while 70 contigs encoding putative CAZymes were from the elephant feces samples (Table 6.4). The average contig length from the pCC1 and pCT3FK fosmid pools was 36,543 and 37,544 bps, respectively (Table 6.4), close to the predicted insert size of around 40 kb. *E. coli* genomic DNA contamination and the fosmid backbone sequences were discovered in the generated contigs by manual BLAST analysis of each contig against the NCBI database, and they were eliminated for further analysis. The presence of these sequence contaminations did not interfere with clone sequence retrieval as the retrieval was done using the clones' end sequences, but it may affect "required-depth of coverage calculations." The remaining 81 non-*E. coli* contigs, including 39 from the pCC1 fosmid pool and 42 from the pCT3FK fosmid pool, were further analyzed (Table 3.5). Sanger end sequencing assigned 54 of the 81 contigs to fosmid clones with recognized functions, and 57.40% of these contigs encoded several putative CAZymes by dbCAN annotation (Table 3.5). A sequence gap in the coverage of the central or terminal region of the insert DNA was found in 47% of the recombinant fosmids sequenced.

Table 3.5 Summary of "non-*E. coli*" contigs generated from shotgun sequencing. (Sequenced samples comprise 40 positive fosmid clones with pCC1 fosmid backbone and 15 positive fosmid clones with pCT3FK fosmid backbone)

Retrieved non-*E. coli* contigs [a]	pCC1 Fosmid Pool (40 clones)	pCT3FK Fosmid Pool (15 clones)	Both Pools
Number of contigs	39	42	81
Number of retrieved contigs [b]	32	22	54
No. of contigs with CAZymes [c]	13	18	31
Total length of contigs (bps)	295,166	344,179	639,345
Average length of contigs (bps) [d]	7,568	8,194	7,893

The chromogenic polysaccharide hydrogel (CPH) and insoluble chromogenic biomass (ICB) substrates were applied to metagenomic library screening.

[a] Contigs from vector backbone or from *E. coli* genomic DNA contamination were confirmed by the size of contig (far more than 40 kb) and sequence identity searched against NCBI database.

[b] Retrieved contigs refer to the contigs which could be aligned to at least one Sanger end sequence.

[c] CAZymes were annotated by dbCAN database.

[d] Average length of contigs was determined by using the sum of the contig lengths divided by the amount of contigs after filtering out the *E. coli* originated contigs.

This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.4 Putative CAZymes analysis

Within the metagenome dataset, putative GH-encoding genes from 26 distinct GH families comprising 71 putative CAZymes were identified using the dbCAN database including members from GH43, GH10, GH51, GH2, and GH11 (Table 6.5). The results revealed that the most common putative GHs identified in the microbiome of environmental samples were GH43 and GH10 (Table 6.5, Table 3.6). Members of the GH43 and GH10 families are involved in the breakdown of hemicellulose and pectic substrates (http://www.cazy.org/). Compared to the vast number of CMCase positive clones (31 out of 55 clones) revealed by functional screening, only a few genes (in total 12 genes) were identified on the contigs, which encode putative enzymes associated with cellulose/CMC hydrolysis (Table 3.6). Even though the microbial communities from the fermenter samples can successfully digest the pectin-rich biomass SBP, surprisingly no potential pectinase-encoding genes were discovered, neither through functional library screening nor through sequence analysis of the assembled contigs.

Table 3.6 Statistics of genes encoding putative fibrolytic enzymes from Shotgun sequencing.

CAZy family [a]	Putative activity [b]	Gene Number
Hemicellulose degrading ability		
GH10	Endo-1,4-beta-xylanase [EC:3.2.1.8]	13
GH2	β-mannosidase [EC:3.2.1.25]	4
GH11	Endo-1,4-beta-xylanase [EC:3.2.1.8]	4
GH67	α-glucuronidase [EC:3.2.1.139]	1
CE1	Endo-1,4-beta-xylanase [EC:3.2.1.8]	2
CE4	Endo-1,4-beta-xylanase [EC:3.2.1.8]	3
CE6	para-nitrobenzyl esterase [EC:3.1.1.-]	1
GH43	Galactosidase [EC:3.2.1.23]	2
GH2	β-galactosidase [EC:3.2.1.23]	1
GH4	α-galactosidase [EC:3.2.1.22]	1
GH127	α-galactosidase [EC:3.2.1.22]	1
Cellulose degrading ability		
GH3	β-glucosidase [EC:3.2.1.21]	3
GH9	Endo-glucanase [EC:3.2.1.4]	3
GH13	Starch synthase (maltosyl-transferring) [EC:2.4.99.16]	2
GH94	Cellobiose phosphorylase [EC:2.4.1.20]	1
GH1	β-glucosidase [EC:3.2.1.21]	1
GH8	Endoglucanase [EC:3.2.1.4]	1
GH16	Endo-1,3-1,4-beta-glycanase ExoK [EC:3.2.1.-]	1
Pectic substrate degrading ability		
GH43	Arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	6
GH43	Endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	5
GH51	α-L-arabinofuranosidase [EC:3.2.1.55]	2
GH51	α-L-arabinofuranosidase [EC:3.2.1.55]	2
Others		
GH24	Lysozyme [EC:3.2.1.17]	1
GH73	N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	1

The dbCAN database annotated 71 potential CAZyme-encoding genes (Table 6.5). This table only displays the combined findings of both dbCAN and KEGG annotations.

[a] The glycoside hydrolase (GH) and carbohydrate esterase (CE) families were predicted using the dbCAN database,

[b] Putative CAZymes were annotated by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.5 Discovery of CAZyme gene clusters (CGC)

As already outlined above, the genomic information of individual fosmid inserts was revealed by mapping Sanger end sequences onto contigs obtained from Illumina shotgun sequencing, allowing the investigation of the content, organization and structure of fibrolytic genes on the inserts. Bioinformatic analysis of 54 contigs showed that 31 of them contained multiple CAZyme-encoding genes. Many

CAZyme-encoding genes were identified to be clustered in CAZyme gene clusters (CGC), which indicates bundled genes encoding functionally related enzymes, putatively involved in the concerted breakdown of substrate polymers to monomers (Figure 3.13). Some of the gene clusters also include genes encoding carbohydrate-binding domains, transport systems and regulatory proteins. For example, contig 40 (assigned to fosmid clone BM1T-2 H7), carrying five genes that encode putative GHs, including a GH67 glucuronidase, a GH3 glucosidase, a GH10 endo-xylanase and two GH43 endo-arabinanases, has the potential to deconstruct both the backbone and side chains of arabinoxylan (Figure 3.13).

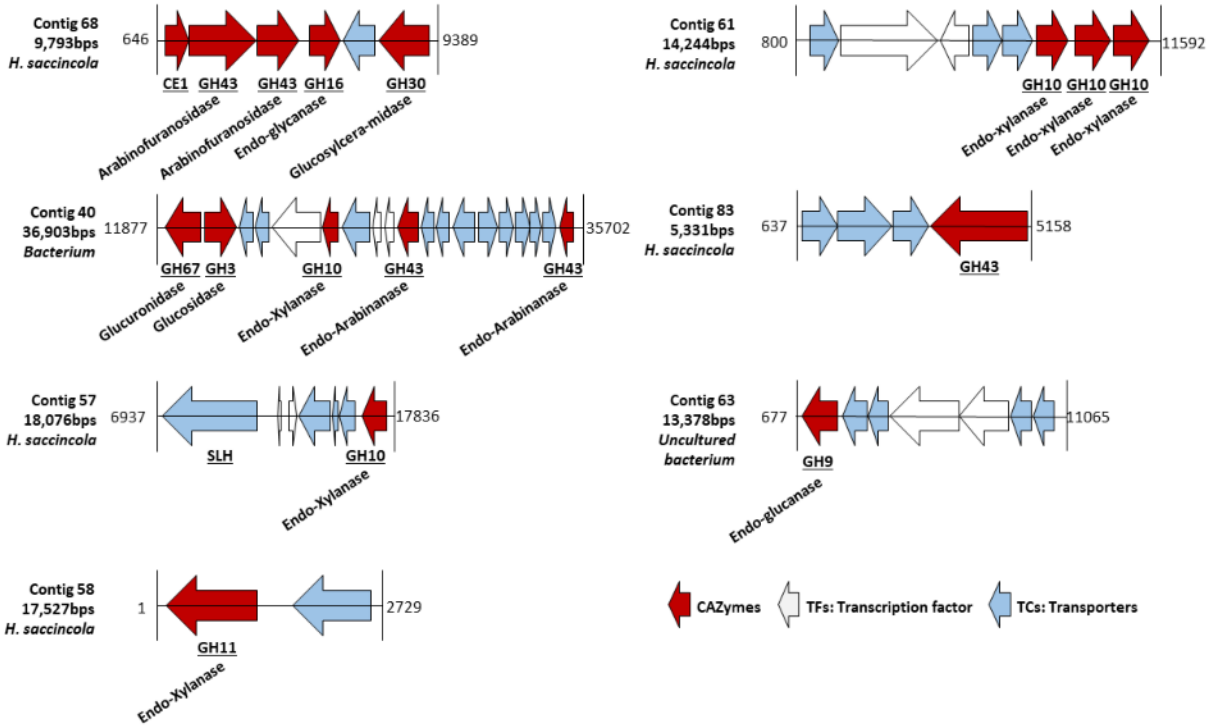


Figure 3.13 Representative gene clusters encoding fibrolytic enzymes that target different plant polysaccharides. CAZyme genes were annotated using the dbCAN and KEGG databases. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.6 Evaluation of the pool sequencing results

By alignment of the end sequence tags acquired from Sanger sequencing to the contigs generated by Illumina shotgun sequencing, finally all the 51 fosmid clones (four clones failed sanger end sequencing as mentioned before) could be assigned to corresponding contigs. We further categorized the 51 clones into 4 different clone types (Type I, II, III and IV) to assess clone coverage based on the number of retrieved end-tags (Figure 3.14). Six fosmid clones were in the group “Type I,” which represents the

expected outcome, in which the two end-tags retrieved the same contig from the pool. In this case, pooled sequencing resulted in 100% coverage for the clones (Figure 3.14, Table 3.7). “Type II” describes a situation in which end-tags aligned to different contigs due to a gap in coverage in the middle of the inserts, which was the case for 18 fosmid clones (Figure 3.14, Table 3.7). Besides, eight clones belong to “Types III,” which represents cases in which coverage was variable and likely underestimated, given that one of the two end-tags was either missing or failed to retrieve a contig (Figure 3.14, Table 3.7). By contrast, in type IV, more than two end-tags were aligned to the same contig, which was possibly caused by the conserved sequence similarity of some functional genes (Figure 3.14). As a result, assembling similar reads into correct contigs may be challenging. There are 19 clones in this category, and most of them were CMCase positive clones (Figure 3.14, Table 3.7).

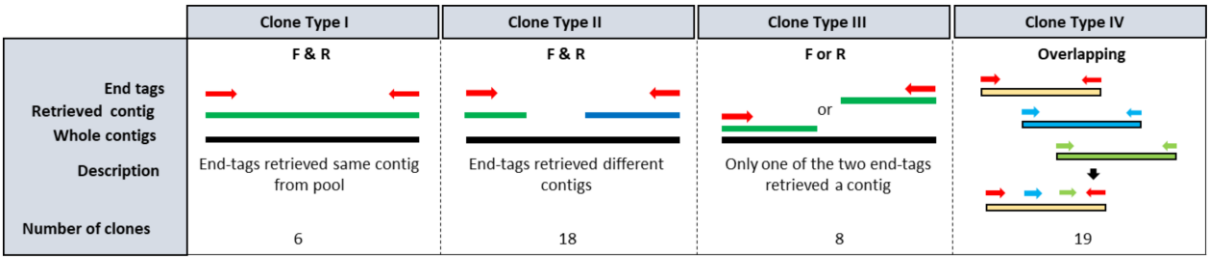


Figure 3.14 Four types of clone sequencing retrieval results.

Table 3.7 Evaluation the coverage of each fosmid clone.

4 Types	Fosmid clones
Type I	T1T2_33, BM1T-2 G5, BM1T-2 B12, BM1T-2 H7, T1T2_8, T1T2_31
Type II	MOD18_28, T1T2_14, MOD18_7-2, MOD18_8, T1T2_20, T1T2_30, MOD18_24, T1T2_16, MOD18_29, BM1T-6 D5, BM1T-7-1 A11, MOD18 Ara3, BM1T-3 D12, BM1T-4 G4, BM1T-9 D12, BM1T-2 B9, BM1T-4 B7, ED-26m C9-1
Type III	T1T2_6, MOD18_14, MOD18_26, MOD18_7-1, MOD18_11, BM1T-10 B11, ED-21S B6/B1, T1T2_24
Type IV	T1T2_2, T1T2_1, T1T2_4, MOD18_17, MOD18_13, T1T2_26, T1T2_23, MOD18_18, T1T2_21, T1T2_29, T1T2_17, T1T2_22, T1T2_18, MOD18_12, MOD18 Ara1, MOD18 Ara2, MOD18 Ara4, BM1T-3 A7, BM1T-3 A11

* Four types of fosmid clones as described in Figure 3.14.

3.2.7 Fosmids without annotated enzymes relevant to its function

The genetic information of the clones’ fosmid inserts can be further predicted by comparing sequence similarity to the known database using various bioinformatics tools, such as searching against the dbcan, KEGG, NCBI and CAZyme databases. Fosmid inserts without any annotated function or related function, but with proven activities of the corresponding fosmid clones may contain novel CAZyme-encoding genes with little sequence similarity to the current database. Table 3.8 lists the screened fosmid clones

that may contain novel enzymes. However, the pool sequencing method had constraints such as under-coverage due *to E. coli* genomic DNA contamination or faulty post-sequencing assembly, making the "potential novel enzymes" ambiguous. By conducting primer-walking along the fosmid inserts in principle the sequence gaps in the inserts could be figured out. However, due to time limitations, we decided to focus on the other interesting genes related to pectic substrate degradation.

Table 3.8 Summary the retrieved contigs without harboring the relevant activities.

Name	Forward end tag	Size(bps)	Reverse end tag	Size(bps)	Activity	Type
BM1T_7-1 A11	pCT3FK_Contig73	8,789	pCT3FK_Contig63	13,378	Arabinoxylanase Xylanase &	II
BM1T_4 G4	pCT3FK_Contig67	11,165	pCT3FK_Contig108	1,992	Arabinoxylanase	II
T1T2_14	pCC1_Contig99	1,806	pCC1_Contig76	5,633	CMCase	II
MOD18_24	pCC1_Contig76	5,633	pCC1_Contig90	2,535	CMCase	II
MOD18_7-2	pCC1_Contig76	5,633	pCC1_Contig90	2,535	CMCase	II
MOD18_8	pCC1_Contig76	5,633	pCC1_Contig90	2,535	CMCase	II
T1T2_16	pCC1_Contig101	1,629	pCC1_Contig76	5,633	CMCase	II
T1T2_30	pCC1_Contig122	1,062	pCC1_Contig108	1,238	CMCase	II
T1T2_20	pCC1_Contig122	1,062	pCC1_Contig108	1,238	CMCase	II
T1T2_14	pCC1_Contig99	1,806	pCC1_Contig76	5,633	CMCase	II
MOD18_18	pCC1_Contig90 pCC1_Contig100	2,535 1,696	pCC1_Contig94	2,097	CMCase	IV
T1T2_22	pCC1_Contig99	1,806	pCC1_Contig76 pCC1_Contig97 pCC1_Contig116	5,633 1,908 1,099	CMCase	IV

3.2.8 Expression and purification of the seven arabinosyl hydrolases

Among all the obtained contigs, contig 68, contig 57, contig 72 and contig 60 can be assembled into a ~56 kb DNA fragment via around 125 bps overlapping regions (the presence of contig 68 in the DNA fragment was confirmed by PCR). The existence of co-linearly organized genes encoding putative enzymes involved in pectic arabinan degradation, sugar-binding, transport, and subsequent pentose metabolism was annotated using bioinformatic tools (KEGG and dbCAN) (Figure 3.15).

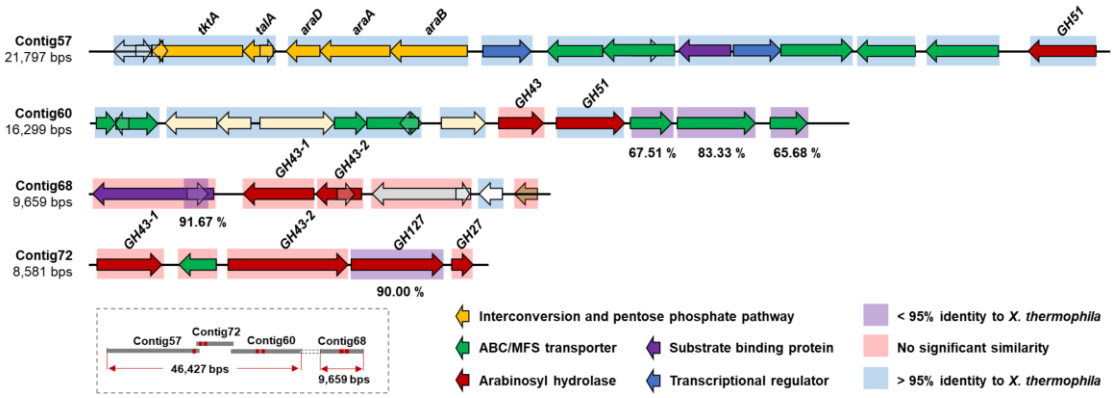


Figure 3.15 Genetic organization of putative gene clusters on an about 56 kb assembled metagenomic DNA fragment. Schematic organization of four contigs in the box represents the assembled DNA fragment composed of contig68, contig57, contig72 and contig60 with around 125 bps overlapping region in two contigs. Additionally, the presence of contig72 in the assembled DNA fragment was confirmed by PCR. The fragment includes genes encoding putative enzymes involved in arabinan and arabino-oligosaccharide degradation, substrate binding, transport, regulatory proteins and pentose metabolism. Green color: genes encoding putative ABC-transporter cassettes/MFS transporter. Red color: genes encoding putative arabinosyl hydrolases including family GH43, GH51, GH127 and GH27. Orange color: gene encoding major enzymes involved in interconversion and pentose phosphate pathway. Blue shadow: more than 95% nucleotide sequence identity with *X. thermophyla*. Red shadow: no significant similarity. Purple shadow: lower than 95% nucleotide sequence identity with *X. thermophyla*. This result has been included in the manuscript by Liu *et al.*, (2022).

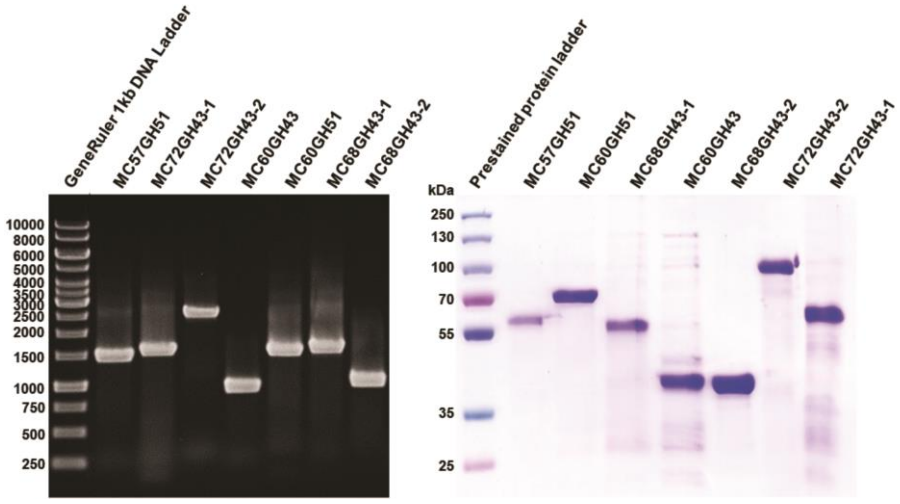
Among them, seven putative arabinosyl hydrolases from families GH51 and GH43, designated MC57GH51, MC60GH43, MC60GH51, MC68GH43-1, MC68GH43-2, MC72GH43-1, MC72GH43-2, which encode a putative α-L-arabinofuranosidase, α-L-arabinofuranosidase, endo-1,5-α-L-arabinosidase, endo-1,5-α-L-arabinosidase, endo-1,5-α-L-arabinosidase, endo-1,5-α-L-arabinosidase, endo-1,5-α-L-arabinosidase, respectively, were chosen for further characterization. Table 3.9 shows the basic characteristics of each enzyme.

Table 3.9 Basic characteristics of characterized arabinosyl hydrolases.

Name	KEGG annotation	Gene Size (bps)	MW (kDa)	pI	SignalP
MC57GH51	α-L-arabinofuranosidase [EC: 3.2.1.55]	1,500	56.65	5.56	N
MC60GH51	Arabinan endo-1,5-α-L-arabinosidase [EC: 3.2.1.99]	1,533	58.40	5.10	N
MC68GH43-1	Arabinan endo-1,5-α-L-arabinosidase [EC: 3.2.1.99]	1,575	58.8	5.06	N
MC60GH43	α-L-arabinofuranosidase [EC: 3.2.1.55]	960	36.71	5.21	N
MC68GH43-2	Arabinan endo-1,5-α-L-arabinosidase [EC: 3.2.1.99]	1,002	39.23	6.29	N
MC72GH43-1	Arabinan endo-1,5-α-L-arabinosidase [EC: 3.2.1.99]	1,407	52.18	4.36	N
MC72GH43-2	Arabinan endo-1,5-α-L-arabinosidase [EC: 3.2.1.99]	2,580	95.88	4.51	N

The function of each enzyme was predicted by BLAST analysis against dbCAN and KEGG databases, gene size, the molecular weight of the encoded protein and isoelectric point (pI) were obtained from Clone manager, the signal peptide was predicted by Signal-5.0 database.

To investigate the biochemical characteristics and hydrolysis products produced by these putative arabinosyl hydrolases, DNA sequences containing the entire ORFs encoding these putative arabinosyl hydrolases were amplified using PCR with primer-based incorporation of His_6 tags at the C-termini of the target proteins. Figure 3.16a shows the PCR products with the expected size of each target gene. *E. coli* BL21 was used as heterologous host to produce the proteins. Small scale cultures were utilized to determine the best IPTG concentration and incubation duration for protein expression. Purification of the generated proteins with molecular masses of 56.65, 58.40, 36.71, 58.80, 39.23, 52.18 or 95.88 kDa for MC57GH51, MC60GH43, MC68GH43-1, MC68GH43-2, MC72GH43-1, MC72GH43-2, respectively, was carried out using His-tag based affinity chromatography followed by anion exchange chromatography to obtain an adequately high yield and purity of the proteins. SDS-PAGE was used to determine the purity of the proteins, revealing a single band for all proteins with electrophoretic mobilities compatible with the estimated molecular weights of the hexa-histidine fusion proteins based on their amino acid sequences (Figure 3.16b).



(a) (b)

Figure 3.16 Cloning and expression of seven putative arabinosyl hydrolases from an assembled DNA fragment. (a) PCR amplification of the putative arabinosyl hydrolase-encoding genes was performed by using fosmid clone MOD18_Abn4 as a template. Lane 1: GeneRuler 1kb, Lane 2: MC57GH51 (1,500 bps); Lane 3: MC72GH43-1 (1,407 bps); Lane 4: MC72GH43-2 (2,580 bps); Lane 5: MC60GH43 (960 bps); Lane 6: MC60GH51 (1,533 bps); Lane 7: MC68GH43-1 (1,575 bps); Lane 8: MC68GH43-2 (1,002 bps). (b) SDS-PAGE analysis of the purified enzymes. The proteins were purified by immobilized metal affinity chromatography with a Ni-TED packed column and subsequent anion exchange chromatography with a QFF column. Lane 1: Prestained protein standards, Lane 2: MC57GH51 (56.65 kDa); Lane 3: MC60GH51 (58.40 kDa); Lane 4: MC68GH43-1 (58.80 kDa); Lane 5: MC60GH43 (36.71 kDa); Lane 6: MC68GH43-2 (39.23 kDa); Lane 7: MC72GH43-2 (95.88 kDa); Lane 8: MC72GH43-1 (52.18 kDa). This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.9 Phylogenetic relationship of the seven arabinosyl hydrolases

A multiple amino acid sequence alignment was conducted using each enzyme and each of its most closely related sequences (top two to six best matches in the NCBI-nr database) to highlight differences in sequences and phylogenetic relationships, followed by the construction of a neighbor-joining tree (Figure 3.17). All putative GH43 family members were well separated from GH51 proteins in the evolutionary tree. The sequence identities of the putative GH43 enzymes varied from 0 to 60.65%, whereas the two putative GH51 enzymes had 28.27% identity. Based on a BLAST search against the NCBI RefSeq database, MC72GH43-2 shared 74.94% amino acid sequence identity with an S-layer domain protein (ADH61400.1) from *Thermoanaerobacter mathranii*. In comparison, the other GH43 enzyme encoded on the same contig shared 63.68% amino acid sequence identity with a GH43 protein (WP 102710436.1) from *Paenibacillus castaneae*. Two GH43 proteins encoded on contig68 (MC68GH43-1 and MC68GH43-2) had the closest phylogenetic relationship (81.35% and 84.31%, respectively) with two GH43 enzymes from *Thermoclostridium stercorarium* (WP 015358800.1 and WP 015358799.1) (Figure 3.17). Multiple alignments of GH51 amino acid sequences from Contig60 and Contig57 with proteins from the RefSeq database found high identities of 97.42% and 99.59%, respectively, with two protein sequences from *Xylanivirga thermophila* (WP 129721540.1 and WP 129721536.1) (Figure 3.17).

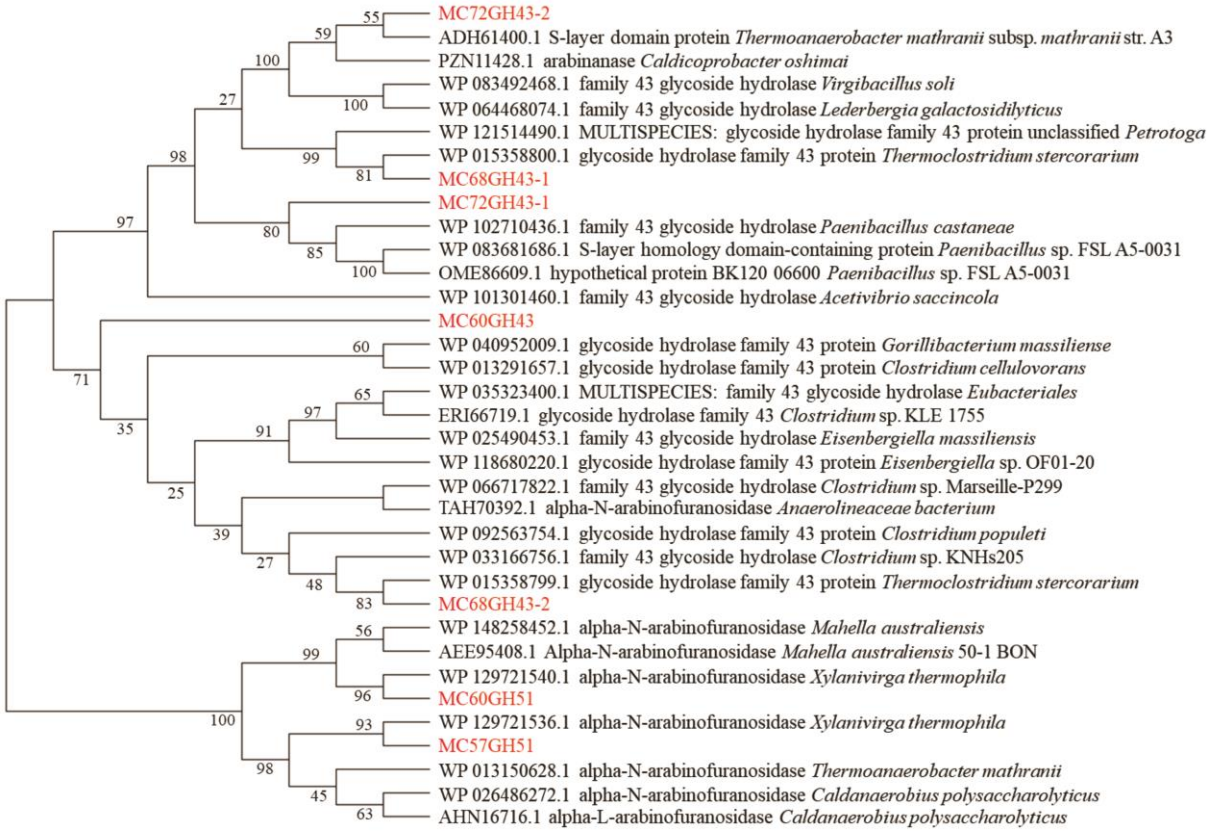


Figure 3.17 Phylogenetic analysis of the diversity of seven putative arabinosyl hydrolases. Multiple sequence alignments were conducted with MUSCLE, using full-length sequences of GH43 or GH51 enzymes and their top 2 to 6 BLAST hits in the NCBI-nr database. Then, using MEGAX and bootstrap values from 500 repetitions, a Neighbor-Joining (NJ) tree was constructed. Proteins from this study are highlighted in red, whereas proteins from other sources are colored in black. For each sequence, the database accession number and organism are displayed. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.10 Determination of the temperature and pH optimum

Using either *p*NP-AF (MC57GH51 and MC60GH51) or DA (MC68GH43-1, MC60GH43, MC68GH43-2, MC72GH43-2) as substrates, the best conditions regarding temperature and pH for each enzyme's activity were established. The maximum activity of MC60GH43 was at 34.4°C (Figure 3.18d), while the maximum activities of the remaining enzymes ranged from 49°C to 56°C (Table 3.10). In a temperature range of 25.0 to 41.9°C, MC60GH43 sustained more than 60% activity (Figure 3.18d). MC68GH43-2 and MC72GH43-2 were more than 60% active at temperatures ranging from 42.7 to 56.9°C and 41.9 to 59.4°C, respectively (Figure 3.18e,f). MC57GH51, MC60GH51 and MC68GH43-1 finally displayed more than 60% of their maximum activity at 45.6 - 70.6°C, 38.1 - 63.1°C, and 37.6 - 57.5°C, respectively (Figure 3.18a,b,c). All enzymes favored a slightly acidic pH, with an activity maximum between pH 4.5

and pH 6.5 (Table 3.10). The stability of the enzymes was investigated at conditions similar to their temperature and pH optimum after the optimal conditions for enzyme activity were identified. For this, each enzyme was incubated for up to 24 hours at the temperature and pH where it had shown maximal activity, and residual activities were measured using DNS assays with SBA or DA as substrate. The results showed that after being kept at their temperature and pH optimum for 24 hours, MC68GH43-2 and MC60GH43 lost up to 30% of their original activity after 24 h under temperature and pH conditions of their maximum activity (Figure 3.19). The remaining four enzymes were more stable. During the same time span, they maintained more than 80% of their activity (Figure 3.19).

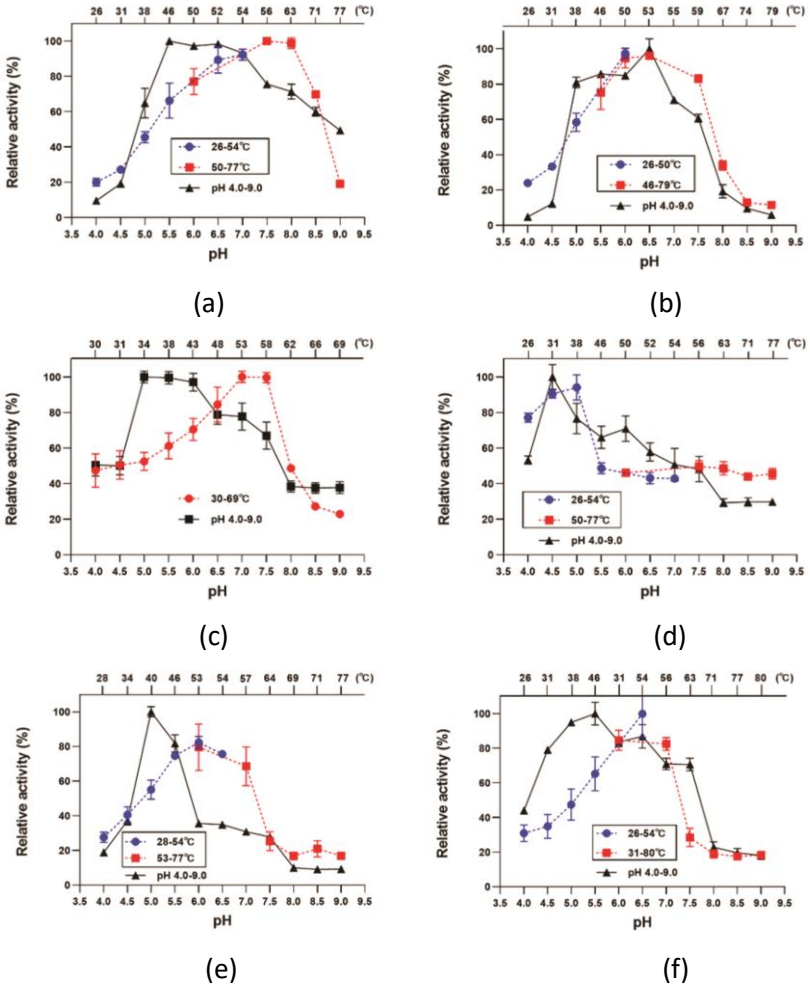


Figure 3.18 Temperature and pH dependence of enzyme activities. (a) MC57GH51, (b) MC60GH51, (c) MC68GH43-1, (d) MC60GH43, (e) MC68GH43-2, (f) MC72GH43-2. The graphs show the relative activities of the characterized arabinosyl hydrolases with 1 mM *p*NP-AF (143.69 nM MC57GH51 and 19.86 nM MC60GH51) or with 5 g L^{-1} DA (875.62 nM MC60GH43, 367.05 nM MC68GH43-2 and 465.25 nM MC72GH43-2) or SBA (MC68GH43-1) at different pH values and temperatures. Assays were carried out with 5 g L^{-1} of DA (debranched arabinan), SBA (Sugar beet arabinan) at pH 4.0 to 9.0 and between 25 ℃ to 80 ℃ for 30 min, or with 1 mM *p*NP-AF (*p*-Nitrophenyl-α-L-arabinofuranoside) for 10 min. Enzyme activities were determined by the reducing end-specific DNS assay (with DA and SBA as substrates) or *p*NP assay (with *p*NP-AF as substrate) by measuring of the absorbance at OD_{540} and OD_{405}, respectively. The assays were performed in triplicates.

Table 3.10 Summary the optimal reaction conditions of arabinosyl hydrolases.

Enzymes	pH *opt*	pH range	T opt (°C)	T range (°C)	Substrates
MC57GH51	5.5	5.0-8.5	55.7	45.6-70.6	*p*NP-AF
MC60GH51	6.5	5.0-7.5	50.8	38.1-63.1	*p*NP-AF
MC68GH43-1	5.0	5.0-7.5	52.5	37.6-57.5	SBA
MC60GH43	4.5	4.5-6.5	34.4	25.0-41.9	DA
MC68GH43-2	5.0	5.0-5.5	49.3	42.7-56.9	DA
MC72GH43-2	5.5	4.5-8.0	55.0	41.9-59.4	DA

The enzymes (143.69 nM MC57GH51, 19.86 nM MC60GH51, 875.62 nM MC60GH43, 367.05 nM MC68GH43-2, 465.25 nM MC72GH43-2, 297.62 nM MC68GH43-1) were incubated in 200 µL reaction volume with 5 g L^{-1} of the DA (MC60GH43, MC68GH43-2 and MC72GH43-2), SBA (MC68GH43-1) at pH 4.0 to 9.0 and between 25 ℃ to 80 ℃ for 30 min or in 50 uL reaction with 1 mM *p*NP-AF (MC57GH51 and MC60GH51) for 10 min. Enzyme activities were determined by the reducing-end DNSA assay (with DA/SBA as substrates) or *p*NP assay (with *p*NP-AF as substrate) by measuring of absorbance at OD_{540} and OD_{405},respectively, and represent mean values of triplicate assays. Borders for pH and temperature ranges were set to 60% of the highest activity. This result has been included in the manuscript by Liu *et al.*, (2022).

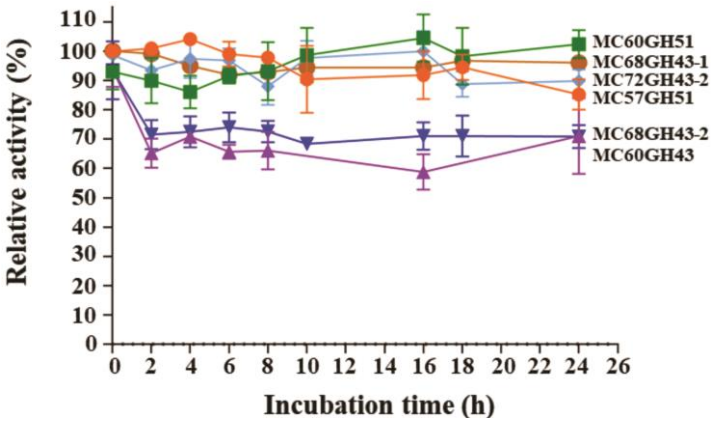
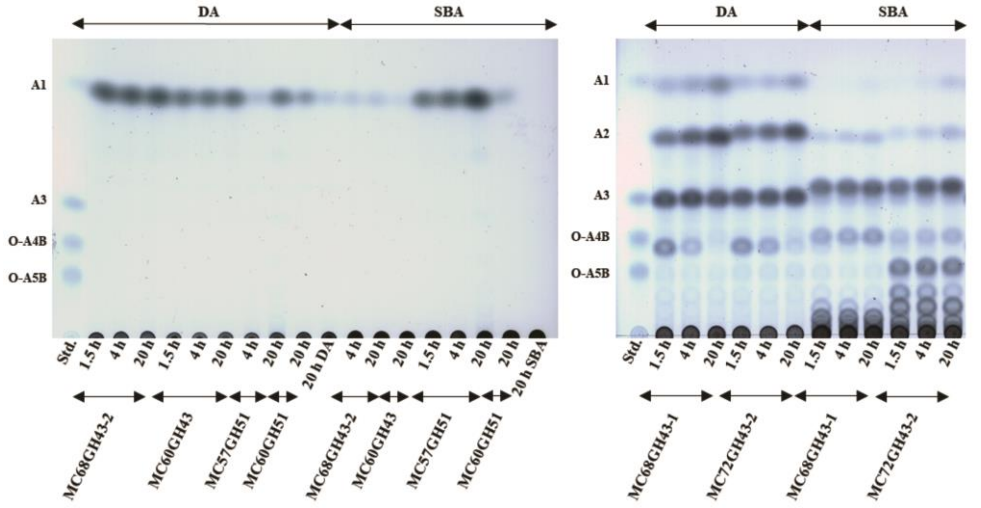


Figure 3.19 Effect of long-term incubation of arabinosyl hydrolases at temperature and pH conditions of their maximum activity. Relative activities were calculated from DNS assay with SBA or DA as substrates. Standard reactions included 5 g L^{-1} of DA, 25 mM citrate phosphate buffer with 283.3 nM MC60GH43, 547.54 nM MC68GH43-2, 99.5 nM MC72GH43-2 or 5 g L^{-1} the SBA, 25 mM citrate phosphate buffer with 120.04 nM MC57GH51, 410.96 nM MC60GH51, 892.86 nM MC68GH43-1. The maximal activity was set as 100%. Error bars represent standard deviations (n=3).

3.2.11 Activity of the arabinosyl hydrolases on polysaccharides and kinetic parameters

The Michaelis-Menten model was used to calculate the kinetic parameters of each enzyme, including K_m and V_{max} values. Only MC72GH43-1 did not exhibit any activity on any of the substrates examined. All of the remaining enzymes were active against the synthetic substrates *p*NP-AF and/or arabinan. Both MC57GH51 and MC60GH51 were active on *p*NP-AF and capable of releasing arabinose as the sole

hydrolysis product from SBA and DA (Figure 3.20a), indicating both of them are exo-acting-α-L-arabinofuranosidases. On SBA, MC57GH51 had a specific activity of 2.02 U mg^{-1}, but its activity on DA was challenging to detect using the DNS assay (Table 3.11). With SBA as a substrate, MC60GH51 demonstrated lower activity than MC57GH51 (0.06 U mg^{-1}) but showed more significant activity on DA (0.15 U mg^{-1}). With DA as the substrate, MC68GH43-2 and MC60GH43 had specific activities of 1.74 U mg^{-1} and 1.65 U mg^{-1}, respectively, but no activity with SBA. These enzymes' activity on both substrates produced arabinose as sole hydrolysis product (Table 3.11, Figure 3.20a), indicating that MC68GH43-2 and MC60GH43 are both exo-L-arabinofuranosidases. The activity of both enzymes on *p*NP-AF, however, was minimal. Regarding MC68GH43-1 and MC72GH43-2, little activity was found on *p*NP-AF, while they displayed a marked activity on SBA and DA. TLC evaluation of MC68GH43-1 and MC72GH43-2 hydrolysis products from various arabinan substrates indicated that the hydrolysis products comprised various AXOs and a tiny quantity of arabinose with SBA as the substrate (Figure 3.20b). Both enzymes are thus categorized as endo-arabinanases. After 20 hours of incubation, the hydrolysis products of DA were quite comparable, including arabinose, arabinobiose, and arabinotriose, but MC72GH43-2 released larger oligosaccharides from SBA than MC68GH43-1 (Figure 3.20b). Finally, none of these enzymes showed any action against wheat arabinoxylan, even though this heteroxylan, just as SBA, includes arabinose side chain substitutions.



(a) (b)

Figure 3.20 TLC analysis of hydrolysis products released by arabinosyl hydrolases characterized in this study. The hydrolysis products released by (a) exo-arabinofuranosidases and (b) endo-arabinanases from SBA or DA were separated with TLC. Standard reactions were carried out by incubating 5 g L^{-1} of SBA or DA with 2.64 µM MC68GH43-2, 457.64 nM MC60GH43, 211.83 nM MC57GH51, 19.69 nM MC60GH51, 130.37 nM MC72GH43-2, or 1.22 µM MC68GH43-1 at the enzymes' optimal working condition (25 mM citrate phosphate buffer was used to maintain the desired pH in the reaction) for different periods. 10 µL hydrolysis products were separated three times on TLC silica plates using chloroform: acetate: water (6:7:1, v/v/v) as a solvent. This result has been included in the manuscript by Liu *et al.*, (2022).

Where possible, the kinetic characteristics of the enzymes were investigated using DA or SBA as a substrate. MC57GH51 was the only exo-arabinofuranosidase with measurable affinity for SBA, with a K_m of 117.02 g L^{-1} and a V_{max} of 52.67 mol min^{-1} mg^{-1} (Figure 3.21g). In contrast, the kinetic characteristics of MC60GH43 and MC68GH43-2 could only be assessed with DA, where the former enzyme had a lower apparent binding affinity for DA than the latter (Figure 3.21e,f). The apparent binding affinity of both endo-arabinanases (MC68GH43-1, MC72GH43-2) for SBA was slightly higher than for DA (Figure 3.21a,b,c,d).

Table 3.11 Specific activities of the characterized arabinosyl hydrolyases on DA and SBA.

	Specific activity[a]		Released soluble sugars[b]
Enzymes	**SBA**	**DA**	
MC57GH51	2.02 ± 0.11	ND	Arabinose
MC60GH51	0.06 ± 0.00	0.15 ± 0.01	Arabinose
MC60GH43	NA	1.65 ± 0.17	Arabinose
MC72GH43-1	NA	NA	NA
MC72GH43-2	3.81 ± 0.25	5.21 ± 0.21	AOS, Arabinose
MC68GH43-1	7.51 ± 0.31	2.39 ± 0.26	AOS, Arabinose
MC68GH43-2	ND	1.74 ± 0.02	Arabinose

[a] Specific activity in µmol min^{-1} mg^{-1}. The data shows the mean values and standard deviations from 3 experiments.

[b] Soluble sugars released, analyzed by HPAEC-PAD and TLC.

ND means not determined due to insufficient activity or lack of detectable activity. NA means no activity.

SBA: sugar beet arabinan, DA: debranched arabinan, AOS: arabinooligosaccharides.

This result has been included in the manuscript by Liu *et al.*, (2022).

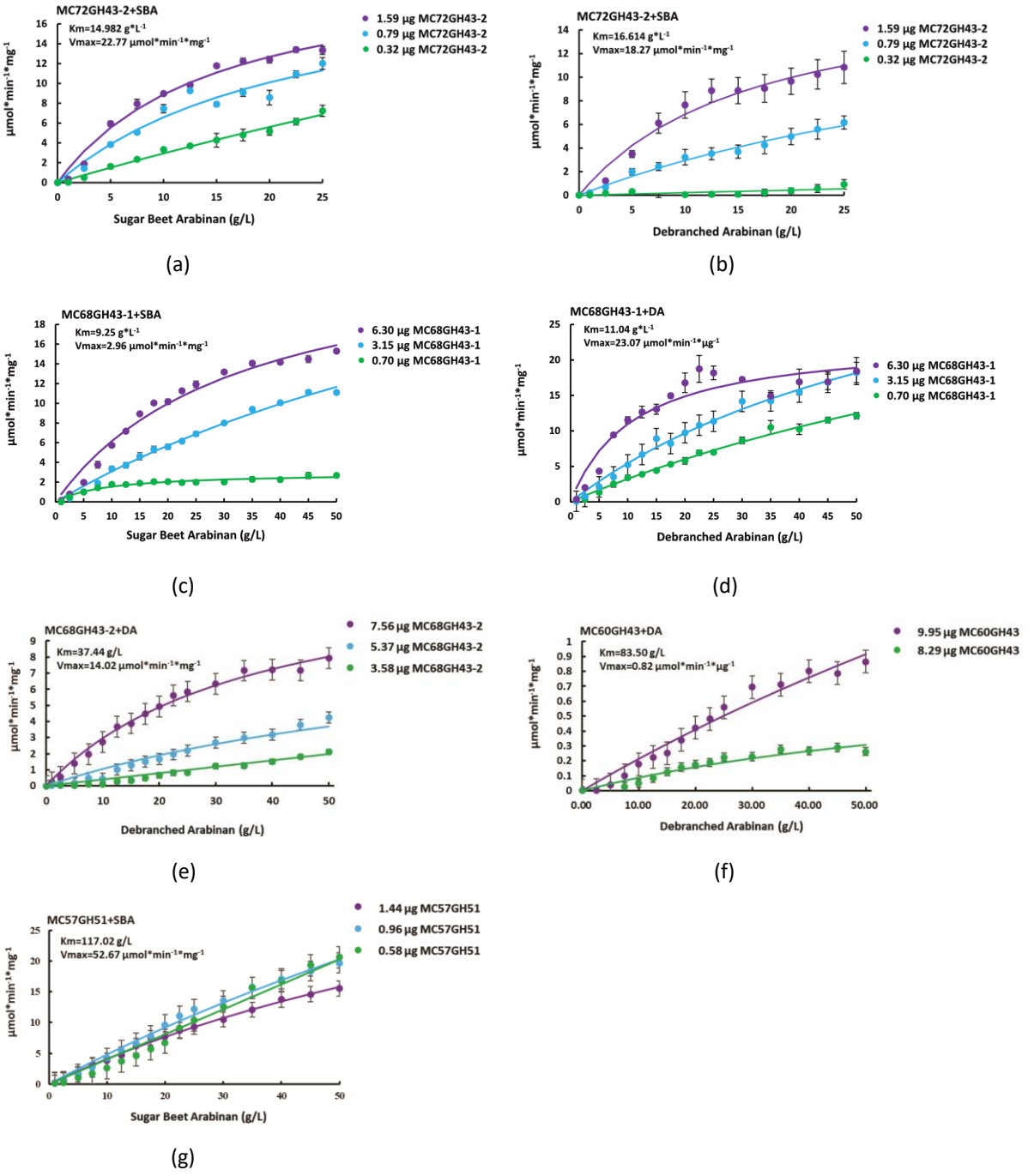


Figure 3.21 Kinetic characteristics of arabinosyl hydrolases with SBA or DA as substrates. Standard reactions were performed by using three different concentrations of each enzyme (as indicated in figures) and various concentration of substrates (between 1 to 50 g L^{-1}) at each enzyme's optimal condition for different time periods according to requirement (MC68GH43-1 and MC60GH43 for 2 h incubation, MC60GH51, MC68GH43-2, MC72GH43-2 for 40 min incubation, MC57GH51 for 1 h incubation). Error bars represent standard deviations of triplicates. K_m and V_{max} were calculated by using Microsoft Excel Solver, as described in Material and methods.

3.2.12 Cleavage specificity of the arabinosyl hydrolases towards oligosaccharides

Except for the polysaccharides, various AOS and AXOS were incubated with the recombinant arabinosyl hydrolases for 12 h and the hydrolysis products obtained were identified by HPAEC-PAD. MC57GH51

and MC60GH51 were shown to be exo-L-arabinofuranosidases and able to cleave α-1,2, α-1,3, and α-1,5-linked arabinofuranosyl residues from purified arabino-oligosaccharides, such as A3, AA^3A, AAA^3A, and $AA^{2,3}A$, resulting in oligosaccharides with a reduced degree of polymerization (Figure 3.22a,b). Because more arabinose was released by an equivalent quantity of enzyme over the same incubation period, MC60GH51 revealed a higher AOS-cleaving activity than MC57GH51 (Figure 3.22a,b). Even though neither MC60GH51 nor MC57GH51 had any activity against the polysaccharide WAX, they were able to liberate arabinose from AXOS, including single 1,2- or 1,3-linked arabinofuranosyl residues, as seen in XA^3XX, XA^2XX, and A^2XX, resulting in X4, X4, and X3, respectively. There was also a tiny amount of arabinose released when MC60GH51 and MC57GH51 were exposed to terminal doubly substituted $A^{2,3}XX$. However, with internal double substituted AXOS, such as $XA^{2,3}XX$, on the other hand, no observable cleavage action was observed (Figure 3.22c,d). MC60GH43 and MC68GH43-2, on the other hand, were only able to liberate arabinose from terminal 1,5-linked arabinofuranosyl residues from A3 and AAA^3A but not from other branched oligosaccharides, which is in agreement with their high activity against polymeric DA, a linear 1,5-linked L-arabinofuranosyl backbone lacking side chains (Figure 3.22e,f).

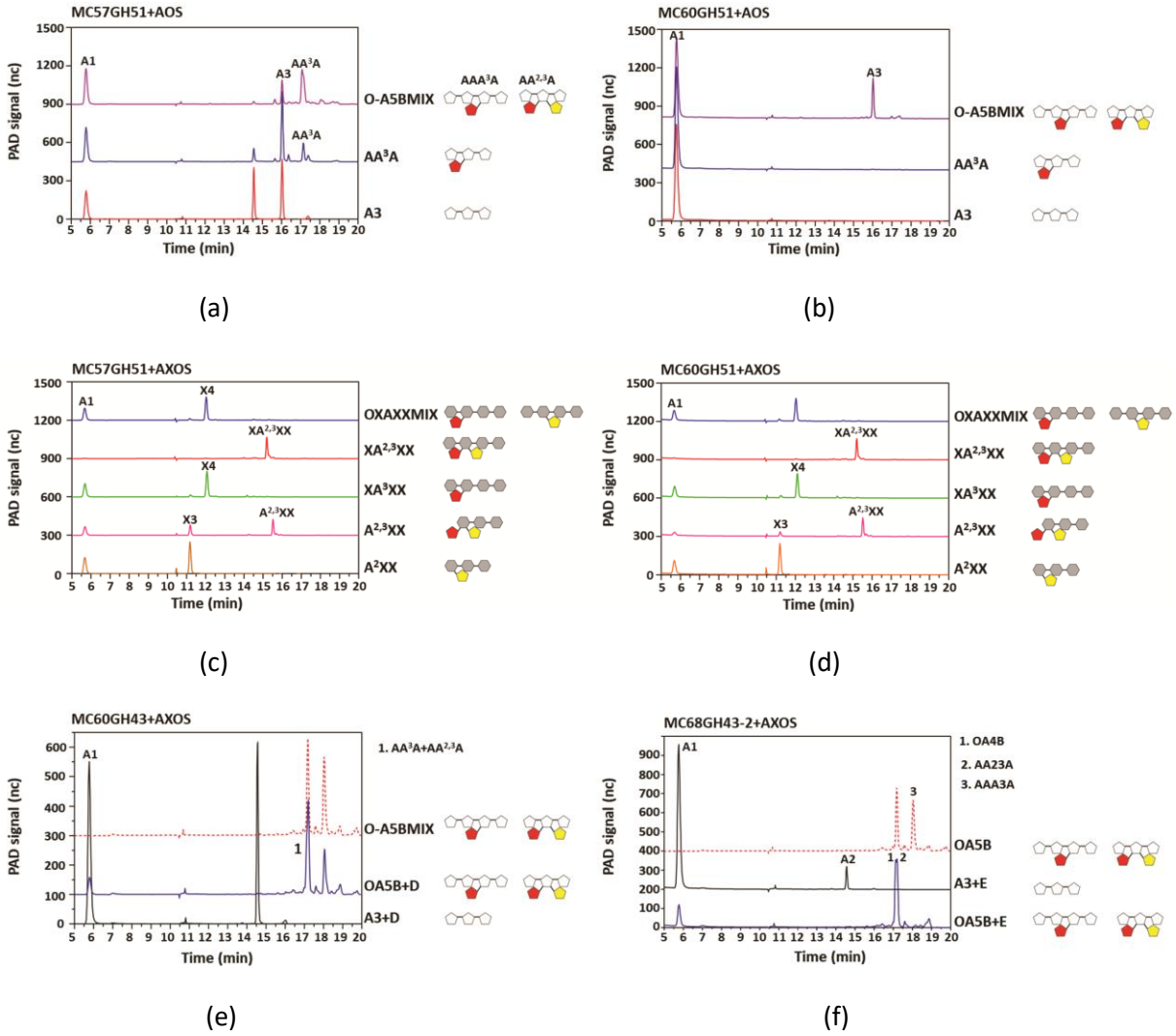


Figure 3.22 HPAEC-PAD analyses and structures of AOS and AXOS hydrolysis products. (a-b) AOS as substrates (c-f) AXOS as substrates. In 10 µL reactions, standard reactions were carried out by incubating 0.1 µg of each enzyme with 2.5 mg mL^{-1} AOS or AXOS for 12 hours at optimum temperature and pH. For HPAEC-PAD analysis, the hydrolysis products were diluted at 1:50 in water with a final volume of 500 µL. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.13 Synergistic action of arabinosyl hydrolases in arabinan degradation

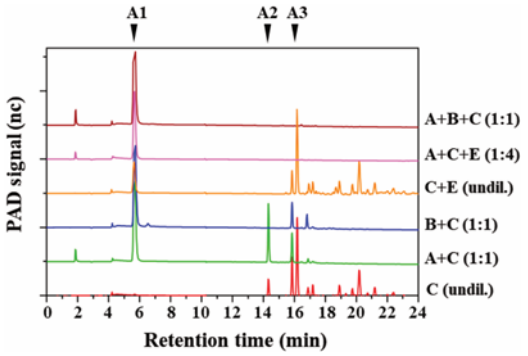
The activity of the single enzyme and various enzyme combinations, including up to six distinct enzymes, was evaluated using SBA and DA as substrates to investigate synergistic effects. HPAEC-PAD was used to measure the quantity of arabinose released. With a maximal production of arabinose of 3,073 mg L^{-1} from SBA and 1,855 mg L^{-1} from DA, a synergistic effect could be detected with various enzyme cocktails used, especially with endo- and exo-enzyme activities (Figure 3.23a). The maximum amount of arabinose released in single enzyme reactions reached only 0.16-26% (from SBA) and 5.54-44.66% (from DA) of what was produced when the maximum yield was reached using enzyme

combinations (Figure 3.23a). Because of the high concentration of reducing sugar but the limited amount of arabinose in the hydrolysis products, MC68GH43-1 and MC72GH43-2 presumably mainly released AXOs from arabinan (Figure 3.23a). Furthermore, not all the combinations of endo-arabinanase and exo-arabinofuranosidase could increase the yield of arabinose released from diverse arabinan substrates considerably. Combining one endo-arabinanase, either MC68GH43-1 or MC72GH43-2, with the exo-arabinofuranosidase(s) MC57GH51 and/or MC60GH51 resulted in a strong synergistic effect for SBA degradation. The yield of arabinose did not improve much when these endo-arabinanases were coupled with MC60GH43 and/or MC68GH43-2. Supposedly, MC57GH51 and MC60GH51 had a broader cleavage specificity and better cleavage efficiency towards different side chains exclusively present in SBA but not in DA. In comparison, MC60GH43 and MC68GH43-2 can only cleave 1,5-arabinosidic bonds in DA. Instead, combining the exo-arabinofuranosidase (MC60GH43 and/or MC68GH43-2) with one of the endo-arabinanases (MC68GH43-1 or MC72GH43-2) enhanced DA hydrolysis significantly. Because the specific activities of both endo-arabinanases against SBA and DA are not dramatically different (all in the range between 2.39 and 7.51 U mg^{-1}; Table 3.11), the results suggested that the synergism of the endo- and exo- arabinosyl hydrolases characterized in our study was mainly dependent on which exo-acting enzyme was chosen rather than which endo-arabinanase was included in the assay reaction. Furthermore, it is noteworthy that the arabinose equivalent of some enzyme assay is higher than theoretically expected with DNS assay (higher than the substrate concentration used in the reaction), which may be due to deviations from the assay methods used.

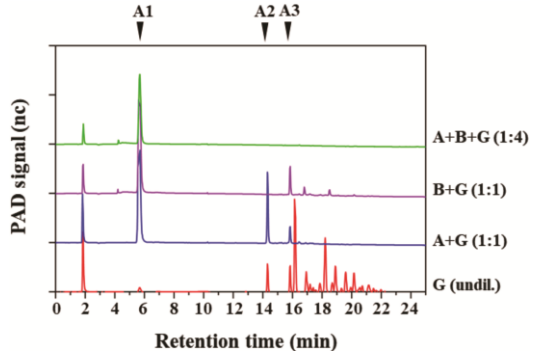
Different enzyme combinations were tested to find out the most efficient one in terms of arabinose production. SBA was efficiently converted to arabinose using a cocktail that included MC57GH51, MC68GH43-1 and MC60GH51. MC68GH43-2 and MC72GH43-2 were good substitutions for MC60GH51 and MC68GH43-1, respectively (Figure 3.23b,c). The most effective cocktail for DA degradation was MC57GH51, MC60GH43, MC68GH43-2 and MC72GH43-2 (Figure 3.23a). After achieving the enzyme's maximal turnover, we saw a slight reduction in arabinose output (Figure 3.23a). This might be explained by the production of large quantities of mono- and oligosaccharides, which may result in end-product inhibition in certain enzymes, as reported previously (Rohman *et al.*, 2018; Wagschal *et al.*, 2009).

	Reducing sugar		Arabinose	
	SBA	DA	SBA	DA
A	8	0	807	109
B	2	2	232	269
C	4	5	10	94
D	0	1	5	182
E	0	7	70	757
G	7	7	8	211
AB	9	3	629	367
AC	33	11	1651	801
AD	8	6	757	724
AE	19	7	1479	877
AG	36	13	2252	930
BC	35	14	1716	1112
BD	2	7	221	658
BE	3	8	260	645
BG	34	23	1900	1484
CD	5	15	160	1238
CE	6	17	263	1438
CG	8	8	34	115
DE	0	8	103	846
DG	7	15	152	1297
EG	8	17	246	1365
ABC	37	17	2160	1370
ABD	10	7	680	657
ABE	17	7	1039	617
ABG	39	24	2224	1443
ACD	34	19	2177	1352
ACE	41	20	2789	1556
ACG	37	17	2270	1090
ADE	18	9	1492	897
ADG	35	21	2248	1512
AEG	37	22	2879	1657
BCD	29	17	2062	1249
BCE	33	18	2140	1281
BCG	33	23	2133	1633
BDE	2	8	343	629
BDG	32	24	2164	1444
BEG	35	21	2284	1466
CDE	5	18	244	1406
CDG	9	18	210	1377
CEG	9	18	276	1471
DEG	8	17	272	1430
ABCD	37	17	2527	1370
ABCE	38	18	2457	1310
ABCG	36	24	2808	1696
ABDE	17	9	1496	774
ABDG	37	26	2721	1611
ABEG	38	25	2754	1590
ACDE	39	20	2921	1430
ACDG	35	23	2396	1730
ACEG	42	24	3073	1695
ADEG	43	22	2999	1774
BCDE	34	19	2225	1326
BCDG	32	24	2306	1622
BCEG	33	24	2402	1762
BDEG	34	25	2403	1624
CDEG	9	22	275	1423
ABCDE	38	20	2926	1414
ABCDG	40	28	2651	1747
ABCEG	40	27	2610	1579
ABDEG	41	27	2671	1844
ACDEG	37	26	2922	1855
BCDEG	31	25	2332	1660
ABCDEG	41	24	2480	1614
Only substrate	0	0	0	0

(a)



(b)



(c)

Figure 3.23 Hydrolysis of SBA or DA by various combinations of endo-arabinanases and exo-arabinofuranosidases. 5 g L^{-1} SBA or DA were incubated in 200 µL reactions containing 25 mM citrate phosphate buffer pH 5.5 with single enzyme or enzyme combinations using enzyme concentrations of 75 nM (endo-arabinanases MC68GH43-1 and MC72GH43-2) and 150 nM (exo-arabinofuranosidases MC57GH51, MC60GH51, MC60GH43 and MC68GH43-2) at 50°C for 24 h. (a) Evaluation the yield of arabinose (HPAEC-PAD) and reducing sugar (DNS assay) released from SBA and DA, in mg L^{-1}. (b-c) HPAEC-PAD analysis of reactions comprising the simplest enzyme combination which can fully hydrolyze SBA. A: MC57GH51, B: MC60GH51, C: MC68GH43-1, D: MC60GH43, E: MC68GH43-2, F: MC72GH43-1, G:MC72GH43-2. This result has been included in the manuscript by Liu *et al.*, (2022).

3.2.14 Bioinformatic identification of additional arabinan degrading enzymes

To identify other genes encoding arabinan degrading enzymes adjacent to the genes for the described arabinosyl hydrolases on the assembled DNA fragment (around 56 Kbp). Using the KEGG and dbCAN

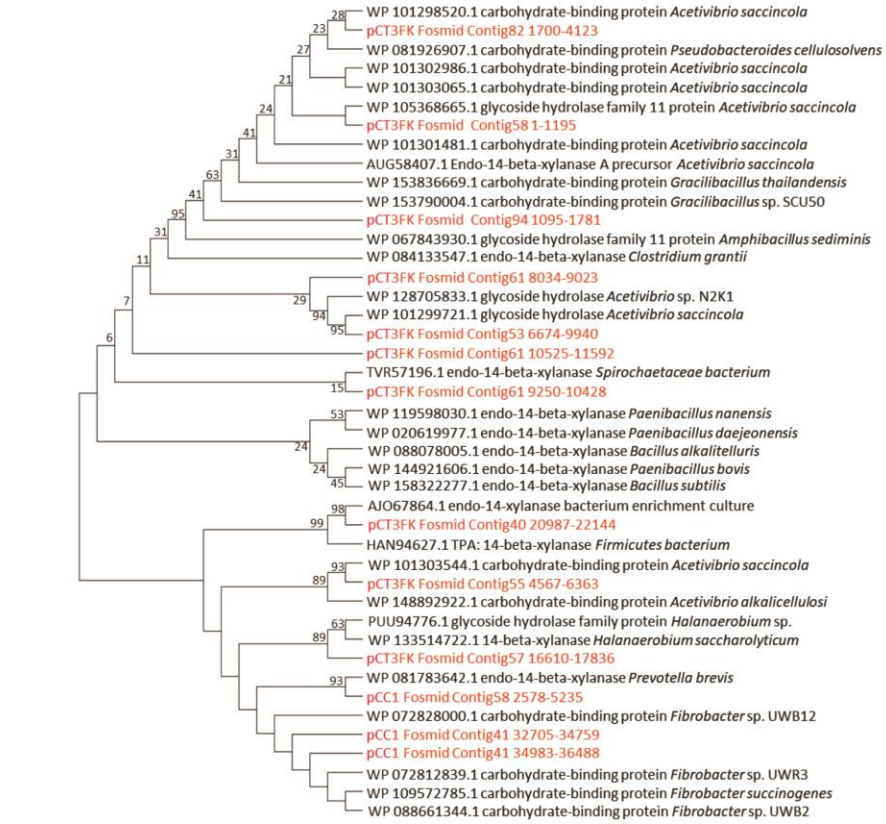
analysis tools, in total 48 ORFs on were annotated. In addition to the six identified and one inactive arabinosyl hydrolases, genes encoding two potential CAZymes from the GH127 and GH27 families were discovered, which might be associated with arabinan degradation. Additionally, the assembled DNA fragment included 22 different genes that encoded sugar transport components, regulatory proteins and proteins involved in the pentose phosphate pathway, including ribulokinase, transaldolase, transketolase, fructose-6-phosphate aldolase, L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase (Figure 3.15, Table 6.6). Most ORFs showed over 99% nucleotide sequence similarities with proteins from *X. thermophila* (Figure 3.15, Table 6.6). As a result, we suggested that this assembled metagenomic DNA fragment, which was recovered from the metagenome of a lab-scale biogas fermenter fed with maize silage, came from *X. thermophila* or a similarly related bacterium. The findings implied that the PPP plays an important role in *X. thermophila*'s pentose consumption.

3.2.15 Analysis of the putative hemicellulases from metagenomic library screening

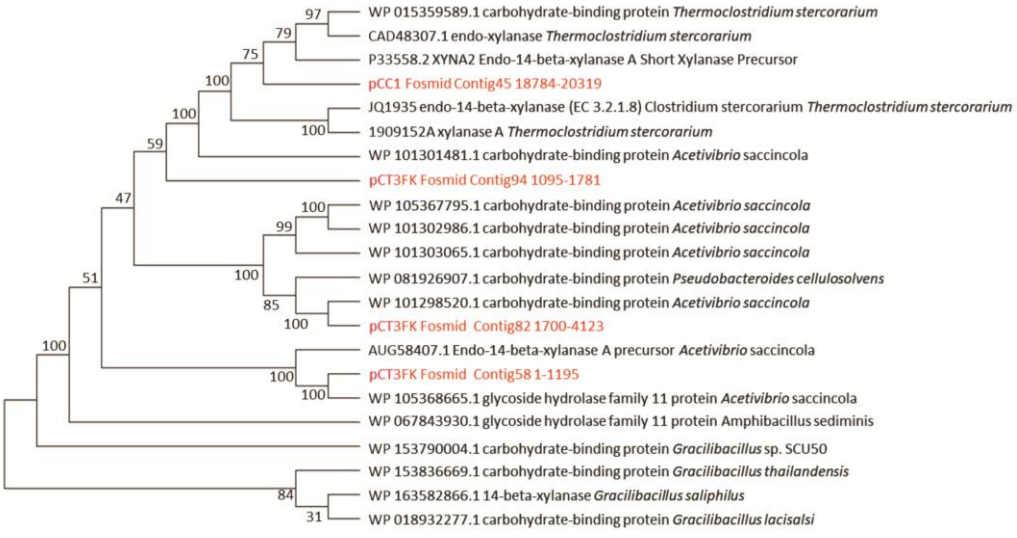
3.2.15.1 Phylogenetic relationship of the putative xylanase

Except for arabinofuranosidase activities, 18 fosmid clones with xylanase/arabinoxylanase activities were also identified from metagenomic library screening (Table 3.6). Further shotgun sequencing of the corresponding fosmid inserts revealed the variety of genes encoding putative xylanases from GHs family 10 and 11, including 13 putative GH10- and 4 putative GH11-encoding genes. The amino acid sequences deduced from the genes were compared to those of their top 2 to 6 BLAST hits in the NCBI-nr database by multiple sequence alignments in MEGAX conducted with the MUSCLE method. The analysis results showed that 13 putative endo-xylanase genes were identified from pCT3FK fosmid inserts, which were mainly screened from a metagenomic library constructed with the microbial communities from elephant feces (Figure 3.24a,b, Table 6.5). By contrast, 4 endo-xylanase encoding genes were derived from the pCC1 fosmid library constructed with SBP enriched fermenter microbial communities (Figure 3.24a,b, Table 6.5). Phylogenetic analysis of the diversity of all these putative hydrolases showed that most of the putative endo-xylanases from the pCT3FK library shared the highest sequence similarity with proteins from *Acetivibrio saccincola*. In contrast, the only putative GH11 xylanase from the pCC1 fosmid library showed a high amino acid sequence similarity with the endo-xylanase from *Thermoclostridium stercorarium (C. stercorarium)*, while three putative GH10 endo-xylanases from the same metagenomic library exhibited the highest identity with different

carbohydrate-binding proteins from *Fibrobacter*, such as WP 072828000.1 and WP 072812839.1 (Figure 3.24).



(a)



(b)

Figure 3.24 Phylogenetic analysis of putative GH10 and GH11 xylanase sequences identified from metagenomic library screening. Multiple sequence alignments were performed with the algorithm MUSCLE against full length sequences of GH10 or GH11 enzymes with each of their top 2 to 6 BLAST hits in the NCBI-nr database. Then a Neighbor-Joining (NJ) tree was calculated in MEGAX with bootstrap values from 500 replicates. Proteins from this study are colored in red, those from other sources in black. Database accession numbers and organisms are shown for each sequence.

3.2.15.2 A GH10 xylanase (M_Xyn10) with unique catalytic properties

TLC analysis of the hydrolysis products of beechwood xylan (BWX) by the xylanase-active fosmid clones revealed different cleavage patterns. Among all of the xylanolytic activities, the fosmid clone named "BM1T-2 H7" drew our attention due to the unique hydrolysis products from TLC analysis, comprising xylose, xylobiose and an unidentified oligosaccharide (Figure 3.26a,c).

By aligning the Sanger end sequences to contigs generated from Illumina shotgun sequencing, contig 40, with the size of 36,903bps, was assigned to fosmid clone BM1T-2 H7. All the putative CAZymes on contig40, according to dbCAN annotation, including a GH67, GH3, GH10 and two GH43 (Figure 3.25b). Since family GH10 is a prominent CAZyme family that includes endo-xylanase activity, the single putative GH10 on this contig (Designated M_Xyn10) is most likely related to the xylanolytic phenotype of the fosmid clone BM1T-2 H7. Another arabinoxylan with known structure, namely wheat flour arabinoxylan (WAX), was also applied to our study, the hydrolysis products, including the xylose, xylobiose and two unidentified oligosaccharides (Figure 3.25d). According to the substrate structure, BWX and WAX have a backbone of 1,4-linked β-D-xylopyranosyl residues. The former has the feature with MeGlcA side-chain substitution, while WAX has the arabinosyl residue decoration. Therefore, the cleavage specificity of the fosmid clone "BM1T-2 H7" is differed based on the structure of xylan substrates adopted, but with the common feature of hydrolysis products comprising xylose, xylobiose and some unidentified oligosaccharide(s).

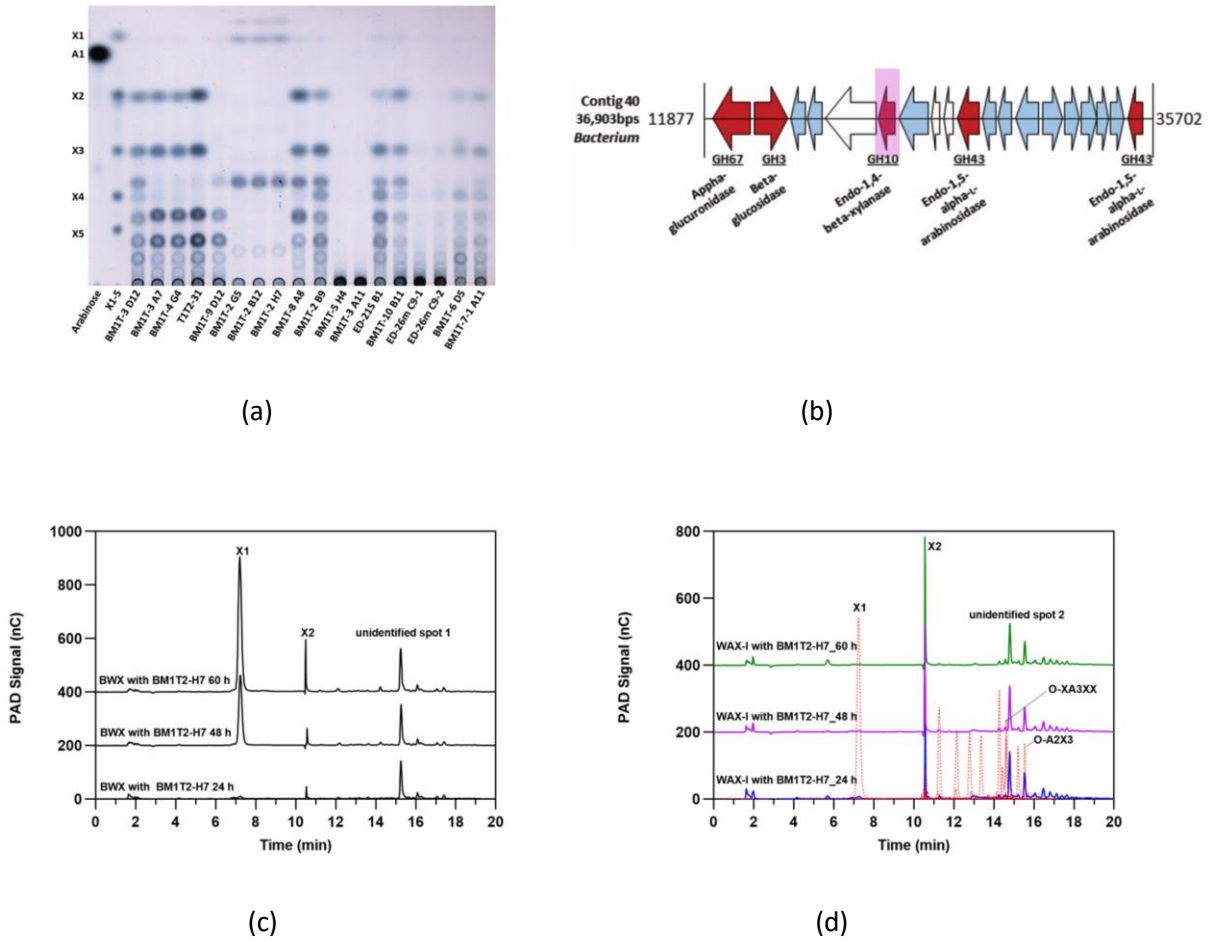


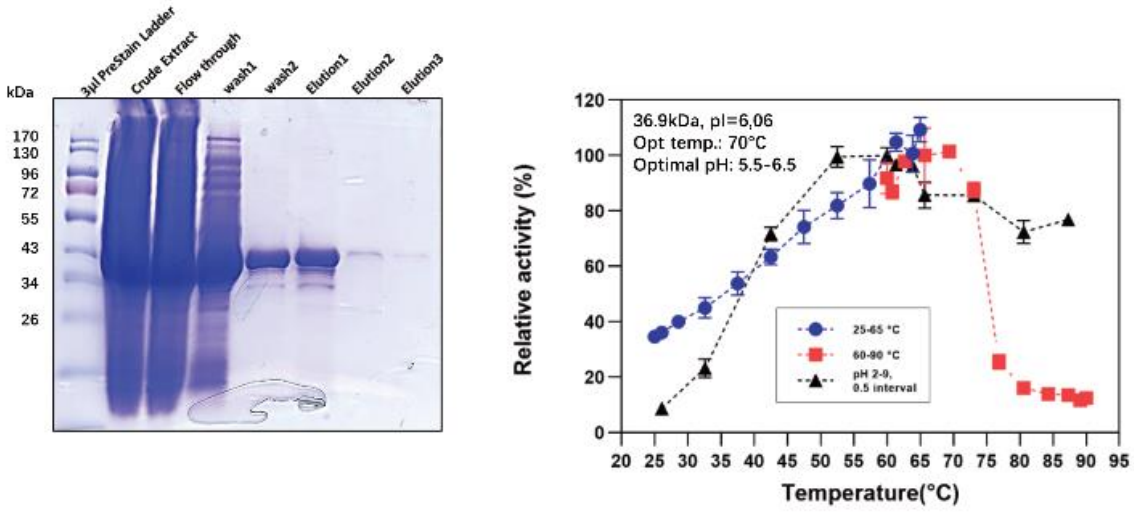
Figure 3.25 The activity of fosmid clone BM1T-2 H7 against various xylan substrates and the gene organization on its clone inserts (a) TLC analysis of the hydrolysis products released from BWX by the crude extract of each positive fosmid clone. (b) Gene organization of the putative CAZyme gene cluster (CGC) on contig 40 (functional annotation was performed by KEGG database and dbCAN database), red color represents the putative CAZymes, M_Xyn10 was marked by a purple square. (c-d) HPAEC-PAD analysis of hydrolysis products of BWX and WAX-I by crude extract of fosmid clone BM1T-H7, respectively. (40 µL cell crude extract was incubated with 0.5% BWX/WAX for 24 h, 48 h and 60 h (c, d) or for 48 h (a) in a 200 µL of enzyme reaction.

3.2.15.3 Cloning, Expression and Purification of M_Xyn10

The target gene encoding M_Xyn10 was amplified with PCR from the fosmid clone BM1T2-H7 with the primers listed in the Material and methods section using Q5® High-Fidelity DNA Polymerase. The resulting PCR product was 1,078 bps, as expected. The amplicon was further cloned into Nde I/Xho I digested pET24c vector by Gibson assembly after the PCR products were purified. Successfully cloned plasmids were confirmed by restriction digestion and Sanger sequencing, which were then transformed into *E. coli* BL21 (DE3). Recombinant cells were grown in LB medium with 50 µg mL^{-1} kanamycin for selection and expression of the recombinant protein was induced overnight at 30°C at 180 rpm with 0.05 mM IPTG. The target protein was extracted by IMAC with a Nickel column. The presence and purity

of the recombinant protein with the size of 41 kDa were detected by SDS-PAGE (Figure 3.26a). Interestingly, the recombinant protein was overexpressed even without IPTG induction, which was probably caused by leaky promoter activity.

To find out the best working condition of the M_Xyn10, the best temperature and pH condition of the enzyme were determined with WAX-RS as substrate. The results showed M_Xyn10 has an optimal temperature of 70°C and optimal pH between 5.5-6.5 (Figure 3.26b). Based on its optimal reaction conditions, further characterization focused on clarifying the cleavage pattern and catalytic mechanism of M_Xyn10 against different xylan, xylooligosaccharide and *p*NP-substrates.



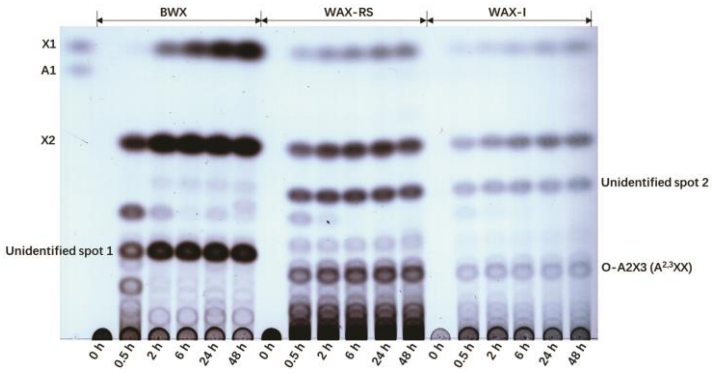
(a) (b)

Figure 3.26 Recombinant M_Xyn10 purification and its enzymatic characterization. (a) SDS-PAGE gel shows overexpression and purification of M_Xyn10. The protein was purified by IMAC with a Ni-TED packed column. Lane 1: Prestained protein standards; Lane 2: Crude cell extract prepared using sonification method; Lane 3: Flow-through; Lane 4,5: Two times washing fraction from the Ni-TED packed column; Lane 6-8: Elution fraction of purified sample. (b) Optimal temperature and pH of purified M_Xyn10. The temperature for maximum activity was determined by incubating 0.5% BWX with 168.78 nM M_Xyn10 at pH 6.0 at different temperatures in a range between 25-90°C for 1 h, before determining the reducing sugar released from the reaction with the DNS assay. The optimal pH was determined by using the same reaction but employing various pH values within a pH range between 2-9 with 25 mM citrate phosphate buffer at the temperature of maximum activity).

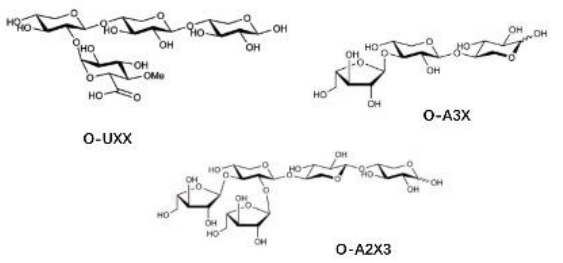
3.2.15.4 Hydrolysis of different xylan polymers by M_Xyn10

Hydrolysis products produced by fosmid clone BM1T-2 H7 from various xylan substrates seem to be specific. However, using the clone's crude extract only reflects the combined activity of all the xylanolytic enzymes encoded by the genes on the clone's recombinant fosmid insert. For cleavage mode assessment and to precisely evaluate the specific enzyme(s) responsible for the fosmid clone's

activity, M_Xyn10 was chosen as the most likely candidate to show xylanase activity. After incubating purified M_Xyn10 with WAX and BWX, the hydrolysis products were further analyzed by TLC. The results showed various oligosaccharides released from xylan substrates initially, which can be further degraded into 3 or 4 different oligosaccharides after 2 h incubation (Figure 3.27). The activity of M_Xyn10 was nearly identical to that of the whole fosmid clone. The unidentified compounds released from BWX and WAX were determined using a method that combined the TLC and MALDI time-of-flight (ToF) mass spectrometry, which showed that O-UXX with a molecular mass of 604.29 (unidentified spot 1) was the unknown hydrolysis product from BWX (Figure 3.28c). The hydrolysis product from WAX, on the other hand, was A^3X and $A^{2,3}XX$. The former compound has a molecular mass of 414, and the latter matched the standard by HPAEC-PAD (Figure 3.29c). As a result, it is hypothesized that M_Xyn10 might cleave xylan substrates into branched XOS with the side-chain substitutions exposed terminally, accompanied by the production of xylose and xylobiose.



(a)



(b)

Figure 3.27 Activity of purified M_Xyn10 against different xylan substrates. (a) Time course analysis of hydrolysis products released from beechwood xylan (BWX), wheat arabinoxylan for reducing sugar assay (WAX-RS) and insoluble wheat arabinoxylan (WAX-I) by M_Xyn10. The reaction was performed by incubating 168.78 nM M_Xyn10 with 0.5% BWX, WAX-RS and WAX-I at pH 6.0 and 70°C for 0 h, 0.5 h, 2 h, 6 h, 24 h, 48 h, 10 µL of the hydrolysis products were separated three times on TLC plates using a solvent mixture of chloroform, acetate and water in a ratio of 6:7:1 (v/v/v).

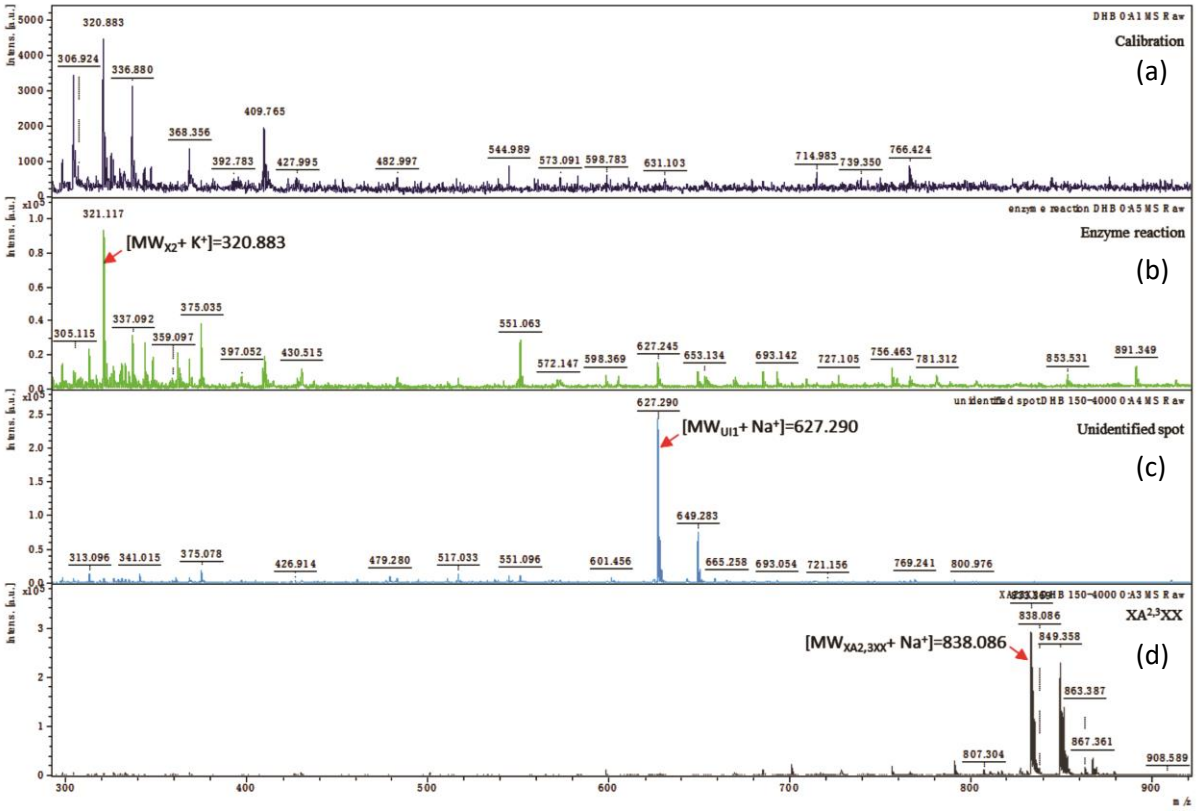


Figure 3.28 MALDI-TOF mass spectral profiles of unidentified oligosaccharides in the hydrolysis products of BWX. (a) Blank control only contains calibration buffer, (b) All the hydrolysis products from enzyme reaction, (c) An unidentified oligosaccharide (designated UI1) isolated from the TLC silica plate, (d) Positive control XA^{2,3}XX. For each spectrum, 0.5 µL of the sample was directly deposited onto the 0.5 µL DHB layer after it dried. The enzyme reaction was performed by incubating 168.78 nM M_Xyn10 with BWX at its optimal pH and temperature for 24 h. For isolation of unidentified oligosaccharide, the hydrolysis products in enzyme reaction from (c) was separated on a TLC plate in the solvent system including chloroform, acetate and water with the ratio of 6:7: 1(v/v/v) for three times, the spot on the plate was scraped and dissolved in 200 µL water, after centrifugation, the supernatant was further concentrated to around 20 µL.

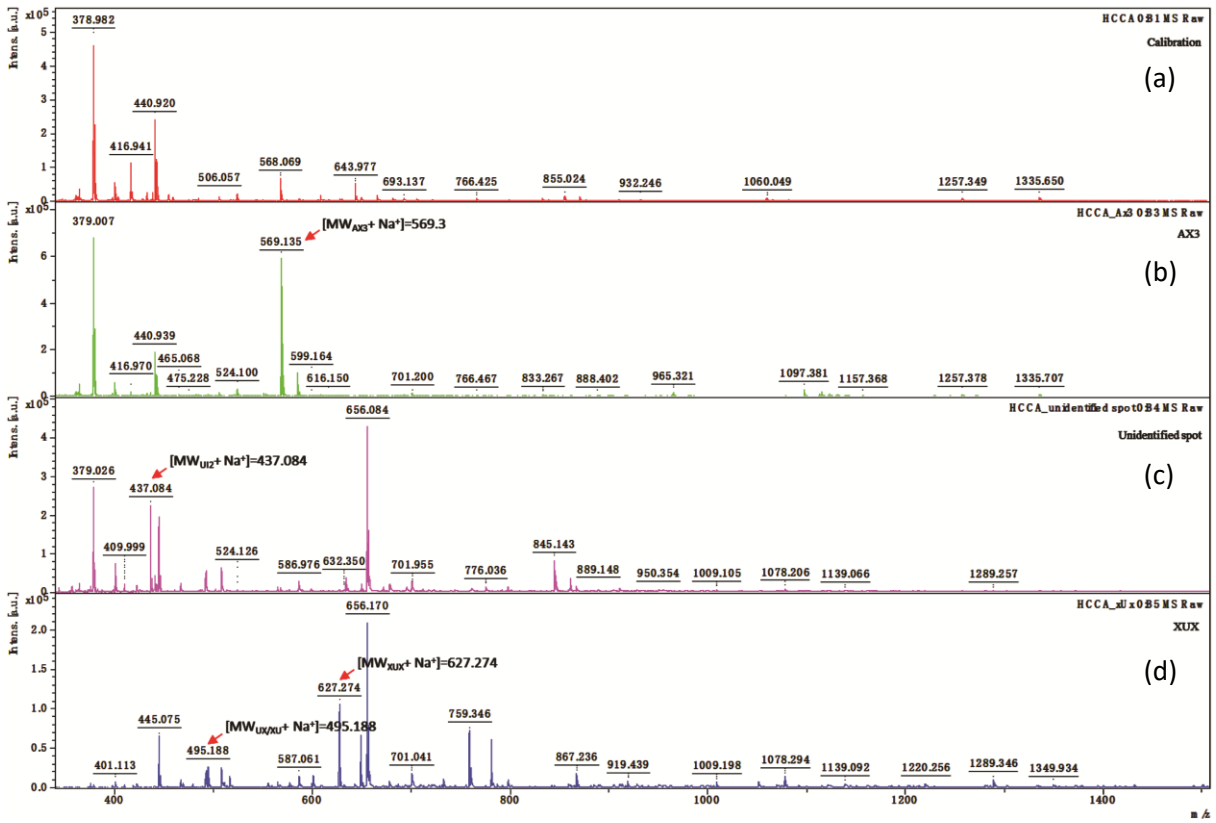


Figure 3.29 MALDI-TOF mass spectral profiles of unidentified oligosaccharides in the hydrolysis products of WAX and XUX. (a) Blank control only contains calibration buffer, (b) Positive control AX3 (c) An unidentified oligosaccharide (designated UI2) isolated from the TLC silica plate, (d) Enzyme reaction with XUX as substrate. For each spectrum, 0.5 µL of the sample was directly deposited onto the 0.5 µL HCCA layer after it dried. The enzyme reaction was performed by incubating 168.78 nM M_Xyn10 with 0.5% BWX or 0.25% XUX at its best working condition for 24 h. For isolation of unidentified oligosaccharide, the hydrolysis products in enzyme reaction from was separated on a TLC plate in the solvent system including chloroform, acetate and water with the ratio of 6:7: 1(v/v/v) for three times, the spot on the plate was scraped and dissolved in 200 µL water, after centrifugation, the supernatant was further concentrated into around 20 µL.

3.2.15.5 Hydrolysis of various XOS by M_Xyn10

With WAX and BWX as substrates, the presence of xylose in the hydrolysis products further suggested that this enzyme has both endo- and exo-activity. Different XOS, including MeGlcA decorated XOS, linear XOS and arabinosyl decorated XOS, were used to evaluate the cleavage mode of M_Xyn10. With X2, O-UXX, O-A3X, O-A2X3 and O-A3X as substrates, M_Xyn10 did not show activity, which was consistent with the findings that four of these oligosaccharides were the end products released from xylan substrates and thus cannot be further degraded by the enzyme.

However, when O-XUX was used as a substrate, which is a 4-O-methyl-α-D-glucuronyl (MeGlcA) decorated XOS with the same degree of polymerization (DP) as O-UXX, the substrate was further degraded into two unidentified oligosaccharides (Figure 3.30a). However, MALDI-TOF analysis of the

hydrolysis products only detected the presence of compounds with the molecular mass of 604.5 and 472.4 (Figure 3.29d). The former compound can be assigned to O-XUX, which is the substrate used in the enzyme reaction, while the latter matched the molecular mass of a MeGlcA decorated xylobiose. Since another possible oligosaccharide was not detected in MALDI-TOF's analysis, the results suggested that O-XUX was degraded into O-XU and O-UX (same molecular mass of 472.4), i.e., M_Xyn10 apparently cleaves off xylose from the substrate from both the reducing and the non-reducing ends. The release of X2 from $XA^{3}XX$ and $XA^{2,3}XX$ further proves such a conclusion (Figure 3.30b).

Furthermore, the release of $A^{2,3}XX$ and $A^{3}X$ from $XA^{2,3}XX$ and $XA^{3}XX$, respectively, was consistent with the result we reached using WAX as a substrate and hydrolysis products containing a single or double substituted arabinofuranosyl side chain at the terminal non-reducing end (Figure 3.30b). It is noteworthy that $XA^{3}XX$ and $XA^{2}XX$ contain a single arabinosyl side chain at the O-2 and O-3 positions, respectively. However, the hydrolysis products of these two oligosaccharides were distinct. They were degraded into $A^{2}XX$ and $A^{3}X$, respectively (Figure 3.30b), indicating the same sugar substitution at different positions may result in a different steric hindrance, a substitution at the O-2 position may be less able to approach the cleavage site of the enzyme.

For the linear xylooligosaccharides, after 24 h incubation, the hydrolysis products in the reaction only included xylobiose and a small amount of xylose. It was unlikely the enzyme only cleaved every other glycosidic bond on the backbone. If this were the case, then X4 would only be degraded into X2 since X2 cannot be further degraded. The presence of xylose in the hydrolysis products, on the other hand, suggested that M_Xyn10 may bind X4 (and larger oligosaccharides) in different ways. It seems at least a total of three xylose-binding subsites (the smalles substrate is xylotriose) and that at least two xylose-binding subsites next to the catalytic center must be occupied for hydrolysis to occur. This mode of action explains all product patterns observed with the xylo-oligosaccharides, i.e. that X3 as well as X4 cleavage results in the hydrolysis products X2 and X1 (Figure 3.30c).

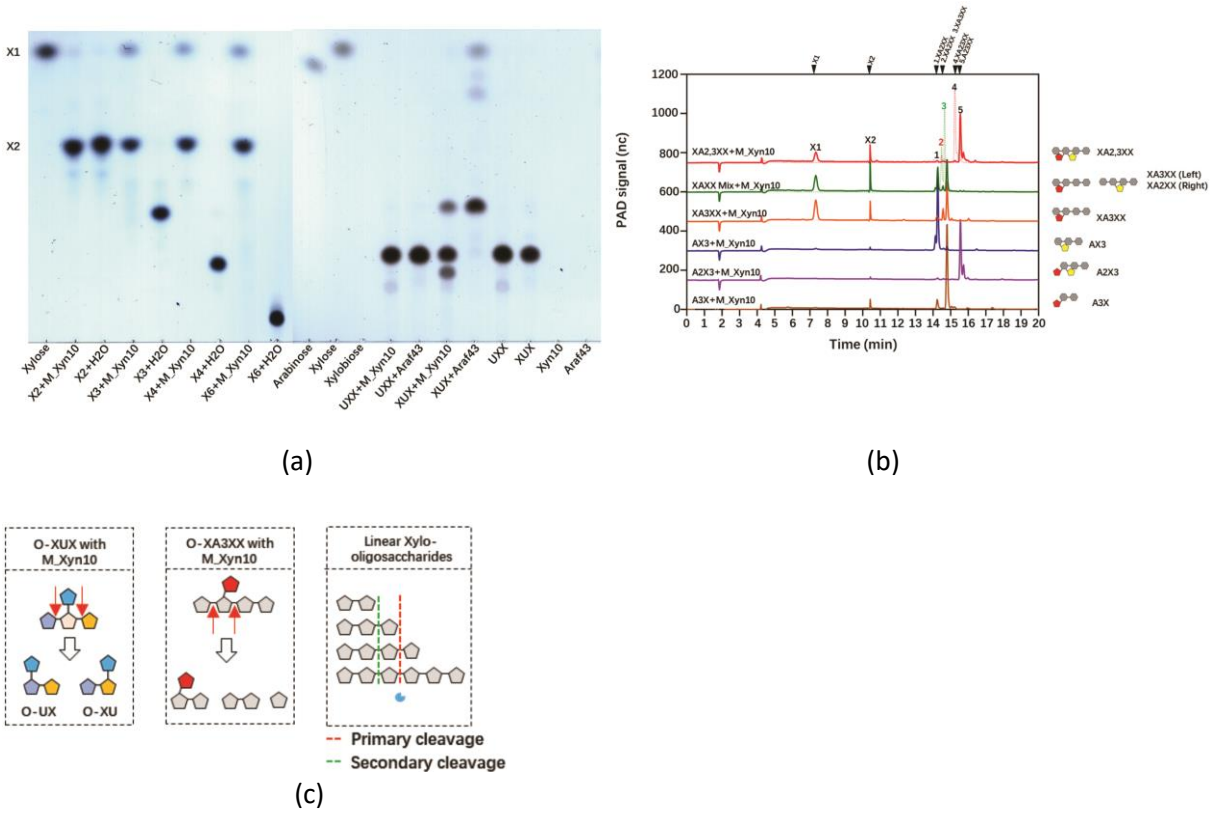
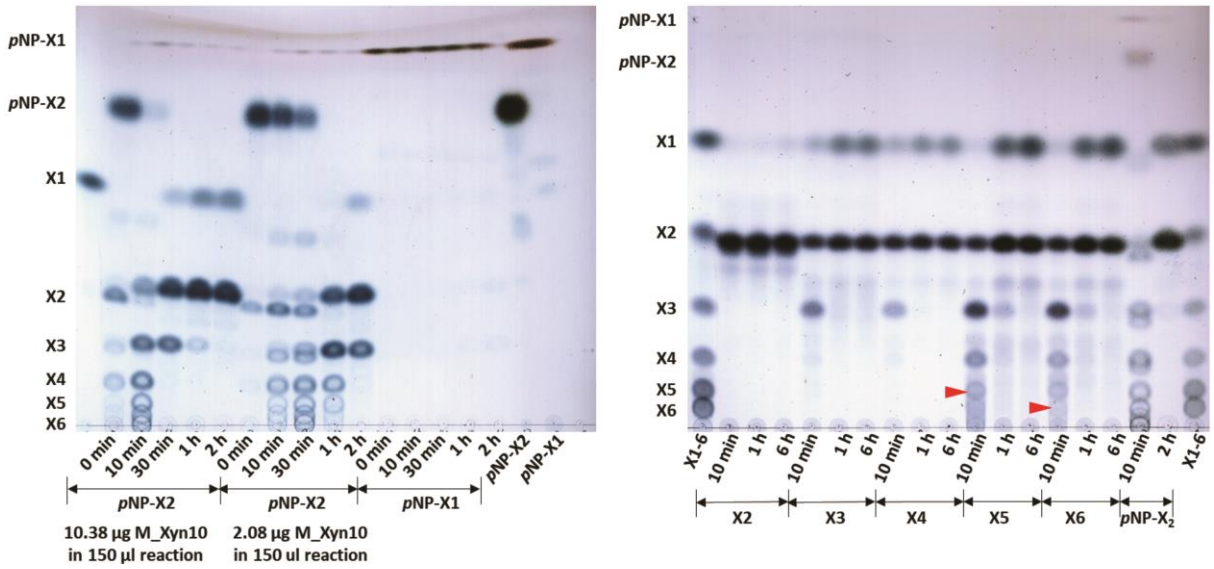


Figure 3.30 The activity of the purified M_Xyn10 against different XOS. (a) TLC analysis the hydrolysis products released from MeGlcA decorated XOS (UXX and XUX) and linear XOS (DP from 2 to 6) by M_Xyn10. (b) HPAEC-PAD analysis the hydrolysis products released from arabinosyl decorated XOS by M_Xyn10. (c) Putative cleavage mode of M_Xyn10 on O-XUX, O-XA3XX and linear xylooligosaccharides. The standard reaction was performed by incubating 168.78 nM M_Xyn10 with 0.25% oligosaccharides at 50°C, pH 6.0 for 24 h. 2.5 µL of the hydrolysis products were separated three times on TLC plates using a solvent mixture of chloroform, acetate and water in a ratio of 6:7:1 (v/v/v).

3.2.15.6 Hydrolysis of *p*NP-substrates by M_Xyn10

The activity of M_Xyn10 was also explored with *p*NP substrates. *p*NP-X2 (4-Nitrophenyl-β-xylobioside) was efficiently cleaved into *p*NP-X1 (4-Nitrophenyl-β-xylopyranoside), X2 and X1, but no activity could be observed with *p*NP-X1 as a substrate (Figure 3.31a). Furthermore, at a low concentration of M_Xyn10 in the reaction, transglycosylation activity of M_Xyn10 could be observed. Transglycosylation activity of xylanase has been studied previously, which could be useful to synthesize various oligosaccharides. However, different xylanases have varying transxylosylation ability with various acceptors. In this study, *p*NP-X1 or *p*NP-X2 served as an acceptor, while xylose was a donor to develop various *p*NP-xylooligosaccharides with different degrees of polymerization, and nitrophenol was cleaved off by the exo-activity of M_Xyn10, leading to corresponding linear XOS which in turn can be digested into xylobiose and xylose. Nevertheless, with *p*NP-X1 as a substrate, the enzyme did not

exhibit any transglycosylation activity, probably because *p*NP-X1 cannot act as a donor substrate. Xylobiose was not likely to be a donor since no synthesized XOS were generated when X1 and X2 were incubated with M_Xyn10. The transglycosylation ability of the enzyme was also tested with linear XOS. The TLC results showed a weak transglycosylation activity towards X5 and X6. However, no clear results could be obtained with X2, X3 and X4, although weak X4 spot could be observed after 10 min with X3 as the substrate (Figure 3.31b).



(a) (b)

Figure 3.31 The transglycosylation activity of M_Xyn10 against different *p*NP substrates and XOS. (a) Time-course analysis of the hydrolysis products released from *p*NP-xylobioside and *p*NP-xylopyranoside and (b) from linear XOS by M_Xyn10. Standard reaction was performed by incubating 1.69 µM or 338.21 nM M_Xyn10 with 1 mM *p*NP substrate at RT for 10 min or by incubating 168.78 nM M_Xyn10 with 0.25% oligosaccharides at 50°C, pH 6.0 for 24 h. 20 µL of the hydrolysis products from *p*NP assay or 2.5 µL of the hydrolysis products from reactions with XOS as substrate were separated three times on TLC plates using a solvent mixture of chloroform, acetate and water in a ratio of 6:7:1 (v/v/v).

3.3. Characterization of microbial communities from biogas fermenter

3.3.1 Enrichment of microbial community on cellulose, hemicellulose and SBP

In addition to isolating single bacterial strains with polysaccharide-degrading ability from nature/manufactured habitats and mining the CAZyme encoding genes by constructing metagenomic libraries, microbe selection by enrichment on plant biomass has been described as a viable method for developing novel lignocellulose saccharification methods. Because polysaccharide pools in plant cell walls are variable regarding polysaccharide composition and chemical structure, the breakdown is

generally aided by mixed bacterial populations rather than a particular strain. To identify the primary bacterial community involved in the breakdown of SBP plant biomass and analyze the dynamic changes in the bacterial population comparing plant biomass and pure polysaccharides, a two-month enrichment culture experiment was set up in the laboratory with microbial communities from a biogas fermenter at the Bayerische Landesanstalt für Landwirtschaft (LfL) that was fed with sugar beet silage (SBS). Since the major fraction in sugar beet silage includes pectin, cellulose and hemicellulose, a broad repertoire of pectin-, cellulose- and hemicellulose- digesting microbes were expected to be present and enriched under appropriate conditions.

The enrichment processes were established as described in the Material and methods. For data analysis, the original sample from the biogas fermenter was designated "O." The enriched samples were named "A," "B," "C,"... "F," consecutively, based on the times of enrichment. To make it easier to understand the results in the post-sequencing analysis and to better differentiate the samples that were sent for 16S rRNA amplicon sequencing, sugar beet pulp, filter paper, and xylan were symbolized by capital letters "S," "FP," and "X," respectively. Individual series of the triplicate experiment are denoted by the Arabic numbers "1", "2" and "3" (see workflow in Fig. 3.32).

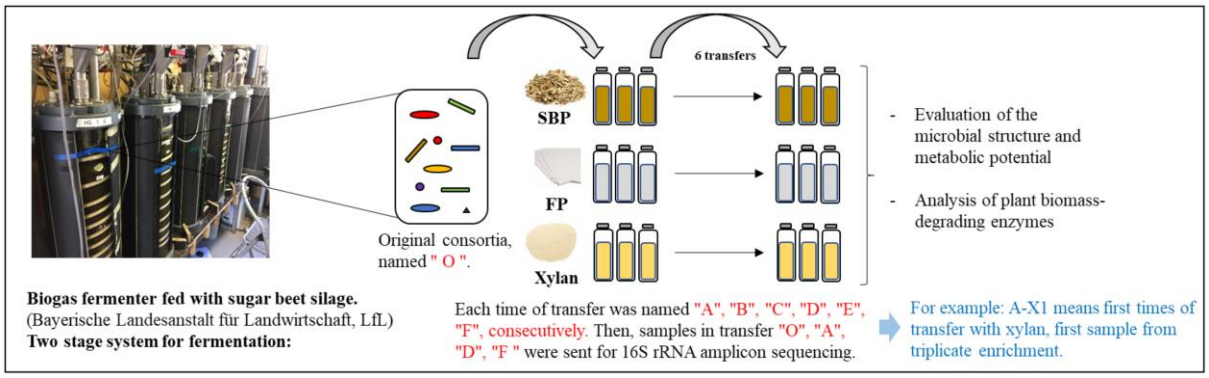


Figure 3.32 Schematic representation of the enrichment strategy. SBP: Sugar beet pulp, FP: Filter paper; "A", "B", "C"... "F" represent the times of transfer; "O" means original sample from fermenter. The enrichment processes were performed triplicates, but only parallel samples were sent for 16S rRNA amplicon sequencing and named by the Arabic numbers "1" and "2."

3.3.2 Hydrolysis of chromogenic polysaccharides by the secreted enzymes

Using chromogenic substrates, including self-made CPH-arabinan, CPH-xylan, CPH-arabinoxylan, CPH-CMC, ICB-HM pectin, ICB-SBP, CPH-avicel, CPH-rhamnogalacturonan and commercial AZCL-cellulose,

AZCL-β-glucan, the ability of the secreted enzymes from the selected microbial consortia to break down plant biomass or polysaccharides following six consecutive transfers (‘’F’’) was evaluated. The standard assay was performed by incubating an equal amount of secreted protein (from transfer “F”), quantified by Bradford assay, with different labeled substrates (mainly test endo-active enzymes). After overnight incubation at 37°C, the supernatant of each reaction was obtained by centrifugation and transferred into a new microtiter plate for measuring the absorbance values under differentwavelengths. All assays were performed in triplicate.

On diverse CPH/ICB substrates, the enzymatic activities of supernatants from SBP and xylan enrichment cultures were similar (except for CPH-arabinan), which may point to similar microbial community structures. The secreted enzymes from SBP and xylan enrichment cultures displayed maximum activity when CPH-arabinoxylan and CPH-xylan were used as substrates (Figure 3.33b,c). Moreover, both secreted protein fractions from enrichment cultures also exhibited weak activity against CPH-CMC, ICB-SBP and AZCL-β glucan (Figure 3.33 d,g,i). In addition, it is noteworthy that the enzymatic activity of the metasecretomes from the SBP enrichment culture specifically showed high activity against CPH-arabinan. The filter paper consortial metasecretome, on the other hand, showed low or no activity against most of the chromogenic substrates tested, but on a few substrates, such as CPH-arabinoxylan and CPH-CMC, there was low activity (Figure 3.33 c,d).

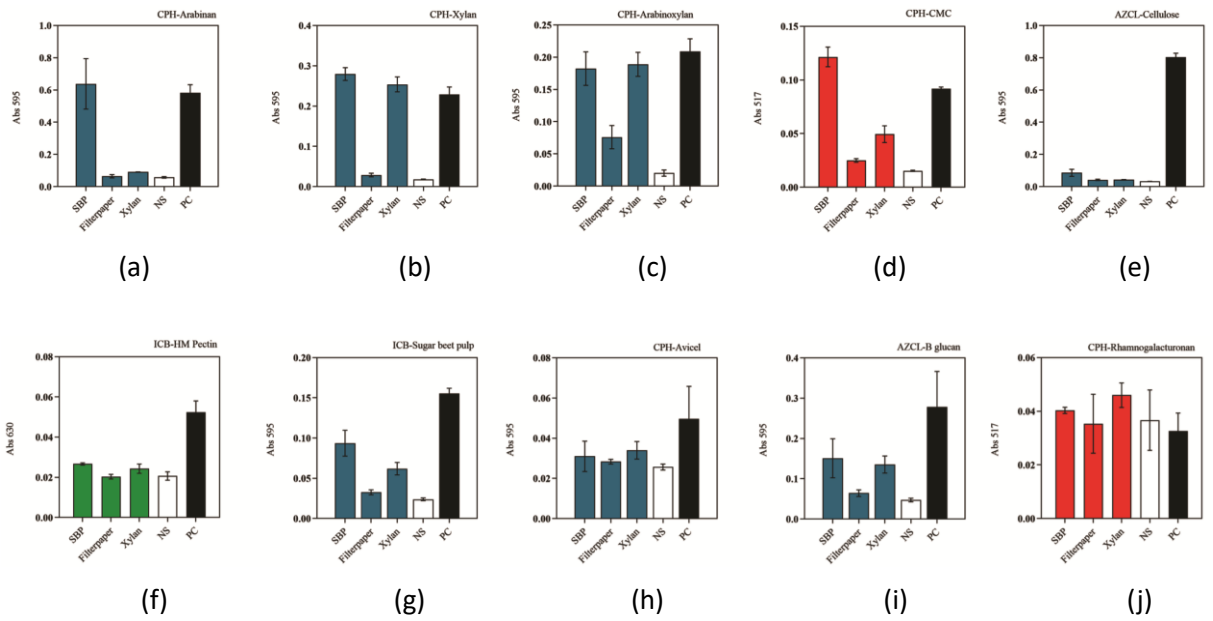


Figure 3.33 Absorbance values obtained from enzyme reactions by incubating chromogenic substrates with supernatant proteins from enrichment cultures. The enzyme assays were performed by incubating the supernatants of the enrichment cultures (normalized to contain equal amounts of secreted protein) with chromogenic substrates at 37°C overnight. The concentration of the secreted protein was measured via Bradford assay. Positive controls and negative controls were set up using commercial enzymes from Megazyme (1 U mL^{-1} enzyme from Megazyme or 10 µL 1:5 diluted viscozyme) or ddH_2O instead of supernatant of enrichment culture, respectively. The enzyme reactions contain a 50 µL chromogenic substrate, 150 µL 50 mM phosphate buffer (pH=7.0) and a 10 µL supernatant adjusted to the same total protein concentration.

3.3.3 Alpha-diversity of the enriched microbial communities

When the parallel samples of the 16S rRNA gene amplicons from the enrichment cultures were sequenced, many OTUs were obtained in the seed samples (3,723), while considerably fewer were detected after a long-term enrichment process (328-3,477) (Table 3.12). The rarefaction curves suggested that mostly the sequencing depth was sufficient to reasonably capture the diversity of the microbial community, except for five under-sampled cases, including A-S1, A-FP1, F-FP2, D-FP1, F-X2 (Figure 3.34). Nevertheless, these data were used for further analysis. If there were substantial variations between duplicate samples, another sample was sent for 16S rRNA amplicon sequencing to obtain a repeatable result.

Table 3.12 Number of reads and observed species in 16S rRNA gene amplicon sequencing.

Sample name	Number of reads	Number of OTUs
Filter paper		
Seed sample	6,467	3,723
A-FP1	8,377	1,812
A-FP2	9,565	2,990
D-FP1	6,775	1,122
D-FP2	8,066	2,461
F-FP1	6,791	3,477
F-FP2	6,407	1,543
Sugar beet pulp		
Seed sample	6,467	3,723
A-S1	6,121	502
A-S2	5,826	1,411
D-S1	13,127	1,070
D-S2	8,083	1,557
F-S1	6,387	2,464
F-S2	9,296	1,767
Xylan		
Seed sample	6,467	3,723
A-X1	8,635	1,171
A-X2	9,931	328
D-X1	7,436	974
D-X2	7,686	1,342
F-X1	9,543	1,762
F-X2	2,005	2,097

"A," "D," "F" represented the times of transfers, The enrichment processes were performed triplicates, but only parallel samples were sent for 16S rRNA amplicon sequencing and named by the Arabic numbers "1" and "2"; "S" represented sugar beet pulp enrichments, "X" represented xylan enrichments, "FP" represented filter paper enrichments.

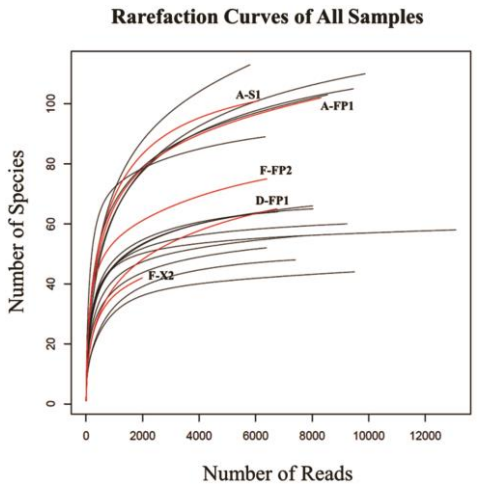


Figure 3.34 Rarefaction curves of all samples. The x-axis represented the number of sequences sampled, while the y-axis represented a measure of the species richness detected, the red color curves represented the under-sampled cases. "A," "D," "F" represented the times of transfers; The enrichment processes were performed triplicates, but only parallel samples were sent for 16S rRNA amplicon sequencing and named by the Arabic numbers "1" and "2"; "S" represented sugar beet pulp enrichment, "X" represented xylan enrichment, "FP" represented filter paper enrichment.

Four α-diversity parameters (Richness, Evenness, Shannon, Simpson) were calculated to explore further the outcome of enrichment on the microbial community within the sample. Richness and evenness reflected the quantity of detected OTUs and the distribution of single OTUs in the sample. Therefore, we concluded that the amount and distribution of bacteria in the bacterial communities decreased significantly after around two months of enrichment across all samples (Figure 3.35a,b). Shannon and Simpson's effective value calculates the microbial community diversities, which further demonstrated that the diversity of fermenter microbes was dramatically reduced when enriched on different carbon sources (Figure 3.35c,d). The decrease occurred mainly after the first transfer (A) but continued during the fourth (D) and sixth transfer (F) but to a smaller extent.

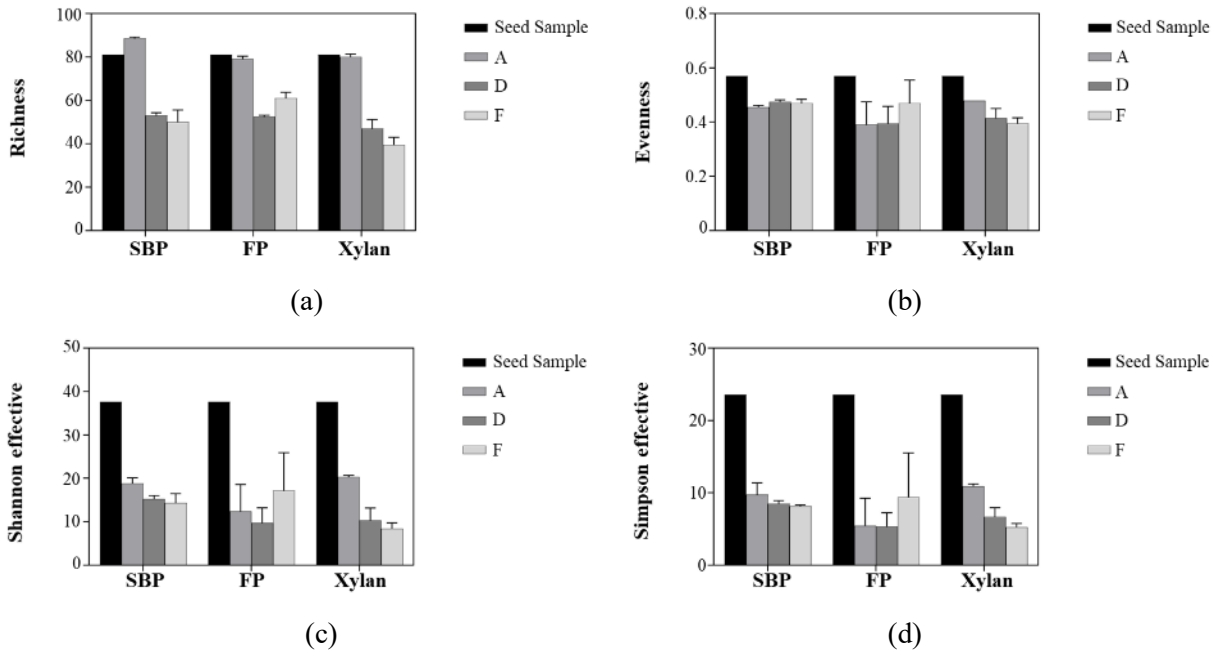


Figure 3.35 Histogram of alpha-diversity parameters for the bacterial communities. (a) Richness (b) Evenness (c) Shannon effective (d) Simpson effective of microbial communities. The alpha diversity of each sample was computed from duplicate biological samples. "A," "D," and "F" represented the times of transfers; The enrichment processes were performed triplicates, but only parallel samples were sent for 16S rRNA amplicon sequencing and named by the Arabic numbers "1" and "2"; "S" represented sugar beet pulp enrichment, "X" represented xylan enrichment, "FP" represented filter paper enrichment.

3.3.4 Beta-diversity of the enriched microbial communities

To explore if the different carbon sources (SBP, filter paper, xylan) can specifically shape a mature microbial community in a two-month enrichment process, β-diversity was calculated between samples.

The MDS plot of microbial profiles showed microbial communities significantly differed in the enrichment cultures than that of seed sample (p-value=0.001), which indicated that bacterial community was highly selected when enriched with different substrates, including SBP, xylan and filter paper (Figure 3.36a). We did not observe considerable changes in the selected bacterial consortia between xylan- and SBP-digesting microbial systems regarding the composition of the microbial community. They were very similar and even partially overlapped on the MDS plot (Figure 3.36a). In comparison, the cellulose-enriched microbial communities were significantly different from the xylan- and SBP- degrading bacterial systems (Figure 3.36a). Time-course analysis of the microbial community under each carbon source was also performed, but no significant difference between three different stages of enrichments (first, fourth, sixth transfer) with a p-value > 0.05 (Figure 3.36b,c,d). With SBP and xylan as substrates, the shaping of the microbial communities was largely completed after the fourth time of transfer (Figure 3.36b,c). The filter paper digesting microbial system was not stable during the enrichment processes compared with the SBP- and xylan- enrichments (Figure 3.36d).

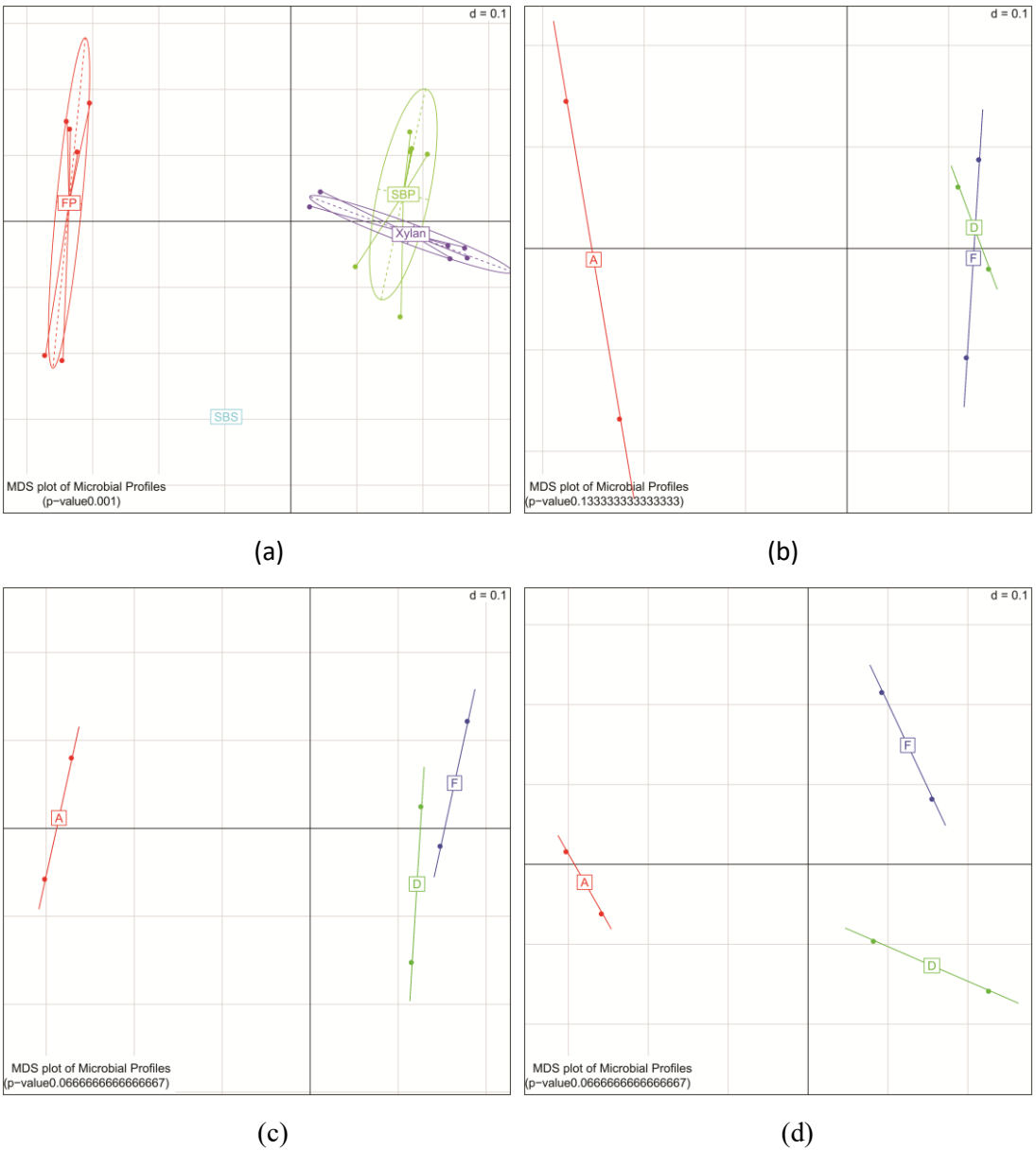


Figure 3.36 MDS plot of microbial profiles between samples. Two-month enrichment of microbial consortia on xylan, sugar beet pulp and filter paper were performed within six transfers. (a) Overview of the changes of bacterial communities in seed sample (SBS) and SBP-, xylan-, filter paper- enrichments; (b-d) The changes of bacterial communities in SBP- (b), xylan- (c), filter paper- (d) enrichments under different transfers, respectively. "SBS" represented the seed sample from biogas fermenter with sugar beet silage as substrate, "A," "D," and "F" represented the microbial community from first-, third- and sixth- time of transfers.

3.3.5 Taxonomic classification of OTUs

Regarding the composition of microbial communities, it was noted that the major phyla shifted to mainly *Firmicutes* and *Bacteroidetes* in the SBP-enriched microbial communities (61.80% *Firmicutes* and 36.82% *Bacteroidetes*) and xylan-enriched microbial communities (41.66% *Firmicutes* and 57.31% *Bacteroidetes*) after a two-month enrichment process (Figure 3.37, Table 3.13). By contrast,

these two phyla accounted for 49.94% (*Firmicutes*) and 37% (*Bacteroidetes*) of the whole bacterial community in the seed sample, respectively. In the filter paper-digesting microbial system, the bacterial community displayed a relatively high diversity which was composed of *Firmicutes* (55.67%), *Spirochaetes* (13.76%), *Synergistetes* (12.50%) and *Bacteroidetes* (10.29%) (Figure 3.37, Table 3.13). Interestingly, the *Fibrobacteres* were increased dramatically at the initial enrichment stage (first-time transfer), but after around one month of the enrichment process (fourth transfer) only 2.1% of *Fibrobacteres* were still present in the microbial system (Figure 3.37, Table 3.13). Therefore, *Fibrobacteres* may play an important role in the early stage of cellulose degradation.

As the major phyla in three different bacterial communities enriched after long-term propagation (six-time transfer) on three different carbon sources, *Bacteroidetes* and *Firmicutes* were further analyzed to unveil the bacterial composition at their family level. *Prevotellaceae* accounted for 49.23% of the bacterial community under the phylum *Bacteroidetes* in SBP enrichments, while at the family level *Prevotellaceae* and *Porphyromonadaceae* make up 31.96% and 12.76% of the taxa found in xylan enrichments (Figure 3.38a). As another major phylum in all the enriched bacterial consortia, *Firmicutes* included the dominant families *Lachnospiraceae* (49.56%) and *Ruminococcaceae* (22.11%) in the SBP-enriched microbial community. In contrast, the xylan-digesting microbial system had almost the reverse ratio of these two bacterial families, with 25.14% *Lachnospiraceae* and 65.39% *Ruminococcaceae*. However, in the filter paper-enriched bacterial community, *Lachnospiraceae* was the prominent family within the phylum *Firmicutes* and accounted for 75.74% of the bacterial community (Figure 3.38a).

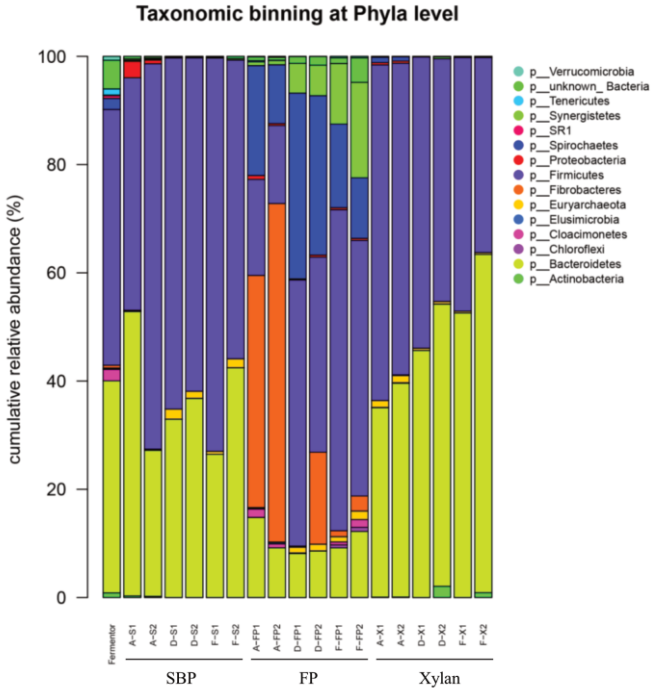
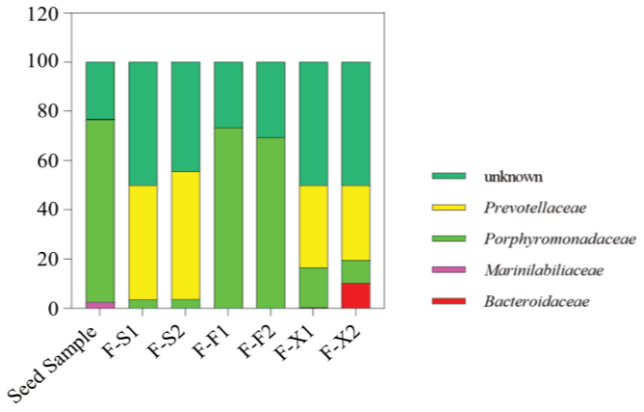


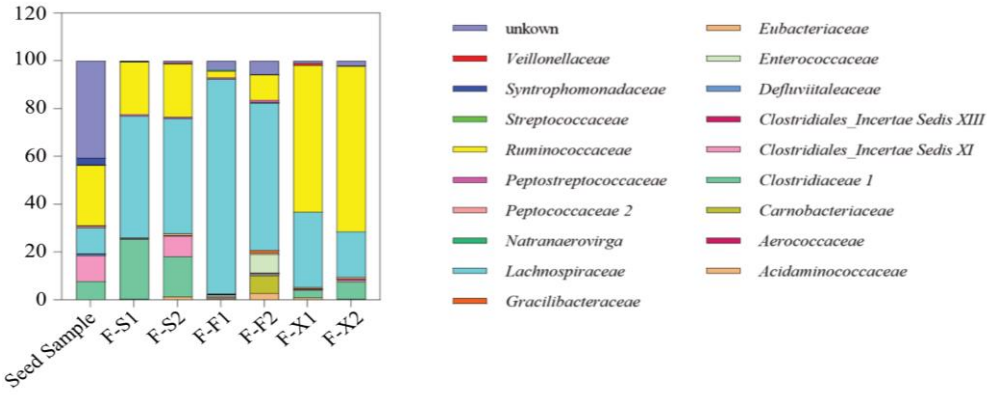
Figure 3.37 Taxonomic classification of OTU's at phyla level by 16S rRNA gene amplicon sequencing after two-month enrichment on different carbon sources.

Table 3.13 Taxonomic classification of OUTs at phyla level by 16S rRNA gene amplicon sequencing after two-month enrichment on different carbon sources.

Phyla	Original	SBP	FP	Xylan
Firmicutes	49.49%	61.80%	55.67%	41.66%
Bacteroidetes	37.00%	36.82%	10.29%	57.31%
Unknown	4.38%	0.30%	2.38%	0.05%
Cloacimonetes	2.27%	0.00%	1.07%	0.00%
Spirochaetes	1.77%	0.00%	13.76%	0.00%
Actinobacteria	1.46%	0.00%	0.01%	0.58%
Verrucomicrobia	1.05%	0.00%	0.19%	0.00%
Tenericutes	0.88%	0.00%	0.00%	0.00%
SRI	0.45%	0.00%	0.00%	0.00%
Fusobacteria	0.31%	0.00%	0.00%	0.00%
Fibrobacteres	0.25%	0.00%	2.10%	0.00%
Synergistetes	0.12%	0.10%	12.50%	0.07%
Elusimicrobia	0.10%	0.00%	0.00%	0.00%
Euryarchaeota	0.01%	0.98%	1.29%	0.30%
Proteobacteria	0.00%	0.00%	0.17%	0.04%
Chloroflexi	0.00%	0.00%	0.58%	0.00%



(a)



(b)

Figure 3.38 Percentage of major families within the phyla *Bacteroidetes* and *Firmicutes* in different enrichments. (The data was calculated by the number of the individual OTUs divided by the sum of the whole OTUs from either *Bacteroidetes* or *Firmicutes*).

4 Discussion

4.1 Sequence-based analysis of the isolated novel *Acetivibrio mesophilus* strain

4.1.1 Hemicellulolytic and pectinolytic potential of *Acetivibrio mesophilus*

Acetivibrio mesophilus was identified as an important cellulolytic degrader in thermophilic fermenter microbiota (Rettenmaier *et al*., 2019). With regard to hemicellulose substrates, it could also degrade WAX and OSX, but not BWX. When comparing the structural differences between these substrates, it is clear that they all have a repetitive beta-1,4-linked xylopyranosyl backbone, which in BWX is substantially substituted with 4-O-methyl-D-glucuronic acid (MeGlcA) while in WAX 1,2- and/or 1,3-linked arabinofuranosyl side chains prevail (Barron *et al.*, 2006; Daas *et al.*, 2017). By contrast, OSX contains 84% xylopyranosyl residue, along with a minor amount of arabinose, glucose and galactose (Morais *et al.*, 2010). The results of this study implied that the MeGlcA side chains acted as a significant barrier, reducing the accessibility of the xylan backbone to secreted enzymes of *A. mesophilus*. The arabinosyl side-chain substitutions, on the other hand, were supposed to be cleaved mainly by arabinofuranosidases, resulting in the exposed xylan backbone, which can be further digested by endo-xylanase. All these assumptions were supported by the strong xylanase and arabinofuranosidase activity of the secreted enzymes (Figure 3.1). However, because *A. mesophilus* exhibited limitations to utilize pentoses, including arabinose and xylose, it is more likely that not pentose monomers but rather the extracellularly released xylooligosaccharides are transported intracellularly by yet unknown sugar transporters, and further hydrolysis of XOS occurs primarily in the periplasm or the cytoplasm, similar to *Bacteroides thetaiotaomicron* VPI-5482 (Hehemann *et al.*, 2019). The tiny amount of pentose released extracellularly was insufficient to maintain *A. mesophilus* growth. However, in a co-culture, *A. mesophilus* can support the non-cellulolytic but saccharolytic bacterium *D. raffinosedens*, which is supported by the presence of the entire pentose phosphate pathway in the draft genome of *D. raffinosedens* (Rettenmaier *et al.*, 2020).

4.1.2 Bioinformatic analysis of a CAZyme gene cluster involved in hemicellulose/pectic substrate degradation

Degradation of recalcitrant polysaccharides by microbes is mainly based on the CAZyme encoding genes in their genomes. CAZymes, based on amino acid sequence similarity and their 3D folding

structure characteristics (Consortium, 2018; Lombard *et al.*, 2014), are divided into different families, including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyltransferases (GT) and some auxiliary activities (http://www.cazy.org/). Often, the CAZyme encoding genes are organized together with other genes, such as genes encoding sugar transporters and regulators, to hydrolyze the complex carbohydrate substrates. This kind of organization was designated CAZymes gene cluster (CGC) and has been extensively characterized in bacteria with strong lignocellulose degradation ability such as *Bacteroidetes* (Ausland *et al.*, 2021; Grondin *et al.*, 2017). Some anaerobic cellulolytic bacteria have developed a unique strategy for digesting lignocellulosic components, forming a multi-component, multienzyme complex known as a cellulosome. Cellulosome complexes containing various enzymatic functionalities are assembled on non-catalytic scaffold proteins via protein-protein interactions between the dockerin domain of each enzyme or accessory protein and the cohesion domains of scaffolding proteins. This special apparatus increases the hydrolysis efficiency towards various polysaccharides (Hyeon *et al.*, 2013). Cellulosome complexes are commonly present in some species from the *Clostridium* genus. For example, a 32-kb *xyl-doc* gene cluster encoding 14 putative cellulosome enzymes has been found on the genome of *Ruminiclostridium cellulolyticum* and functionally annotated to degrade hemicellulose (Mroueh *et al.*, 2019).

According to Prokka's automated annotation, the genome of the recently isolated bacterial strain *Acetivibrio mesophilus* (basonym *Hungateiclostridium mesophilum*) included five putative arabinofuranosidases, four of which belong to family GH43 and one to family GH51. Two of them (RXE58498.1 and RXE58509.1) together with a putative xylanase-encoding gene (RXE58512.1) from the same gene cluster were chosen for further characterization, but the gene RXE58509.1 was not further studied since the encoded protein did not show any enzymatic activity. In addition, the genes for a single putative GH28 polygalacturonase (locus tag: EFD62_RS12165) and another putative GH43 arabinofuranosidase (RXE59351.1) were also heterologously expressed, but the former failed to be produced and the latter showed minimal activity. Thus, the further work only focused on the genes RXE58498.1 and RXE58512.1, encoding *Am*Araf51 and *Am*Araf43, respectively.

Except for the *Am*Araf43 and *Am*Araf51, genes encoding a putative β-galactosidase (RXE58508.1) and four putative endo-xylanases (RXE58510.1, RXE58511.1, RXE58512.1, RXE58513.1), which were clustered downstream of the gene RXE58509.1 in the same orientation, were also annotated. The putative GH51 α-L-arabinofuranosidase encoding gene was found upstream of these CAZyme genes,

separated from these co-linearly organized genes by multiple genes encoding an ABC transporter and regulatory proteins (Figure 3.2). Based on the predicted activities, it can be speculated that the CAZymes gene cluster participates in hemicellulose degradation and incorporation of the released products, which is consistent with the capability of strain N2K1 to degrade complex lignocellulosic substrates such as WAX, OSX and SBP (Rettenmaier *et al.*, 2019). All the ORFs between RXE58508.1 and RXE58513.1 have signal peptide-encoding sequences and dockerin domain at the N-terminus and C-terminus of the protein, respectively, suggesting that these genes encode cellulosome proteins that were secreted extracellularly. Because N2K1 can poorly utilize the externally supplied monosaccharides (Rettenmaier *et al.*, 2019), the co-existence of ABC transporter-encoding genes in the hemicellulose degrading gene cluster indicates that oligosaccharides released from substrate polymers are further transported into the cell before being degraded to monosaccharides. Regarding the absence of a signal peptide, it can be assumed that *Am*Araf51 (RXE58498.1) is an intracellular enzyme and, like other cytosolic arabinofuranosidases, may be involved in the cleavage of arabinose-containing oligosaccharides after their internalization into the cytoplasm (Inacio *et al.*, 2008b). However, it could also play a (partial) extracellular role, which can be exemplified by the extracellular proteomic analysis of a *B. subtilis* strain which revealed almost half of secreted proteins without signal peptides in the culture supernatant (Antelmann *et al.*, 2001), possibly released by different secretion mechanisms or by bacterial cell lysis.

4.1.3 Characterization of *Am*Araf51 and *Am*Araf43

Arabinofuranosidases, as accessory enzymes, are essential for the complete deconstruction of arabinose-containing polysaccharides, such as arabinan and arabinoxylan. WAX-RS and WAX-I contain almost the same Araf/Xylp ratio, with 38:62 in WAX-RS and 36:51 in WAX-I, but the degradation of WAX-RS is much easier than WAX-I due to the alkaline extraction condition of WAX-RS which leads to the break of ferulic acid cross-links between two arabinofuranosyl residues and acetyl groups in the substrate polymer (Ma *et al.*, 2017; Mroueh *et al.*, 2019). SBA used in this study is a relatively homogeneous substrate where around 60% of the α-1,5-linked arabinan backbone residues are substituted with arabinosyl side chains at O-3 positions, while fewer arabinosyl residues in the backbone are modified with arabinose moieties at O-2 positions (Cartmell *et al.*, 2011). In comparison, debranched arabinan mainly has a linear arabinan backbone with few side-chain substitutions.

Regarding their substrate preference, *Am*Araf51 and *Am*Araf43 showed converse activities against arabinan or arabinoxylan substrates: *Am*Araf51 displayed a decreasing order of activity on linear arabinan (LA) > branched arabinan (DA) > sugar beet arabinan (SBA), suggesting that *Am*Araf51 has a significant capacity to cleave arabinan α-1,5-linked arabinofuranosyl residues in the backbone, but its hydrolysis ability against arabinoxylan (both WAX-I and WAX-RS) was lower than the limit of detection of the assay used (Table 3.2). Additionally, its ability to cleave different oligosaccharides also reflected the preference of *Am*Araf51 against arabinan over arabinoxylan. For example, *Am*Araf51 can efficiently cleaved off the internally double substituted arabinosyl side chains from the arabinose-based backbone ($AA^{2,3}A$), but not from the xylose-based backbone in $XA^{2,3}XX$ (Figure 3.6b, Figure 3.7c). The strong activity of *Am*Araf51 towards arabinosyl side chains in both arabinan and arabino-oligosaccharides suggests that *Am*Araf51 plays an important role in the utilization of arabinan and AOS by *A. mesophilus* strain N2K1. By contrast, *Am*Araf43 showed an opposite order of substrate preferences. It has very weak activity towards pectic arabinan, but high activity against WAX, especially WAX-RS. According to Prokka's automatic annotation, *Am*Araf43 was predicted to be an endo-xylanase. However, it cannot cleave the xylan backbone but only releases arabinose from WAX. No xylooligosaccharides were observed in the hydrolysis products (Figure 6.16b). Therefore, *Am*Araf43 represents an exo-α-L-arabinofuranosidase, which further demonstrated that GH43 enzymes could be misleadingly annotated by the use of automatic annotation. It is necessary to characterize the authentic activities of the enzymes. Similarly, a putative GH43 enzyme (subfamily 29) encoded by the gene axb8 from *A. thermocellus* (syn. *C. thermocellum*) strain B8 was functionally annotated to be an arabinofuranosidase, but it could only liberate xylose from WAX (de Camargo *et al.*, 2018). Generally, the molecular structure of the enzymes, particularly the molecular details of the substrate-binding site as well as the presence of carbohydrate-binding modules such as CBM6, CBM13, CBM35 and CBM42, can determine the binding and cleavage abilities of the enzyme to xylan substrates (Abbott *et al.*, 2009; Mewis *et al.*, 2016). Therefore, the binding ability of *Am*Araf43 to arabinoxylan rather than arabinan might be related to the CBM6 module of the enzyme. When we only cloned the GH43 domain of *Am*Araf43, the enzyme lost all its activity, but this has to be further confirmed experimentally because the GH43_C2 domain may also influence *Am*Araf43*'s* enzymatic activity. Similar to *Am*Araf43, an arabinofuranosidase (designated *Ct*Abf43A) from *C. thermocellum* and its closest homolog *Bs*AXH-m2,3, belonging to family GH43, was shown to contain a narrow and deep V-shaped cleft, which is likely to accommodate the

arabinoxylan backbone (Goyal *et al.*, 2016; Vandermarliere *et al.*, 2009). *Ct*Abf43A only release arabinose from arabinoxylan but not from sugar beet arabinan. *Am*Araf43 in our study showed 59.20% amino acid sequence identity with *Ct*Abf43A from *C. thermocellum*. Therefore, it may also have a similar mechanism regarding substrate specificity.

To further elucidate the action mode of the two arabinofuranosidases, structurally defined arabino- and xylo-oligosaccharides were also applied to this study. Both enzymes can cleave O-2 and/or O-3 position of arabinose moieties, but the enzymatic efficiency was dramatically influenced by the accurate feature of the side chain modification. With AOS as substrates, *Am*Araf51 exhibited cleavage activities that were negatively correlated with the complexity of the substrate oligomers (decreasing order of activities on A3> AA^3A> $AA^{2,3}A$ +AA^3AA), which could be observed by the time required for total hydrolysis of the oligomers. In comparison, *Am*Araf43 showed a converse order of substrate preference ($AA^{2,3}A$ +AA^3AA > AA^3A > A3), the activity towards α-1,5-arabinosyl linkages in A3 was barely observed (Figure 3.6a,b). With AXOS, both enzymes were capable of cleaving arabinosyl side chains from linear and/or branched XOS. *Am*Araf43 preferred to cleave the 1,3-linked arabinose moieties in double substituted xylose residues, including substrates $XA^{2,3}XX$ and $A^{2,3}XX$. By contrast, the 1,3-linked arabinosyl side-chain substitutions in substrates A^3X, XA^3XX are less preferred (Figure 3.7c), which fits to the characteristics of AXHs-m,d3 type arabinofuranosidases (Mroueh *et al.*, 2019). Compared to *Am*Araf43, *Am*Araf51 can hardly cleave the internally double substituted xylotetraose ($XA^{2,3}XX$). However, it showed high activity against internally twofold arabinosylated $AA^{2,3}A$ (Figure 3.6b, Figure 3.7b), which suggested that this enzyme prefers the arabinofuranosyl residues in the arabinose-based backbone. According to the cleavage mode towards the arabinose-substituted XOS, *Am*Araf51 can be grouped into the AXHs-m,d type of arabinofuranosidase, which can degrade the mono-substituted and double substituted xylose located at the terminal and/or internal position of the XOS (Mroueh *et al.*, 2019). Moreover, *Am*Araf51 also exhibited β-xylosidase side-activity with linear XOS and arabinosylated XOS as substrates, where the XOS with different degrees of polymerization were also found in the hydrolysis products (Figure 3.7a). This kind of bifunctional enzyme is likely to appear in evolution by combination from genes that encode functionally tightly related enzymes (Khandeparker *et al*., 2008; Yourno *et al.*, 1970). The bifunctional enzymes from family GH51 were also found in other cases, for example, a GH51 from *Alicyclobacillus* sp. A4 displayed a dual activity of endo-xylanase and arabinofuranosidase. It can hydrolyze the soluble wheat arabinoxylan into arabinose, xylose, xylobiose,

xylotriose and xylotetraose (Yang *et al.*, 2015). Besides, another multi-functional GH51 arabinofuranosidase from *Paenibacillus* sp. THS1, as described by Bouraoui *et al*., also displayed a similar dual activity like the GH51 arabinofuranosidase from *Alicyclobacillus* sp. A4 (Bouraoui *et al.*, 2016). It has also been reported that a GH51 arabinofuranosidase from *G. stearothermophilus* T-6 could bind the xylopyranosidic substrate moieties in its active site, but its xylanase activity has not been determined (Shallom *et al.*, 2002). Usually, multi-functional enzymes from bacteria have multi-domain architectures, with each domain harboring a specific activity. However, *Am*Araf51 only showed a single active site, and this kind of multi-functional activity may be of interest for the development of new biomass-degrading cocktails with significantly reduced enzyme production costs (Bouraoui *et al.*, 2016). *Am*Araf51, which lacks a putative signal peptide at the N-terminus of the protein, is predicted to be an intracellular enzyme. Therefore, the dual α-arabinofuranosidase and β-xylosidase activities combined in this enzyme may spare the necessity to synthesize two separate enzymes and thus may represent a cost-saving way for the bacterial cell to hydrolyze the internalized AOS and/or AXOS/XOS released from arabinan and xylan substrates, respectively.

Usually, the enzyme's thermostability is an inherent property, which is determined by the structure of the protein (Ward *et al.*, 1988). Additionally, it is also a vital parameter to balance the enzymatic stability and activity and regulate the temperature range applied to industrial production. Regarding the resistance of the enzymes against thermo-inactivation, *Am*Araf51 is preferable to *Am*Araf43 in industrial applications since it is very stable at both high temperature and its temperature of maximum activity (Figure 3.5c). The catalytic domain of GH51 enzymes contain an N-terminal (β/α)8-barrel structure, which is followed by a functionally unknown β-sandwich domain (Lagaert *et al.*, 2014). Many studies have shown that the (β/α)8-barrel architecture often related with a relatively high stability of the enzyme. For example, the enzymes with this structure from thermophiles are commonly very stable under harsh environmental conditions, e.g. coming from increased association state, additional salt bridges or hydrogen bonds (Sterner *et al*., 2005). Half-lives of GH51 arabinofuranosidase at high temperatures have been reported ranging from hours up to several days, which can be exemplified by an *A. clariflavus* DSM 19732 GH51 arabinofuranosidase, which at its optimal temperature of 60°C (Geng *et al.*, 2019) had a half-life of seven days. In contrast, the fivefold β-propeller structure of family GH43 enzymes is less studied in detail regarding thermostability (Yeoman *et al.*, 2010). Additionally, both *Am*Araf51 and *Am*Araf43 have a higher tolerance against high concentrations of hydrolysis

products than other previously reported arabinofuranosidase (IC50=100 mM to 500 mM) (Gao *et al.*, 2018), and the hydrolysis ability of *Am*Araf51 and *Am*Araf43 against *p*NP-α-L-arabinofuranosidases was not influenced by up to 533 mM of the monosaccharide arabinose.

4.1.4 Synergism of *Am*Araf51, *Am*Araf43 and *Pp*Abn43, M_Xyn10 on SBA and WAX-RS degradation

*Am*Araf51 and *Am*Araf43 showed specific activity on pectin substrate (arabinan) and hemicellulose substrate (arabinoxylan), respectively. To further explore the potential role of these enzymes as the components of enzyme cocktails for the deconstruction of plant biomass, the synergistic effects of both enzymes were studied together with two endo-active enzymes available in the laboratory, including an endo-xylanase from a fermenter metagenomic library (M_Xyn10) and an endo-arabinanase from *Paenibacillus polymyxa* DSM 292 (*Pp*Abn43). Presumably, as an exo-acting enzyme, α-L-arabinofuranosidase can cleave off the arabinosyl side-chain substitutions from sugar beet arabinan and wheat arabinoxylan, thus generating more easily accessible intermediates for endo-acting enzymes such as endo-arabinanase and endo-xylanase. Likewise, endo-active enzymes which randomly cleave the substrate polymers into oligosaccharides, resulting in more poly-/oligosaccharide ends, thus improving cleavage efficiency of exo-acting enzymes, such as xylosidase and arabinofuranosidase activity of *Am*Araf51 towards terminal xylosidic and arabinosidic residues of oligosaccharides. In recent years, studies of synergism between endo-active and exo-active enzymes have increasingly been carried out with isolated polysaccharides and even with complex plant biomass degradation. However, each portion of enzyme cocktails from different GH families also significantly influence the yield of the hydrolysis products, which is mainly determined by the cleavage pattern of each enzyme. For example, the concerted action of a GH51 (Abf51) arabinofuranosidase from *Acetivibrio clariflavus* DSM 19732 (previously *Hungateiclostridium clariflavum*) and a GH11 β-1,4-xylanase (XynA) from *Thermomyces lanuginosus* along with a GH43 β-1,4-xylosidase (Xyl43C) from *Clostridium clariflavum* were studied by Geng *et al*. for the degradation of arabinoxylan (18.5 g L^{-1}). The saccharification level improved six times compared with the single enzyme treatment (Geng *et al.*, 2019; Le *et al.*, 2014). By contrast, another study did not show the synergistic effect between GH51 and GH11 (Bouraoui *et al.*, 2016). Regarding the GH11 endo-xylanase which degraded the substrate arabinoxylan into internally double substituted xylo-oligosaccharides instead of oligosaccharides with terminally substituted xylopyranosyl residues

(McCleary *et al.*, 2015), the former decorations usually resist cleavage by most of exo-arabinofuranosidases, including GH51 (Koutaniemi *et al.*, 2016), as is also shown in this study, i.e., $XA^{2,3}XX$ is hardly degraded by *Am*Araf51 (Figure 3.7b). An exception also exists, which was reported by Dos Santos *et al.*, who reported an arabinofuranosidase, named *Xac*Abf51, to be capable of cleaving off the internal two-fold arabinosyl substitutions of xyloparanosyl units. The crystallographic structure of *Xac*Abf51 revealed it contains a pocket arranged near to subsite -1 that has capacity for a second arabinofuranosyl group. However, the weak catalytic activity of GH51 arabinofuranosidases (including *Am*Araf51) towards internally double substituted xylopyranosyl residues in $XA^{2,3}XX$ suggested that GH10 endo-xylanase could be superior over GH11 endo-xylanase as a partner of GH51 since the hydrolysis products of GH10 endo-xylanases often contain single or double arabinosyl side-chain substitutions on terminal xylopyranosyl residues (McCleary *et al.*, 2015). The results of the work presented here show that by combination of *Am*Araf51 and M_Xyn10 with WAX-RS as substrate, the yields of arabinose and xylose increased up to 14-fold and 5-fold, respectively, compared with single enzyme treatment (only *Am*Araf51 or M_Xyn10 as a catalyst) (Figure 3.9b,c). Regarding sugar beet arabinan degradation, a six-hour co-incubation of the substrate with *Am*Arf51 and *Pp*Abn43 in one batch treatment increased the arabinose yield by 3.4-fold compared to *Am*Arf51 alone (Figure 3.9a), implying that *Am*Arf51 could be a good candidate for SBA monomerization alongside endo-arabinanase. Further studies are of need to target the development of efficient enzyme cocktails to degrade complex plant biomass, especially SBP. Compared with polysaccharides and oligosaccharides, it will be a great challenge to clarify the synergistic effects between several or many enzymes to develop low-cost defined enzyme cocktails with high catalytic efficiency for the complete breakdown of recalcitrant agricultural residues.

4.2 Searching novel enzymes by function-based metagenomic library screening

4.2.1 Fermenter metagenomic library construction

As a cultivation-independent method, metagenomics partially overcomes the difficulty of simulating natural conditions in the laboratory and the limited cultivability of many bacteria, allowing researchers to investigate microbial diversity directly from the entire DNA of environmental samples. As a result, metagenomics is a crucial tool for determining the connection between microbial community structure and function. Specifically, functional-based screening approaches based on the construction of a

metagenomic library and sequence-driven analysis based on high-throughput sequencing of the whole ambient DNA, are two of the most commonly utilized methodologies in metagenomic studies (Leis *et al.*, 2013). The present study used the function-based strategy by creating metagenomic libraries with the fermenter microbial population (named T1T2 and MOD18) that showed significant degradation ability towards SBP. In general, lignocellulolytic biomass in biogas fermenters can select a mature microbial community or enhance excellent copies of particular genomes involved in deconstructing the main polysaccharides in plant biomass (Jiao *et al*., 2006; Montella *et al.*, 2016). The soluble sugar from hydrolysis products can serve as "food" for bacterial growth. Meanwhile, the by-products, including methanol or ethanol, can be applied to industrial productions.

Usually, *Firmicutes* and *Bacteroidetes* are the most common phyla in biogas digesters (Gullert *et al.*, 2016), but they can be influenced by various other factors (Li *et al.*, 2016), such as temperature or inoculum. Our analysis of the bacterial community by using 16S rRNA amplicon sequencing showed that *Firmicutes*, *Thermotogae*, and *Bacteroidetes* were the dominating phyla in SBP enriched fermenter samples T1T2 and MOD18 (Table 3.3). According to prior research, all three phyla contain powerful bio-degraders with excellent lignocellulose degradation capacity (Benedetti *et al.*, 2019; Gavande *et al.*, 2021). They were regarded as promising biotechnological resources to create energy alternatives to traditional fossil fuels. Since T1T2 and MOD18 showed a strong lignocellulolytic activity towards SBP, the carbohydrate-active enzymes can degrade the glycosidic bonds in the major components of SBP, including pectin, cellulose and hemicellulose, were supposed to be present in the metagenomic library and became the major targets of this study.

4.2.2 Screening of metagenomic libraries by using CPH-/ICB substrates

To improve the screening efficiency, a high-throughput assay system was established which was based on house-made chromogenic substrates prepared by labeling polysaccharides (CPH substrates) /plant biomass (ICB substrates) with chlorotriazine dyes (Kračun *et al.*, 2015). The chromogenic substrates were prepared in three colors (red, green and blue). By mixing different color-labeled substrates in the reaction plates, single enzymes with side activities or multiple enzymes from the same clone inserts can be identified in a single well, thus improving the screening efficiency. However, this screening method has limitations regarding its use to search for esterase and exo-acting activities because alkaline treatment during the substrate preparation process can break the di-ferulate bridges in the

substrates and the steric hindrance of the substrates limits the accessibility of exo-acting enzymes(Kračun *et al.*, 2015). However, since CAZyme encoding genes are frequently clustered in genome sequences, bioinformatics tools can identify auxiliary enzymes in the proximity of other genes for polysaccharide hydrolysis and utilization.

4.2.3 Sequencing of positive fosmid clones

After finishing the metagenomic library screening, the genetic information of the positive fosmid clone should be unveiled. To this end, target genes encoding enzymes related to the phenotype of the corresponding fosmid clone can be further identified by searching against the known database. The "barcode sequencing method," which involves attaching unique sequence index adapters to DNA fragments from different fosmid clones during NGS library preparation so that each read can be recognized and sorted during post-sequencing data processing, is extensively used for large-scale fosmid clone sequencing. This technique has so far been used to sequence fosmid clones with various activities, such as prebiotic degradation, dietary fiber breakdown, and cellulosic biomass conversion (Cecchini *et al.*, 2013; Tasse *et al.*, 2010; Gong *et al.*, 2012). However, preparing barcode-based samples is always expensive, especially when sequencing a large number of fosmid clones (Lam *et al.*, 2014). Some other strategies, such as Sanger sequencing by primer-walking along with the insert (Ferrer *et al.*, 2005; Jiang *et al.*, 2009; Simon *et al.*, 2009b; Torres-Cortes *et al.*, 2011), subcloning the smaller fragments of fosmid insert as well as transposon mutagenesis (Gloux *et al.*, 2011; Jeon *et al.*, 2011; Rabausch *et al.*, 2013), can also be applied to obtain the sequences of clone inserts (Allen *et al.*, 2009; Culligan *et al.*, 2012). However, these methods are suitable only for small-scale sequencing due to the low throughput. To balance the economic cost and sequencing efficiency, we use a hybrid Sanger/Illumina Shotgun sequencing approach reported by Lam *et al*., (2014). With Shotgun sequencing, the genetic information of the cloned inserts can be obtained in a short time. Further clone-by-clone Sanger end sequencing allows the retrieval of clone-specific sequences. A comparable method combined the 454 pyrosequencing with Sanger sequencing was previously applied to a metagenome derived medium-insert plasmids sequencing, as described by Džunková *et al*., (2012). Regarding the necessity of establishing a low-cost sequencing method for large-scale fosmid sequencing, we choose this method and tried to improve this technology by sequencing the DNA from 55 positive fosmid clones.

Evaluation of the shotgun sequencing result shows the presence of contamination from *E. coli* genomic DNA. Similar problems also were reported with a BAC library or cosmid library sequencing (Bazzani *et al.*, 2011; Lam *et al.*, 2014). This contamination does not influence the post-sequencing retrieval. The retrieval only relies on the clone's end tag. However, due to the undesired DNA template in the reaction, the required depth of coverage can be influenced and miscalculated. That is also why even with 50,000-fold coverage depth, some fosmid clones are still partially sequenced. Thus, to improve the pooled sequencing strategy, sample pretreatment is necessary to remove the interference from *E. coli* genomic DNA, such as using Plasmid-Safe™ ATP-Dependent DNase (Epicentre) to remove the *E. coli* genomic DNA (Bazzani *et al.*, 2011). After obtaining contigs from Illumina shotgun sequencing, Sanger end tags were used to assign contigs to specific fosmid clones. However, due to the insufficient sequencing depth, only 11% of clones were fully sequenced, with both forward and reverse end tags aligned to the same contig. Nearly 47% of the clones were partially sequenced, with a gap in either the middle or at the end of the DNA insert.

4.2.4 Functional annotation of CAZymes and CAZyme gene cluster (CGC)

The majority of putative CAZymes identified from Shotgun sequencing belonged to the GH43, GH51, GH10 and GH11 families. These GH families mainly comprise enzymes engaged in the degradation of hemicellulose or pectin substrates such as arabinan or arabinogalactan, which matches the fermenter microbial community's high lignocellulolytic degradation capacity. In terms of the CMCase activities found in the metagenomic library, 37 fosmid clones with this activity were chosen for sequencing. However, only 12 putative CAZymes associated with CMC (Carboxymethyl cellulose) or cellulose degradation were found (Table 3.4, Table 3.6). CMC is product chemically modified cellulose derivative obtained by treating cellulose with sodium hydroxide and chloroacetic acid (Anita, 2013). The less-than-expected results regarding the identification of cellulolytic enzyme-encoding genes might have to do with conserved sequences in cellulolytic genes, which is a bottleneck in subsequent sequencing reads assembly (Lin *et al.*, 1995). This problem can be avoided by preparing a balanced pool with various functionalities of fosmid clones.

Besides single genes that encode putative enzymes, various CAZyme gene clusters were also functionally annotated by the dbCAN database. Many of them are involved in the hydrolysis of complexed heteropolysaccharides by the coordinated action of different enzymes. Contig40, with the

size of 11,877 bps, derived from the fosmid clone “BM1T-2 H7”, contains five genes encoding putative CAZymes are involved in arabinoxylan degradation (Figure 3.13). Two of them, belonging to GH3 and GH10 (M_Xyn10), were also isolated from a metagenomic library constructed with the microbial community from a thermophilic methanogenic digester. These two enzymes were further biochemically characterized, revealing that they were thermostable and had xylosidases (GH3) and xylanase (GH10) activity, respectively. Moreover, both enzymes played an essential role in a mini enzyme cocktail for steam-exploded corncob degradation (Wang *et al.*, 2015). The broad distribution of these two genes showed that they play a key role in lignocellulose breakdown. The thermostability of these two enzymes also suggested the possibility of other identified CAZymes on the same contig being thermostable.

4.2.5 Pectinolytic enzymes from metagenomic library screening

4.2.5.1 Bioinformatic analysis of the putative arabinosyl hydrolases

The genes encoding the seven putative arabinosyl hydrolases identified in this investigation come from the same fosmid clone insert. Phylogenetic analysis of the enzyme diversity showed these potential CAZymes had the highest sequence similarity with proteins from diverse bacterial strains, including species from the *Clostridium* sp., *Xylanivirga thermophila* and *Caldicoprobacter oshimai* (Figure 3.17). Further analysis of additional genetic information from the same clone inserts revealed that most of the various ORFs shared over 99% similarity with genes/proteins from the less well studied *X. thermophila* (Figure 6.7). So far, the only *X. thermophila* strain has been isolated by Liu *et al*. from an anammox dominant wastewater treatment plant. Phylogenetic analysis of the 16S rRNA gene sequence suggested that this isolated strain is from the order *Clostridiales* and has 89.9% sequence identity with *Caldicoprobacter faecalis* DSM20678T (Liu *et al.*, 2020).

All the putative arabinanases/arabinofuranosidases characterized in the present study do not contain a predicted signal peptide. Therefore, they are likely to be intracellularly localized. Two endo-arabinanases, however, have significant enzymatic activities on SBA and DA, but the large molecular size and steric hindrance prevent these substrate polymers from crossing the cell membrane. As an explanation, the arabinanases may be exported by an alternate secretion mechanism. SignalP exclusively identifies the proteins transported primarily via the Sec or Tat pathways by BLAST analysis of the conserved signal sequences of proteins (Freudl, 2018). However, bacteria can employ diverse

export mechanisms. Some of the secretion pathways are even specific for only a few proteins (Lenz *et al.*, 2003). Certain bifidobacteria, for example, are capable of hydrolyzing a wide range of carbohydrates, but only a handful of the CAZymes produced include signal peptides (Ejby *et al.*, 2019). Aside from that, an extracellular proteome study of *B. subtilis* revealed that only 50% of released protein contained a signal peptide (Antelmann *et al.*, 2001). The signal peptide is not the only way to determine the localization of a protein. An alternate secretory pathway that exports the endo-arabinanase extracellularly might exist. The other four arabinofuranosidases could play an intracellular role since the released AXOS/AOS can be internalized into the cytoplasm/periplasm by sugar transporters (Hehemann *et al.*, 2019; Inacio *et al.*, 2008a), but an extracellular role may also exist.

4.2.5.2 Biochemical characterization of the enzymes

The physiological characteristics of the newly isolated *X. thermophila* reports its growth at an optimal temperature of 50°C in an alkalic environment (Liu *et al.*, 2020). The six arabinosyl hydrolases of the present study that putatively originated from this bacterial species had comparable biochemical properties, such as a slightly acidic optimal pH of 4.5-6.5 and a temperature of maximum activity at approximately 50°C (except MC60GH43 with an optimal temperature at 37°C) (Figure 3.18). The best pH condition was not consistent with *X. thermophila*'s alkaline growth preference, but that temperature condition for maximum activity in the assay system used was close to the bacterial optimal growth temperature and the operation temperature of the biogas fermenter utilized in our investigation (Liu *et al.*, 2020). The physiological characteristics of the host strain are usually consistent with the biochemical properties of the enzymes. For example, enzymes from the thermophile *Thermus thermophilus* can withstand temperatures as high as 90°C (Pantazaki *et al.*, 2002). Also, the optimum conditions of arabinofuranohydrolases isolated from *P. militaris* are similar to the bacterium's growth conditions (Bastien, 2013). Furthermore, all of the arabinosyl hydrolases characterized in this thesis are highly stable at their temperature and pH optimum, which is advantageous for developing enzyme cocktails. Less effort will be needed for evaluating the system conditions to optimize the activity of each biocatalyst.

Both MC68GH43-1 and MC72GH43-2 belong to GH43_4. This subfamily, along with subfamilies 5 and 6, usually exhibits endo-arabinanase activity (EC 3.2.1.99) (Wefers *et al.*, 2017), which is in agreement with the endo-cleavage mode of MC68GH43-1 and MC72GH43-2. Additionally, GH43 enzymes from *T.*

petrophila (Squina *et al.*, 2010), *B. subtilis* (Takao *et al.*, 2002), and *C. polysaccharolyticus* (Wefers *et al.*, 2017) also belong to subfamily 4 and all of them have endo-arabinanase activities similar to MC68GH43-1 and MC72GH43-2. Compared to GH43_4 from *B. subtilis* and *C. polysaccharolyticus*, MC68GH43-1 and MC72GH43-2 exhibited distinct cleavage patterns towards DA. GH43_4 from *B. subtilis* and *C. polysaccharolyticus* hydrolyzed DA to only arabinose and arabinobiose, while MC68GH43-1 and MC72GH43-2 cleaved DA into arabinose, arabinobiose, and arabinotriose as end products (Takao *et al.*, 2002; Wefers *et al.*, 2017). Furthermore, the cleavage modes of MC68GH43-1 and MC72GH43-2 towards SBA were different. In comparison to MC68GH43-1, MC72GH43-2 can release oligosaccharides with a higher degree of polymerization (Figure 3.20b). In enzyme synergism experiments, the produced oligosaccharides were always efficiently digested by exo-acting arabinofuranosidase.

The affinity of enzymes towards either SBA or DA was calculated based on the Michaelis-Menten model. MC68GH43-1 and MC72GH43-2, two endo-arabinanases, had slightly stronger binding affinities for SBA than for DA (Figure 3.21a,b,c,d). The capacity of an enzyme to bind substrates can be associated with the architecture of its specific modules, such as the well-known carbohydrate-binding domain (CBM domain). MC72GH432 contains three domains: GH43, GH43_C and laminin G_3. Previous research found that the laminin G_3 domain is crucial for binding substrates, particularly arabinose-oligosaccharide with a branched arabinosyl side-chain substitution (Sakka *et al.*, 2019). This could explain why MC72GH43-2 had a higher affinity t\o SBA than DA. MC68GH43-1, on the other hand, has a multi-modular structure comprised of a GH43 domain and a GH43_C2 module (module associated with some GH43 modules, Pfam designation PF17851, beta-xylosidase C-terminal concanavalin A-like domain). However, the function of the GH43_C has not been studied well and is unclear.

The other arabinofuranosidases only exhibited a single catalytic module. Two GH51 arabinofuranosidases (MC57GH51 and MC60GH51) preferred SBA over DA as judged from their specific activities. Two GH43 arabinofuranosidases (MC60GH43 and MC68GH43-2), on the other hand, were solely active against DA (Figure 3.20a). SBA contains a backbone of 1,5-linked arabinofuranosyl residues, within which arabinosyl side chains are substituted at 60% of the O-3 positions and a minor percentage at the O-2 positions (Cartmell *et al.*, 2011). By contrast, DA has a linear backbone with a very low degree of substitutions. It was demonstrated that MC60GH43 and MC68GH43-2 could only cleave the α-1,5-linkages of the arabinan backbone while MC57GH51 and MC60GH51, on the other

hand, exhibited broader specific activity against both the side chain and the backbone of SBA.

Using structurally defined arabino- and xylo-oligosaccharides as substrates, this cleavage model was further validated. The results indicated that MC57GH51 and MC60GH51 were unable to cleave the internally double arabinosylated XA^{2,3}X but could cleave the two-fold arabinosyl side chains in A^{2,3}XX and AA^{2,3}A (Figure 3.22 a,b,c,d). This suggested that the backbone composition also influenced the activity of these enzymes. The internally double arabinosylated xylooligosaccharides are recalcitrant for cleavage by most arabinofuranosidases, as in our case (Dos Santos *et al.*, 2018; Liu *et al.*, 2021). Besides, all these enzymes only showed activities against pectic arabinan. However, with the hemicellulose arabinoxylan, which is also decorated with the same side chain, the activity of the arabinofuranosidases was negligible.

4.2.5.3 Bioinformatic identification of additional arabinan degrading enzymes

Bioinformatic analysis revealed a high abundance of arabinosyl hydrolase encoding genes on the same DNA fragment, which encode proteins with robust enzyme activity and various cleavage patterns. This supports the DNA fragment's significant involvement in arabinan breakdown and subsequent arabinose assimilation in the bacterium *X. thermophila* which contains a highly similar DNA segment in its genome. In addition to the seven characterized arabinosyl hydrolases, more genetic information was discovered on the same DNA fragment using dbCAN annotation. Among them, two more putative CAZymes involved in arabinan breakdown were identified. Both putative enzymes belonged to the GH127 and GH27 families (Figure 3.15). However, unlike the other arabinosyl hydrolase containing families, these two GH families are less studied. Family GH27 contains arabinopyranosidase, which can hydrolyze terminal β-arabinopyranose residues, such as the GH27 arabinopyranosidase from *C. polysaccharolyticus* (Wefers *et al.*, 2017). GH127, on the other hand, can cleave arabinosyl groups in pectic structural components, such as dimeric side chains with a terminal β-arabinofuranose (Wefers *et al.*, 2016; Wefers *et al.*, 2014).

Besides putative CAZyme encoding genes, 22 of the 39 ORFs on this DNA fragment were functionally annotated by KEGG Blast KOALA (Table 6.6). Many sugar transport and regulatory proteins were found. Additionally, this clone insert also encodes the enzymes necessary for the entire PPP pathway, including L-ribulokinase, transaldolase, transketolase, L-ribulose-5-phosphate-4-epimerase and L-arabinose isomerase. As a result, the bacterium proposedly degrades arabinan substrates and

metabolizes arabinose produced during the hydrolysis of arabinose-containing substrates.

Therefore, a model of the pentose metabolism pathway starting from hydrolysis of arabinose-containing substrates was proposed in *X. thermophila* (Figure 4.1). According to the model, MC68GH43-1 and MC72GH43-2, as two endo-arabinanases, are presumably secreted and/or covalently anchored to the cell surface extracellularly, where it can degrade arabinan into arabino-oligosaccharide. The free oligosaccharides can be transported intracellularly via the ABC transporters and the major facilitator superfamily transporter (MFS) anchored in the cell membrane's lipid bilayer (Ejby *et al.*, 2019; Pao *et al.*, 1998). Then the intracellularly located exo-arabinofuranosidases (MC57GH51, MC60GH51, M60GH43, MC68GH43-2) can further cleave arabinose moieties from these internalized oligosaccharides. Arabinose can be further transformed into L-ribulose before entering the pentose phosphate pathway (PPP). The PPP can further offset the functions that the regular glycolysis pathway cannot accomplish. It can not only produce energy but also generates essential sugars or precursors, such as 4-carbon sugar (D-erythrose 4-phosphate), a precursor of aromatic amino acids (Herrmann *et al.*, 1999), 5-carbon sugars (D-ribose), which is essential for the biosynthesis of DNA and RNA (Furukawa *et al.*, 2013), as well as the 7-carbon sugar (D-sedoheptulose-7-phosphate), which is a precursor for lipid A (Taylor *et al.*, 2008). So far, the PPP pathway has not been confirmed in *X. thermophila*. As a newly isolated xylan-degrading thermophilic and obligate anaerobic microorganism, the ability of *X. thermophila* to utilize pectic substrates has not been reported yet but seems likely based on the results described here (Liu *et al.*, 2020).

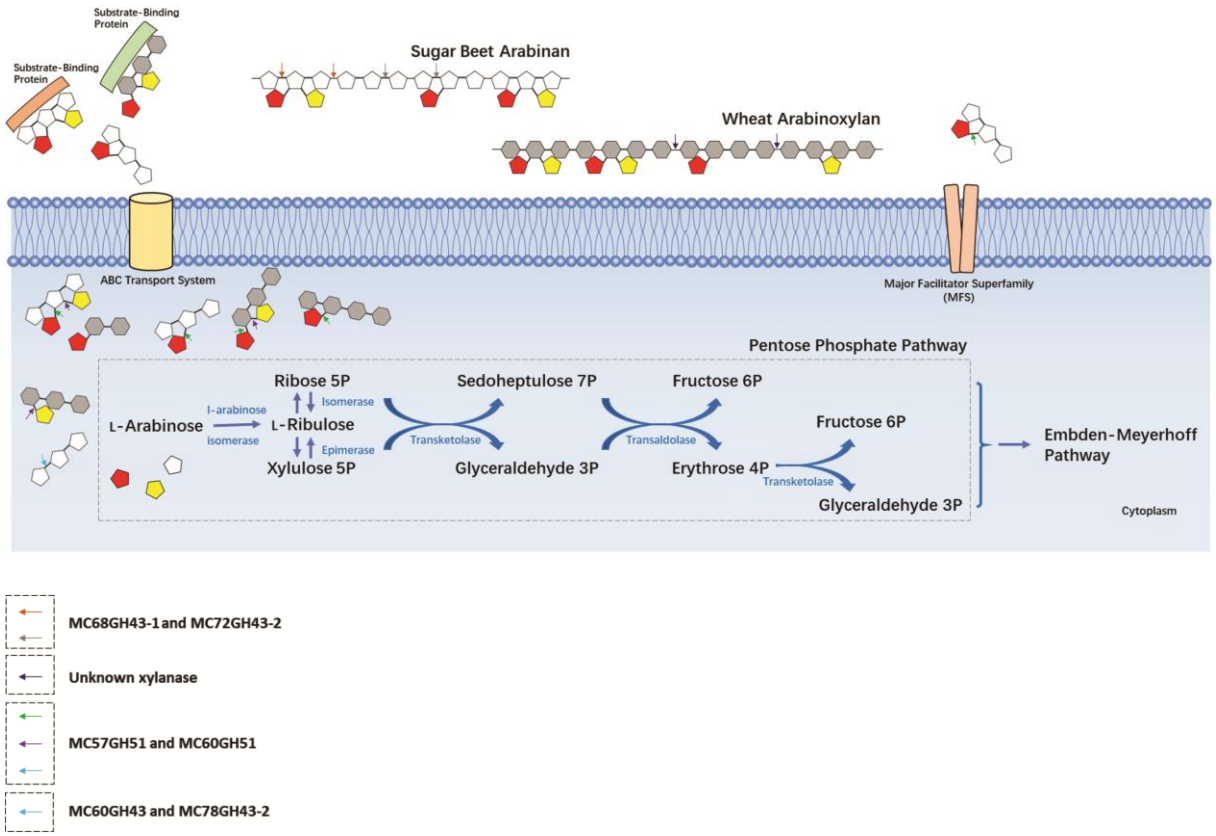


Figure 4.1 A schematic model of the proposed arabinose-containing polysaccharides degradation and arabinose metabolism in *Xylanivirga thermophila*. In this model, SBA/DA are cleaved into arabino-oligosaccharides by endo-arabinanases (MC68GH43-1 or MC72GH43-2), while arabinoxylan (WAX) is degraded into arabinoxylo-oligosaccharides by the unknown xylanases. The arabino-oligosaccharides and arabinoxylo-oligosaccharides are transported into the cell through an ABC transporter or an MFS transporter and further hydrolyzed into arabinose by intracellular enzymes such as MC57GH51 and MC60GH51 (Green, purple and blue color arrow) or MC60GH43 and MC68GH43-2 (Blue color arrow). Arabinose isomerase converts arabinose into ribulose, which then is converted to ribulose-5-phosphate or xylulose-5-phosphate to enter the pentose phosphate pathway (PPP). The gene cluster on contig57 encodes all the enzymes necessary for the pentose phosphate pathway. (As was shown in Fig 3.15).

4.2.6 Hemicellulolytic enzymes from metagenomic library screening

4.2.6.1 Positive fosmid clones with xylanase activity

Fosmid-based metagenomic library screening also allows the identification of novel or interesting xylanolytic enzyme-encoding genes. Various xylanolytic enzymes, such as endo-xylanases, beta-xylosidases, acetyl xylan esterases and alpha-arabinofuranosidases, have been isolated from nature or artificial environments through metagenomic approaches over the last two decades. Some of the enzymes that have been evaluated can meet the demand for xylanases with various enzymatic characteristics for specific industrial applications. For example, thermo-acidophilic and thermo-

alkaliphilic xylanolytic enzymes are ideal biocatalysts for stone washing and biopolishing industries as well as paper bioprocessing industries, respectively (Ariaeenejad *et al.*, 2019; Sae-Lee *et al.*, 2014; Verma *et al.*, 2013). In contrast, halophilic/halotolerant xylanolytic enzymes display important applications in the seafood processing and beverage industries for juices and wine clarification (Alokika *et al.*, 2018; Wang *et al.*, 2016a).

Among all xylanolytic enzymes, endo-xylanase is an essential xylan-degrading enzyme, which can randomly cleave the glycosidic linkages of the xylan backbone, thus generating various xylooligosaccharides. So far, xylanases have been classified in GH families GH5, GH8, GH10, GH11, GH30 and GH43 in the CAZyme database (Collins *et al.*, 2005). Among them, GH10 and GH11 contain most of the known endo-xylanases (Collins *et al.*, 2005; Dodd *et al.*, 2009), which is also the case in our findings, with 13 and 4 putative GH10- and GH11-encoding genes found in the metagenomic dataset, respectively (Table 6.5). However, the production of the CPH substrates used includes steps done under alkaline conditions, which causes the hydrolysis of ester bonds. As a result, chromogenic substrates have limitations in screening esterase enzymes. Besides, because of steric hindrance of cleavage sites, CPH substrates are usually not reactive with exo-acting enzymes. Therefore, only a few beta-xylosidase or esterase genes were found in the study.

Phylogenetic analyses of all the screened xylanases revealed most of them shared high amino acid sequence similarities with the xylanases encoded by genes from *Acetivibrio saccincola* and *Thermoclostridium stercorarium* (Figure 3.24a,b). Considering the high xylanase activities of the contig-retrieved fosmid clones and strong hemicellulose degradation abilities of the corresponding bacterial strains can be speculated. Furthermore, the high abundance of these bacterial strains in both the fermenter microbial community and the elephant dung sample could be responsible for the limited host range of xylanase encoding genes screened. As a result, the presence of a disproportionately large number of *Clostridium*/*Acetvibrio* specific genes in total eDNA increases the likeliness of retrieving these genes from recombinant fosmids of xylanase-positive library clones. Another possible reason is that *E. coli* is not suitable for heterologous expression of all the foreign genes, especially the genes encoding proteins that contain disulfide bonds or require phosphorylation, glycosylation, sulphation, proline cis/trans isomerization, lipidation, disulfide isomerization (Lueking *et al.*, 2000). Therefore, developing novel expression tools and alternative host organisms is necessary to expand the coverage of all bacterial genomes or capture the uncommon novel bacterial enzymes.

According to previous studies, xylanases from the GH10 family has been found to function better for the deconstruction of plant biomass or the polysaccharide xylan than the GH11 endo-xylanases. *Hu et al.* reported, for example, that GH10 endo-xylanase had greater accessibility to the xylan backbone in pretreated biomass, particularly in such with higher acetyl group concentration. Furthermore, GH10 xylanases often have a higher thermostability than GH11 enzymes (Hu & Saddler, 2018). From a different perspective, Geng *et al*. and Liu *et al*. demonstrated certain superior features of GH10 xylanases during hemicellulose degradation (Geng *et al.*, 2019; Liu *et al.*, 2021). Specifically, GH10 xylanase was found to produce AXOS with the double/single arabinosylated substitutions on the terminal xylopyranosyl residues with wheat arabinoxylan as substrate (McCleary *et al.*, 2015). In comparison, GH11 xylanase would generate hydrolysis-resistant oligosaccharides with double substitutions on the inner xylopyranosyl rather than terminal xylopyranosyl residues (McCleary *et al.*, 2015).

The cleavage modes of the screened fosmid clones on BWX were examined further, and the findings revealed that the cell extracts exhibited different activities. One explanation might be that some putative endo-xylanases seem to work synergistically with other enzymes encoded by the genes on the same clone insert, thus forming physically linked gene clusters to digest recalcitrant substrates fully. For example, putative GH10 xylanase encoded on contig40 of the present study was part of a gene cluster containing putative GH67, GH3 and other two GH43 genes (Figure 3.25b). All the corresponding putative enzymes can work synergistically for glucuronoarabinoxylan degradation. As mentioned before, the putative GH10- and GH3-encoding genes were also identified from another metagenomic study of a long-term enriched thermophilic methanogenic digester. Both enzymes were reported to be thermophilic xylanase and xylosidase, respectively, which were further used as the major components of enzyme cocktails for the degradation of a steam-exploded corncob (Wang *et al.*, 2015). That would be why xylobiose was not present in the hydrolysis products of BWX, while the purified enzyme can produce xylobiose. However, the cleavage specificities of the GH10 xylanase were not explored in Wang's study (Wang *et al.*, 2015). Since enzymes from the same origin usually share similar biophysical properties, it is possible that the other three enzymes (GH67 and two GH43) on contig40 are also thermo-resistant. This uniformity can further enhance their compatibility for complex polysaccharide degradation and is of great interest in industrial applications with harsh process conditions. Similarly, many other thermo-resistant GH10 xylanases were also isolated from metagenomic libraries. For

example, Verma *et al*. found the Mxyl gene from a compost-soil metagenome, which encodes a novel xylanase exhibiting maximum activity at 80°C and pH 9.0 (Dutilh *et al*., 2013). These results enhance the importance of functional metagenomic studies for discovering novel enzymes independent of microbial cultivation in pure cultures. Besides, three contiguous putative GH10 endo-xylanase genes were identified on contig61, putatively from *Acetivibro saccincola*, which may have increased the xylanase activity of the assigned fosmid clone (Figure 3.13). Additionally, the various cleavage patterns of the screened xylanase-active clones may also be possible due to the unique properties of the different GH10 xylanase.

4.2.6.2 Cleavage mode of M_Xyn10 against different xylose-based substrates

The cleavage specificity of M_Xyn10 was specifically explored by using various xylan substrates. The present work revealed that M_Xyn10 released different XOS and a relatively low amount of X1 from xylan. The structures of XOS are dependent on the type of xylan applied in the study (Figure 3.27). Both BWX and WAX have β-1,4-linked xylopyranosyl residues as the backbone, where the structural units are substituted with arabinofuranosyl residues at O-2 and/or O-3 position in WAX (Barron *et al.*, 2006), while in BWX, the backbones have the side chains of 4-O-methylglucuronic acid (Lattová *et al.*, 1992). However, the branched XOS in the hydrolysates have common features, with (a) terminal non-reducing end substituted single/double side chain(s), indicating the enzyme can cleave adjacent to, and on the non-reducing end of a D-Xylp residue substituted by either a 1,3- and/or a 1,2- α-L-Araf group in WAX and MeGlcA group at the O-2 position in BWX. It is similar to the GH10 xylanase from *Cellvibrio mixtus*, with only $A^{2,3}XX$ and $A^{3}X$ as hydrolysis products, but no xylose was reported in the hydrolysis products of WAX (McCleary *et al.*, 2015). Since WAX contains only unsubstituted, mono-substituted (1,3-) and di-substituted (1,2- and 1,3-) xylose residues, thus the 1,2-arabinosyl-linkage substituted XOS was not detected in the hydrolysis products (McCleary *et al.*, 2015).

Additionally, M_Xyn10 also has other unique properties which are different from previously reported GH10 xylanase. For example, the release of X1 from both WAX and BWX suggested the presence of xylosidase activity with this enzyme. With oligosaccharides, GH10 xylanase from *Cellvibrio mixtus* degrades $XA^{3}XX$ and $XA^{2}XX$ into $XA^{3}X$ and $XA^{2}X$, in which the arabinosyl substitution on the penultimate D-xylopyranosyl residue at the non-reducing end of $XA^{3}X$ and $XA^{2}X$ further prevents the cleavage of the terminal xylopyranosyl residue (McCleary *et al.*, 2015). However, the same steric

hindrance was not observed with M_Xyn10. It can cleave XA^{3}XX and XA^{2}XX into terminal non-reducing end substituted A^{3}X and A^{2}XX. In addition to the A^{3}X in the hydrolysates of XA^{3}XX, the released oligosaccharides also included X2 and X1 (Figure 3.30b), which indicated M_Xyn10 could also accommodate substitution of D-xylopyranosyl residue at +2 subsite. Additionally, M_Xyn10 can also liberate xylose from X3, but the glycosidic bond in X2 is poorly to be degraded. End hydrolysis products from other linear XOS (DP from 3-6) were always composed of X2 and X1 (Figure 3.30a). Since xylotriose is the smallest oligosaccharide substrate (xylobiose is resistant to degradation), at least two xylose-binding subsites next to the catalytic center can be degraded. In comparison, X3 seems to be a recalcitrant product of *C. mixtus* xylanase.

Therefore, a putative cleavage model was proposed. In this model, M_Xyn10 has a primary cleavage site that cleaves linear XOS into xylotriose. The secondary cleavage site of the enzyme can further cleave the xylotriose into xylobiose and xylose. A^{3}X, A^{2,3}XX and A^{2}XX are stubborn substrates for M_Xyn10, which was consistent with the specific activity of M_Xyn10 towards WAX (Figure 3.30b). A^{2}XX was not present in the hydrolysates of WAX, but it was far more difficult to be degraded than A^{3}XX, which indicated the 1,2-arabinofuranosyl linkages could produce more steric hindrance than 1,3-linkages. With BWX, a 4-O-methyl-D-glucurono-D-xylan, M_Xyn10 cleaved substrate specifically into MeGlcA2Xyl2, which has the same hydrolysis pattern with endo-xylanase from *Cryptococcus albidus* and *Streptomyces lividans* (Biely *et al.*, 1997a). When the MeGlcA decorated XOS was used as the substrate, M_Xyn10 was able to degrade XylMeGlcA2Xyl into two unidentified products (Figure 3.30a). MALDI-TOF analysis of the hydrolysis products reveals only one molecular mass of the compound in the hydrolysis products except for the substrate itself. Probably the enzyme can liberate non-substituted xylopyranosyl residues adjacent to substituted xylopyranosyl residues from both reducing and non-reducing ends, thus generating XylMeGlcA2 and MeGlcA2Xyl. However, when oligosaccharide MeGlcA2Xyl2 was used as a substrate, it resisted degradation, which was similar to the GH10 xylanase from *Caldicellulosiruptor* sp. F32 (Gong *et al.*, 2016). The results also suggested that the total degradation of BWX and WAX by GH10 xylanase was very likely limited by the MeGlcA2 side-chain substitutions and arabinofuranosyl side chains, respectively. Therefore, the other accessory enzymes, such as α-glucuronidases, which can cleave 4-O-methyl-glucuronic acid substituted on the non-reducing end, and arabinofuranosidases which can cleave the arabinofuranosyl residue (Biely *et al.*, 2000; Yeoman *et al.*, 2010), are necessary for increasing substrate accessibility for endo-xylanase and

exo-xylosidase to degrade the xylan efficiently and completely.

4.2.6.3 Transglycosylation by M_Xyn10

In addition to cleaving the glycosidic bonds from the backbone of β-1,4-linked xylopyranosyl residues, M_Xyn10 was also shown to have transglycosylation activity which means transfer of part of a glycoside from a saccharidic donor to a nucleophilic acceptor, which contains at least one hydroxyl group (Withers, 2001) (Figure 3.31a,b). In this study, *p*NP-Xyl2 clearly served as a donor, and *p*NP-Xyl2 and *p*NP-Xyl may be acceptors, resulting in various DP of *p*NP-xylooligosaccharides. Further, the para-nitrophenol group can be removed to generate xylooligosaccharides with different degree of polymerization. However, with *p*NP-Xyl, no transglycosylation activity was observed, probably due to the lack of a donor or because *p*NP-Xyl was not an acceptor. Besides M_Xyn10, other GH10 xylanases were also reported to harbor transglycosylation activity (Han *et al.*, 2019; Biely *et al.*, 1981). All these xylanases can serve as candidates for enzymatic oligosaccharide synthesis.

Other linear oligosaccharides with DP up to 6, were also used for testing the transglycosylation ability of M_Xyn10, but only weak transxylosylation was observed with X5 and X6 and perhaps to an extremely low extent with X3 and X4 which appear to be very poor donor saccharides. The hydrolysis and transglycosylation activities of the enzyme are dependent on the configuration of the binding cleft and cleavage site. GH10 xylanases usually have highly conserved subsites, including -2, -1 and +1. Therefore, variations in cleavage and transgylosylation specificities of different enzymes are mainly determined by the binding cleft of -3, +2, +3, and +4 subsites (Eneyskaya *et al.*, 2020). A previous study revealed the important role of subsites in the aglycone region for transglycosylation of *Pc*Xyn10A (GH10 xylanase from bacterial *Pseudomonas Cellulosa*) (Armand *et al.*, 2001). While in another study, the hydrolysis activity of GH10 xylanase from *Talaromyces leycettanus* was improved after reducing the binding ability at +3 and +4 subsites (Wang *et al.*, 2016b). However, more study must be carried out to reveal the transglycosylation mechanism of M_Xyn10.

4.3 Enrichment of microbial consortia with desirable biodegradation properties

Cellulose, xylan and pectin are three major components of SBP, which stands for a potential energy source for the production of biofuel or food products (Demain *et al.*, 2005). However, due to their recalcitrant structure, a specific repertoire of microbial functions was required to establish an

anaerobic digestion system to degrade SBP and its components completely. So far, many studies have been focused on the proper biodegradation of specific polysaccharides by isolated bacterial strains. However, pure cultures always displayed undesirable biodegradation rates (Koullas *et al.*, 1992; Kim *et al.*, 2006), suggesting that the breakdown of resistant plant biomass necessitates a wide range of specialized microbe combinations. Each microorganism in the community can play distinct roles. Many of them even build a close and long-term biological interaction so that they can benefit each party. The sum of all microbes can further synergistically promote lignocellulose degradation (Lee *et al.*, 2013).

To develop a superior microbial community with bio-deconstruction properties, we adopted a dilution-to-stimulation method using sequential-batch enrichments on three different substrates (cellulose, xylan, sugar beet pulp). Such an approach was also successfully applied to other studies. For example, the soil-derived microbial consortia were enriched with different plant biomass, which further unveiled the role of varying degraders in lignocellulose digestion (de Lima Brossi *et al.*, 2016). Besides, Jia *et al*.'s study showed the convergence of microbial community in long-term cellulose and xylan enrichments from anaerobic digesters (Jia *et al.*, 2016). However, only little research is available that compares the taxonomic differences arising when sugar beet silage-derived microbial communities are enriched on the major components of SBP and the plant biomass SBP (Chen *et al.*, 2009; de Lima Brossi *et al.*, 2016; Jimenez *et al.*, 2014; Lu *et al.*, 2005).

Here, we clearly showed that the substrate types could determine the structure of the microbial community. In terms of composition, the microbial consortia became quite stable after 4-6 times-transfers, the mature cultures were much lower in richness and evenness, this result was in line with a soil-derived microbial consortium enriched with different plant biomass, which also required 4 to 6 transfers until obtaining the stable microbial community (de Lima Brossi *et al.*, 2016). Because the sugar content and substrate structure of xylan, cellulose and sugar beet pulp vary a lot, the degradation efficiency for these compounds differed as well. Even though SBP has a considerably more complicated structure than the other two polysaccharides, enriched microbial consortia can degrade plant biomass SBP efficiently. It is most likely because the inoculum originated from a fermenter fed sugar beet silage. This recalcitrant plant biomass (sugar beet silage) can select a unique microbial community capable of digesting a comparable substrate, SBP. In comparison, filter paper with the major cellulose component was not as fast degraded as the other two compounds, which suggested that the "core" consortium required for cellulose deconstruction is specific and different from the microbial consortia in SBP and

xylan degradation. The MDS plot of microbial profiles corroborated our observations, demonstrating that cellulose-digesting microbial systems differed considerably from xylan/SBP-digesting microbial systems, whereas xylan/SBP-digesting microbial systems were comparable (Figure 3.36). However, Jia *et al*. described the functional and taxonomic convergence of microbial communities in anaerobic digestion sludge enriched in xylan and cellulose, which contradicted our results and further suggested that the inoculum could influence the structure of the microbial community (Jia *et al.*, 2016). Regarding the relatively low efficiency of the enriched microbial community on cellulose degradation. Cellulose, being a major component of SBP, was anticipated to be degraded following xylan digestion. In other words, it suggested that the microbial community prioritized substrate utilization, with xylan being used first, followed by the breakdown of cellulose/monogalacturonic acid. However, our enrichment process was unable to determine whether cellulose or monogalacturonic acid took precedence.

Generally, the microbial community structure in various environmental samples has been described in many already published studies. These data were based on either 16S rRNA amplicon sequencing or metagenome datasets (Jaenicke *et al.*, 2011; Wirth *et al.*, 2012; Stolze *et al.*, 2015; Zakrzewski *et al.*, 2012). According to that, *Firmicutes* and *Bacteroidetes* are summarized as the prevalent phylum. The mean ratio of these two phyla was usually 5.6-6.0: 1 in fermenter samples. In comparison, the average percentage of these two phyla in herbivorous animals changes to 1:1 (Gullert *et al.*, 2016). In our study, almost 1.32-fold higher relative abundance of the *Firmicutes* compared to the *Bacteroidetes* were calculated in the seed sample, and they were the most abundant OTUs at the phyla level (Figure 3.37). Both phyla were reported to contain abundant genes encoding for lignocellulolytic GHs, especially the *Bacteroidetes*. A similar bacterial structure was also shown in the enrichments, especially in SBP and xylan enriched microbial communities. The ratio of these two phyla fluctuated but was always dominants, suggesting that *Firmicutes* and *Bacteroidetes* were the major phyla involved in SBP- or xylan- digestion. However, the most enriched bacterial OUT in cellulose culture were *Firmicutes* (55.67%). Other phyla included *Bacteroidetes*, *Spirochaetes* and *Synergistetes*, with almost an equal amount of 10.29%, 13.76% and 12.50% of the total OUT, respectively (Figure 3.37). RNA-Seq data of a previous study showed that highly transcribed cellulolytic enzymes in the biogas fermenter were more than four-fold associated with the *Firmicutes* compared to the *Bacteroidetes* (Gullert *et al.*, 2016). Thus, there is no doubt to get the conclusion that *Firmicutes* dominate the major cellulolytic activities of a biogas fermenter.

The enzymatic activities of the secreted protein were also tested by using house-made CPH-/ICB-substrates or commercial AZCL-substrates. Secreted proteins from SBP- and xylan- enriched cell culture had similar activities on the tested substrates, including CPH-xylan and CPH-arabinoxylan. These carbon source-derived activities reflected the adaptation of the corresponding bacteria to these polysaccharides in the environment. Additionally, the microorganisms from the SBP enrichment probably have stronger and broader hydrolysis ability towards lignocellulolytic substrates and plant biomass, which the strong CPH-arabinan degradation ability can exemplify (Figure 3.33a). A possible reason is that gene regulation or protein expression of the same organism can differ under different substrate regimes. Thus, even though the xylan- and SBP-enriched microbial community had a similar structure, the secretion can differ with various substrates, similar to the temporal changes in exometabolome composition of *Alteromonas macleodii* when it was cultured on different substrates (Koch *et al.*, 2019). However, just a few clones with endo-active enzymes were detected after the filter paper enrichment procedure, even though the substrate can be fully degraded by the specific microbial community. This was also the case in Diego *et al*.'s study, even though metagenomic data showed cellulase-encoding genes and proteins in the secretory fraction. They claimed that the xylo-oligosaccharides produced from hemicellulose reduced cellulase activity significantly (Jimenez *et al.*, 2016; Kont *et al.*, 2013). However, in our study, the microbial consortia were grown on pure cellulose (filter paper), xylooligosaccharide had a minor influence on enzyme activity, but it was still hard to detect cellulase activity. Therefore, it is speculated that the cellulolytic enzymes were mainly present in the form of cellulosome protein complexes since there was a high proportion of *Firmicutes* in the filter paper enriched microbial communities. These cellulosome complexes mostly anchor to the cell wall of bacteria but are not freely released in the supernatant. Perhaps this led to the result that cellulolytic activity was hard to detect in the culture supernatant.

However, limited data makes it hard to reveal the relationship between the taxonomic change and the lignocellulolytic degradation ability of enriched bacterial consortia. More efforts, such as metasecretomics- and metagenomics-based approaches, must be applied to elucidate this relationship further.

5 References

Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., . . . Boraston, A. B. (2009). Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules. *Biochemistry, 48*(43), 10395-10404. doi:10.1021/bi9013424.

Ahmed, S., Luis, A. S., Bras, J. L., Ghosh, A., Gautam, S., Gupta, M. N., . . . Goyal, A. (2013). A novel alpha-L-arabinofuranosidase of family 43 glycoside hydrolase (Ct43Araf) from *Clostridium thermocellum*. *PLoS One, 8*(9), e73575. doi:10.1371/journal.pone.0073575.

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J, 3*(2), 243-251. doi:10.1038/ismej.2008.86.

Armenteros, J. J. A, Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., . . . Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology, 37*(4), 420-423. doi:10.1038/s41587-019-0036-z.

Alokika, Singh, D., & Singh, B. (2018). Utility of acidic xylanase of *Bacillus subtilis* subsp. subtilis JJBS250 in improving the nutritional value of poultry feed. *3 Biotech, 8*(12), 503. doi:10.1007/s13205-018-1526-2.

Angelov, A., Mientus, M., Liebl, S., & Liebl, W. (2009). A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst Appl Microbiol, 32*(3), 177-185. doi:10.1016/j.syapm.2008.01.003.

Anita, B. B. (2013). Biodegradation of carboxymethyl cellulose using *Aspergillus flavus*. *Science International, 1*(4), 85-91. doi: 10.17311/sciintl.2013.85.91.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J. M., & Hecker, M. (2001). A proteomic view on genome-based signal peptide predictions. *Genome Res, 11*(9), 1484-1502. doi:10.1101/gr.182801.

Ariaeenejad, S., Hosseini, E., Maleki, M., Kavousi, K., Moosavi-Movahedi, A. A., & Salekdeh, G. H. (2019). Identification and characterization of a novel thermostable xylanase from camel rumen metagenome. *Int J Biol Macromol, 126*, 1295-1302. doi:10.1016/j.ijbiomac.2018.12.041.

Armand, S., Andrews, S. R., Charnock, S. J., & Gilbert, H. J. (2001). Influence of the aglycone region of the substrate binding cleft of *Pseudomonas xylanase* 10A on catalysis. *Biochemistry, 40*(25), 7404-7409. doi:10.1021/bi002704r.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., . . . Yin, Y. (2021). dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res, 49*(D1), D523-D528. doi:10.1093/nar/gkaa742.

Barron, C., Robert, P., Guillon, F., Saulnier, L., & Rouau, X. (2006). Structural heterogeneity of wheat arabinoxylans revealed by Raman spectroscopy. *Carbohydr Res, 341*(9), 1186-1191. doi:10.1016/j.carres.2006.03.025.

Bastien, G. (2013). Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnology for biofuels*, 6:78. doi: 10.1186/1754-

6834-6-78.

Bayer, E. A., Belaich, J. P., Shoham, Y., & Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol, 58*, 521-554. doi:10.1146/annurev.micro.57.030502.091022.

Bazanella, M., Maier, T. V., Clavel, T., Lagkouvardos, I., Lucio, M., Maldonado-Gomez, M. X., . . . Haller, D. (2017). Randomized controlled trial on the impact of early-life intervention with *Bifidobacteria* on the healthy infant fecal microbiota and metabolome. *Am J Clin Nutr, 106*(5), 1274-1286. doi:10.3945/ajcn.117.157529.

Bazzani, R. P., Cai, Y., Hebel, H. L., Hyde, S. C., & Gill, D. R. (2011). The significance of plasmid DNA preparations contaminated with bacterial genomic DNA on inflammatory responses following delivery of lipoplexes to the murine lung. *Biomaterials, 32*(36), 9854-9865. doi:10.1016/j.biomaterials.2011.08.092.

Benedetti, M., Vecchi, V., Betterle, N., Natali, A., Bassi, R., & Dall'Osto, L. (2019). Design of a highly thermostable hemicellulose-degrading blend from *Thermotoga neapolitana* for the treatment of lignocellulosic biomass. *J Biotechnol, 296*, 42-52. doi:10.1016/j.jbiotec.2019.03.005.

Bernhard Seiboth, C. I. a. V. S.-S. (2011). *Trichoderma reesei*: A fungal enzyme producer for cellulosic biofuels. *Biofuel production*. doi:DOI: 10.5772/16848. doi: 10.5772/16848.

Berry, D., Ben Mahfoudh, K., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol, 77*(21), 7846-7849. doi:10.1128/AEM.05220-11.

Biely, P., de Vries, R. P., Vrsanská, M., & Visser, J. (2000). Inverting character of alpha-glucuronidase A from *Aspergillus tubingensis*. *Biochim Biophys Acta, 1474*(3), 360-364. doi:10.1016/s0304-4165(00)00029-5.

Biely, P., Singh, S., & Puchart, V. (2016). Towards enzymatic breakdown of complex plant xylan structures: State of the art. *Biotechnol Adv, 34*(7), 1260-1274. doi:10.1016/j.biotechadv.2016.09.001.

Höcker, B., Jürgens, C., Wilmanns, M., Sterner, R. (2001). Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold. *current opinion in biotechnology, 12*, 376-381. doi: 10.1016/s0958-1669(00)00230-5.

Biswabandhu, C., Sanjit, M., & Debabrata, M. (2019). Biodegradation study of cellulose bearing synthetic wastewater in activated sludge system. *J. Indian Chem. Soc., 96 (4)*, 527-532.

Bonnin, E., Garnier, C., & Ralet, M. C. (2014). Pectin-modifying enzymes and pectin-derived materials: applications and impacts. *Appl Microbiol Biotechnol, 98*(2), 519-532. doi:10.1007/s00253-013-5388-6.

Bonnin E., Pelloux J. (2020) Pectin Degrading Enzymes. In: Kontogiorgos V. (eds) Pectin: Technological and Physiological Properties. Springer, Cham. doi:10.1007/978-3-030-53421-9_3.

Bouraoui, H., Desrousseaux, M. L., Ioannou, E., Alvira, P., Manai, M., Remond, C., . . . O'Donohue, M. J. (2016). The GH51 alpha-L-arabinofuranosidase from *Paenibacillus* sp. THS1 is multifunctional, hydrolyzing main-chain and side-chain glycosidic bonds in heteroxylans. *Biotechnol Biofuels, 9*, 140. doi:10.1186/s13068-016-0550-x.

Broeker, J., Mechelke, M., Baudrexl, M., Mennerich, D., Hornburg, D., Mann, M., . . . Zverlov, V. V. (2018). The hemicellulose-degrading enzyme system of the thermophilic bacterium *Clostridium stercorarium*: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. *Biotechnol Biofuels, 11*, 229. doi:10.1186/s13068-018-1228-3.

Carapito, R., Imberty, A., Jeltsch, J. M., Byrns, S. C., Tam, P. H., Lowary, T. L., . . . Phalip, V. (2009). Molecular basis of arabinobio-hydrolase activity in phytopathogenic fungi: crystal structure and catalytic mechanism of *Fusarium graminearum* GH93 exo-alpha-L-arabinanase. *J Biol Chem, 284*(18), 12285-12296. doi:10.1074/jbc.M900439200.

Cartmell, A., McKee, L. S., Pena, M. J., Larsbrink, J., Brumer, H., Kaneko, S., . . . Marles-Wright, J. (2011). The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J Biol Chem, 286*(17), 15483-15495. doi:10.1074/jbc.M110.215962.

Cecchini, D. A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Dore, J., . . . Potocki-Veronese, G. (2013). Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One, 8*(9), e72766. doi:10.1371/journal.pone.0072766.

Chen, X., Li, Z., Zhang, X., Hu, F., Ryu, D. D., & Bao, J. (2009). Screening of oleaginous yeast strains tolerant to lignocellulose degradation compounds. *Appl Biochem Biotechnol, 159*(3), 591-604. doi:10.1007/s12010-008-8491-x.

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

Consortium, C. A. (2018). Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology, 28*(1), 3-8. doi:10.1093/glycob/cwx089.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., Helynck, G., . . . Pernodet, J. L. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol, 69*(1), 49-55. doi:10.1128/aem.69.1.49-55.2003.

Culligan, E. P., Sleator, R. D., Marchesi, J. R., & Hill, C. (2012). Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J, 6*(10), 1916-1925. doi:10.1038/ismej.2012.38.

Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E., Pudlo, N. A., . . . Gilbert, H. J. (2015). Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature, 517*(7533), 165-169. doi:10.1038/nature13995.

Koullas, D. P., Christakopoulos, P., Kekos, D., Macris, B. J., Koukios, E. G. (1992). Correlating the effect of pretreatment on the enzymatic hydrolysis of straw. *Biotechnol Bioeng, 39*(1), 113-116. doi: 10.1002/bit.260390116.

Daas, M. J. A., Martinez, P. M., van de Weijer, A. H. P., van der Oost, J., de Vos, W. M., Kabel, M. A., & van Kranenburg, R. (2017). Biochemical characterization of the xylan hydrolysis profile of the extracellular endo-xylanase from *Geobacillus thermodenitrificans* T12. *BMC Biotechnol, 17*(1), 44. doi:10.1186/s12896-017-0357-2.

de Camargo, B. R., Claassens, N. J., Quirino, B. F., Noronha, E. F., & Kengen, S. W. M. (2018). Heterologous expression and characterization of a putative glycoside hydrolase family 43

arabinofuranosidase from *Clostridium thermocellum* B8. *Enzyme Microb Technol, 109*, 74-83. doi:10.1016/j.enzmictec.2017.09.014.

De La Mare, M., Guais, O., Bonnin, E., Weber, J., & Francois, J. M. (2013). Molecular and biochemical characterization of three GH62 α-L-arabinofuranosidases from the soil deuteromycete *Penicillium funiculosum*. *Enzyme Microb Technol, 53*(5), 351-358. doi:10.1016/j.enzmictec.2013.07.008.

de Lima Brossi, M. J., Jimenez, D. J., Cortes-Tolalpa, L., & van Elsas, J. D. (2016). Soil-derived microbial consortia enriched with different plant biomass reveal distinct players acting in lignocellulose degradation. *Microb Ecol, 71*(3), 616-627. doi:10.1007/s00248-015-0683-7.

Demain, A. L., Newcomb, M., & Wu, J. H. (2005). Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev, 69*(1), 124-154. doi:10.1128/MMBR.69.1.124-154.2005.

Den Haan, R., Rose, S. H., Lynd, L. R., & van Zyl, W. H. (2007). Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab Eng, 9*(1), 87-94. doi:10.1016/j.ymben.2006.08.005.

Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., . . . Derewenda, Z. S. (1994). Crystal structure, at 2.6-A resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of beta-1,4-D-glycanases. *Journal of Biological Chemistry, 269*(33), 20811-20814. doi:10.1016/s0021-9258(17)31892-6.

Ding, S. Y., Bayer, E. A. (2020). Understanding cellulosome interaction with cellulose by high-resolution imaging. *ACS Cent Sci, 6*(7), 1034-1036. doi:10.1021/acscentsci.0c00662.

Dodd, D., & Cann, I. K. (2009). Enzymatic deconstruction of xylan for biofuel production. *Glob Change Biol Bioenergy, 1*(1), 2-17. doi:10.1111/j.1757-1707.2009.01004.x.

Dos Santos, C. R., de Giuseppe, P. O., de Souza, F. H. M., Zanphorlin, L. M., Domingues, M. N., Pirolla, R. A. S., . . . Murakami, M. T. (2018). The mechanism by which a distinguishing arabinofuranosidase can cope with internal di-substitutions in arabinoxylans. *Biotechnol Biofuels, 11*, 223. doi:10.1186/s13068-018-1212-y.

Dutilh, B. E., 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). doi:10.1371/journal.pone.0052459.

Džunková, M., D'Auria, G., Perez-Villarroya, D., & Moya, A. (2012). Hybrid sequencing approach applied to human fecal metagenomic clone libraries revealed clones with potential biotechnological applications. *PLoS One, 7*(10), e47654. doi:10.1371/journal.pone.0047654.

Lattová, E., Ebringerova, A., Toman, R., Kačuráková, M. (1992). Controlled Depolymerization of 4-O-Methyl-D-glucurono-D-xylan Isolated from Wood of Beech. *Chemical papers, 46*(1), 66-69.

Edwards, M. C., & Doran-Peterson, J. (2012). Pectin-rich biomass as feedstock for fuel ethanol production. *Appl Microbiol Biotechnol, 95*(3), 565-575. doi:10.1007/s00253-012-4173-2.

Ejby, M., Guskov, A., Pichler, M. J., Zanten, G. C., Schoof, E., Saburi, W., . . . Abou Hachem, M. (2019). Two binding proteins of the ABC transporter that confers growth of *Bifidobacterium animalis* subsp. lactis ATCC27673 on beta-mannan possess distinct manno-oligosaccharide-binding

profiles. *Mol Microbiol, 112*(1), 114-130. doi:10.1111/mmi.14257.

Ekkers, D. M., Cretoiu, M. S., Kielak, A. M., & Elsas, J. D. (2012). The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol, 93*(3), 1005-1020. doi:10.1007/s00253-011-3804-3.

Eneyskaya, E. V., Bobrov, K. S., Kashina, M. V., Borisova, A. S., & Kulminskaya, A. A. (2020). A novel acid-tolerant beta-xylanase from *Scytalidium candidum* 3C for the synthesis of *O*-nitrophenyl xylooligosaccharides. *J Basic Microbiol, 60*(11-12), 971-982. doi:10.1002/jobm.202000303.

Sanger, F. , Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*(12), 5463-5467. doi: 10.1073/pnas.74.12.5463.

Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., . . . Golyshin, P. N. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol, 7*(12), 1996-2010. doi:10.1111/j.1462-2920.2005.00920.x.

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact, 17*(1), 52. doi:10.1186/s12934-018-0901-3.

Furukawa, Y., Horiuchi, M., & Kakegawa, T. (2013). Selective stabilization of ribose by borate. *Orig Life Evol Biosph, 43*(4-5), 353-361. doi:10.1007/s11084-013-9350-5.

Beldman, G., Schols, H. A., Pitson, S. M., Searle-van Leeuwen, M. J. F., Voragen, A. G. J. (1997). Arabinans and arabinan degrading enzymes. *Advances in macromolecular carbohydrate research*, Vol 1/ Sturgeon, R. J., -p. 1-64.

Gao, J., Zhao, Y., Zhang, G., Li, Y., & Li, Q. (2018). Production optimization, purification, expression, and characterization of a novel α-L-arabinofuranosidase from *Paenibacillus polymyxa*. *Electronic Journal of Biotechnology, 36*, 24-33. doi:10.1016/j.ejbt.2018.09.002.

Gavande, P. V., Basak, A., Sen, S., Lepcha, K., Murmu, N., Rai, V., . . . Ghosh, S. (2021). Functional characterization of thermotolerant microbial consortium for lignocellulolytic enzymes with central role of *Firmicutes* in rice straw depolymerization. *Sci Rep, 11*(1), 3032. doi:10.1038/s41598-021-82163-x.

Geng, A., Wu, J., Xie, R., Wang, H., Wu, Y., Li, X., . . . Sun, J. (2019). Highly thermostable GH51 alpha-arabinofuranosidase from *Hungateiclostridium clariflavum* DSM 19732. *Appl Microbiol Biotechnol, 103*(9), 3783-3793. doi:10.1007/s00253-019-09753-8.

Gloux, K., Berteau, O., El Oumami, H., Beguet, F., Leclerc, M., & Dore, J. (2011). A metagenomic beta-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc Natl Acad Sci U S A, 108 Suppl 1*, 4539-4546. doi:10.1073/pnas.1000066107.

Gong, W., Zhang, H., Tian, L., Liu, S., Wu, X., Li, F., & Wang, L. (2016). Determination of the modes of action and synergies of xylanases by analysis of xylooligosaccharide profiles over time using fluorescence-assisted carbohydrate electrophoresis. *Electrophoresis, 37*(12), 1640-1650. doi:10.1002/elps.201600041.

Goyal, A., Ahmed, S., Sharma, K., Gupta, V., Bule, P., Alves, V. D., . . . Najmudin, S. (2016). Molecular determinants of substrate specificity revealed by the structure of *Clostridium thermocellum* arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16. *Acta Crystallogr D Struct Biol, 72*(Pt 12), 1281-1289. doi:10.1107/S205979831601737X.

Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W., & Brumer, H. (2017). Polysaccharide utilization loci: fueling microbial communities. *J Bacteriol, 199*(15). doi:10.1128/JB.00860-16.

Gullert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B., . . . Streit, W. R. (2016). Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol Biofuels, 9*, 121. doi:10.1186/s13068-016-0534-x.

Han, Z., Shang-Guan, F., & Yang, J. (2019). Molecular and biochemical characterization of a bimodular xylanase from *Marinifilaceae Bacterium* strain SPP2. *Front Microbiol, 10*, 1507. doi:10.3389/fmicb.2019.01507.

Hehemann, J. H., Reintjes, G., Klassen, L., Smith, A. D., Ndeh, D., Arnosti, C., . . . Abbott, D. W. (2019). Single cell fluorescence imaging of glycan uptake by intestinal bacteria. *ISME J, 13*(7), 1883-1889. doi:10.1038/s41396-019-0406-z.

Heinze, T. (2005). *Polysaccharides I structure, characterisation and use*. Advances in Polymer Science, ISSN 0065-3195, doi: 10.1007/b136812.

Heinze, T., & Liebert, T. (2012). Celluloses and polyoses/hemicelluloses. *Polymer Science: A Comprehensive Reference*, 10, 83-152. doi: 10.1016/B978-0-444-53349-4.00255-7.

van Tilbeurgh, H., Claeyssens, M., de Bruyne, C. K. (1982). The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. *FEBS LETTERS, 149*(1), 152-156. doi:10.1016/0014-5793(82)81092-2.

Herrmann, K. M., & Weaver, L. M. (1999). The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol, 50*, 473-503. doi:10.1146/annurev.arplant.50.1.473.

Hoondal, G. S., Tiwari, R. P., Tewari, R., Dahiya, N., & Beg, Q. K. (2002). Microbial alkaline pectinases and their industrial applications: a review. *Appl Microbiol Biotechnol, 59*(4-5), 409-418. doi:10.1007/s00253-002-1061-1.

Hu, J., Saddler, J. N. (2018). Why does GH10 xylanase have better performance than GH11 xylanase for the deconstruction of pretreated biomass? *Biomass and Bioenergy, 110*, 13-16. doi:10.1016/j.biombioe.2018.01.007.

Huang, D., Liu, J., Qi, Y., Yang, K., Xu, Y., & Feng, L. (2017). Synergistic hydrolysis of xylan using novel xylanases, β-xylosidases and an α-L-arabinofuranosidase from *Geobacillus thermodenitrificans* NG80-2. *Appl Microbiol Biotechnol, 101*(15), 6023-6037. doi:10.1007/s00253-017-8341-2.

Hutnan M, Drtil M, & L, M. (2000). Anaerobic biodegradation of sugar beet pulp. *Biodegradation, 11*, 203-211. doi: 10.1023/a:1011139621329.

Hyeon, J. E., Jeon, S. D., & Han, S. O. (2013). Cellulosome-based, Clostridium-derived multi-functional enzyme complexes for advanced biotechnology tool development: advances and applications. *Biotechnol Adv, 31*(6), 936-944. doi:10.1016/j.biotechadv.2013.03.009.

Colquhoun, I. J., Ralet, M. C., Thibault, J. F., Faulds, C. B., Williamson, G. (1994). Structure identification of feruloylated oligosaccharides from sugar-beet pulp by NMR spectroscopy. *Carbohydrate Research, 263*(243-256). doi: 10.1016/0008-6215(94)00176-6.

İlgü, H., Sürmeli, Y., & Şanlı-Mohamed, G. (2018). A thermophilic α-L-arabinofuranosidase from

Geobacillus vulcani GS90: heterologous expression, biochemical characterization, and its synergistic action in fruit juice enrichment. *European Food Research and Technology, 244*(9), 1627-1636. doi:10.1007/s00217-018-3075-7.

Inacio, J. M., Correia, I. L., & de Sa-Nogueira, I. (2008a). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology, 154*(Pt 9), 2719-2729. doi:10.1099/mic.0.2008/018978-0.

Inacio, J. M., Correia, I. L., & de Sa-Nogueira, I. (2008b). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology (Reading), 154*(Pt 9), 2719-2729. doi:10.1099/mic.0.2008/018978-0.

Alkorta, I., Garbisu, C., Llama, M., J., Serra, J., L. (1998). Industrial applications of pectic enzymes: a review. *Process Biochemistry, 33*(1), 21-28. doi.org/10.1016/S0032-9592(97)00046-0.

Jaenicke, S., Ander, C., Bekel, T., Bisdorf, R., Droge, M., Gartemann, K. H., . . . Goesmann, A. (2011). Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS One, 6*(1), e14519. doi:10.1371/journal.pone.0014519.

Jeon, J. H., Kim, S. J., Lee, H. S., Cha, S. S., Lee, J. H., Yoon, S. H., . . . Lee, J. H. (2011). Novel metagenome-derived carboxylesterase that hydrolyzes beta-lactam antibiotics. *Appl Environ Microbiol, 77*(21), 7830-7836. doi:10.1128/AEM.05363-11.

Jia, Y., Wilkins, D., Lu, H., Cai, M., Lee, P. K. H., & Stams, A. J. M. (2016). Long-term enrichment on cellulose or xylan causes functional and taxonomic convergence of microbial communities from anaerobic digesters. *Appl Environ Microbiol, 82*(5), 1519-1529. doi:10.1128/aem.03360-15.

Jiang, C., Ma, G., Li, S., Hu, T., Che, Z., Shen, P., . . . Wu, B. (2009). Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *J Microbiol, 47*(5), 542-548. doi:10.1007/s12275-009-0024-y.

Jiao, J. Y., Wang, H. X., Zeng, Y., & Shen, Y. M. (2006). Enrichment for microbes living in association with plant tissues. *J Appl Microbiol, 100*(4), 830-837. doi:10.1111/j.1365-2672.2006.02830.x.

Jimenez, D. J., de Lima Brossi, M. J., Schuckel, J., Kračun, S. K., Willats, W. G., & van Elsas, J. D. (2016). Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl Microbiol Biotechnol, 100*(24), 10463-10477. doi:10.1007/s00253-016-7713-3.

Jimenez, D. J., Korenblum, E., & van Elsas, J. D. (2014). Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. *Appl Microbiol Biotechnol, 98*(6), 2789-2803. doi:10.1007/s00253-013-5253-7.

Wooley, J. C., Godzik, A., Friedberg, I. (2010). A primer on metagenomics. *Pols Comutational Biology, 6*(2), e1000667. doi:10.1371/journal.

Johnson, E. A., Madia, A., & Demain, A. L. (1981). Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. *Appl Environ Microbiol, 41*(4), 1060-1062. doi:10.1128/aem.41.4.1060-1062.1981.

Knox, J.P. (1995). The extracellular matrix in higher plants. 4. Developmentally regulated proteoglycans and glycoproteins of the plant cell surface. *FASEB, 9*, 1004-1012. doi:10.1096/fasebj.9.11.7544308.

Grondin, J. M., Tamura, K., Dejéan, G., Abbott, D. W., Brumer, H. (2017). Polysaccharide utilization loci: Fueling microbial communities. *J Bacteriol, 199*(15), e00860-00816. doi: 10.1128/JB.00860-16.

Kaji, A., & Shimokawa, K. (2014). New exo-type arabinase from *Erwinia carotovora* IAM 1024. *Agricultural and Biological Chemistry, 48*(1), 67-72. doi:10.1080/00021369.1984.10866094.

Khandeparker, R., & Numan, M. T. (2008). Bifunctional xylanases and their potential use in biotechnology. *J Ind Microbiol Biotechnol, 35*(7), 635-644. doi:10.1007/s10295-008-0342-9.

Koch, H., Durwald, A., Schweder, T., Noriega-Ortega, B., Vidal-Melgosa, S., Hehemann, J. H., . . . Wietz, M. (2019). Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. *ISME J, 13*(1), 92-103. doi:10.1038/s41396-018-0252-4.

Koeck, D. E., Ludwig, W., Wanner, G., Zverlov, V. V., Liebl, W., & Schwarz, W. H. (2015). *Herbinix hemicellulosilytica* gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int J Syst Evol Microbiol, 65*(8), 2365-2371. doi:10.1099/ijs.0.000264.

Kont, R., Kurasin, M., Teugjas, H., & Valjamae, P. (2013). Strong cellulase inhibitors from the hydrothermal pretreatment of wheat straw. *Biotechnol Biofuels, 6*(1), 135. doi:10.1186/1754-6834-6-135.

Koseki, T., Miwa, Y., Mese, Y., Miyanaga, A., Fushinobu, S., Wakagi, T., . . . Hashizume, K. (2006). Mutational analysis of N-glycosylation recognition sites on the biochemical properties of *Aspergillus kawachii* alpha-L-arabinofuranosidase 54. *Biochim Biophys Acta, 1760*(9), 1458-1464. doi:10.1016/j.bbagen.2006.04.009.

Koutaniemi, S., & Tenkanen, M. (2016). Action of three GH51 and one GH54 alpha-arabinofuranosidases on internally and terminally located arabinofuranosyl branches. *J Biotechnol, 229*, 22-30. doi:10.1016/j.jbiotec.2016.04.050.

Kracher, D., Oros, D., Yao, W., Preims, M., Rezic, I., Haltrich, D., . . . Ludwig, R. (2014). Fungal secretomes enhance sugar beet pulp hydrolysis. *Biotechnol J, 9*(4), 483-492. doi:10.1002/biot.201300214.

Kračun, S. K., Schuckel, J., Westereng, B., Thygesen, L. G., Monrad, R. N., Eijsink, V. G. H., & Willats, W. G. T. (2015). A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes. *Biotechnol Biofuels, 8*, 70. doi:10.1186/s13068-015-0250-y.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli. Science, 324*(5924), 255-258. doi:10.1126/science.1170160.

Kühnel, S. (2011). Characterization of cell wall degrading enzymes from *Chrysosporium lucknowense* C1 and their use to degrade sugar beet pulp. PhD thesis Wageningen University, NL (2011). ISBN: 978-90-8585-978-9.

Kühnel, S., Hinz, S. W., Pouvreau, L., Wery, J., Schols, H. A., & Gruppen, H. (2010). *Chrysosporium lucknowense* arabinohydrolases effectively degrade sugar beet arabinan. *Bioresour Technol, 101*(21), 8300-8307. doi:10.1016/j.biortech.2010.05.070.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version

7.0 for bigger datasets. *Mol Biol Evol, 33*(7), 1870-1874. doi:10.1093/molbev/msw054.

Lagaert, S., Pollet, A., Courtin, C. M., & Volckaert, G. (2014). β-xylosidases and α-L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol Adv, 32*(2), 316-332. doi:10.1016/j.biotechadv.2013.11.005.

Lagkouvardos, I., Fischer, S., Kumar, N., & Clavel, T. (2017). Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ, 5*, e2836. doi:10.7717/peerj.2836.

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., & Clavel, T. (2016). IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep, 6*, 33721. doi:10.1038/srep33721.

Lagkouvardos, I., Klaring, K., Heinzmann, S. S., Platz, S., Scholz, B., Engel, K. H., . . . Clavel, T. (2015). Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men. *Mol Nutr Food Res, 59*(8), 1614-1628. doi:10.1002/mnfr.201500125.

Lakhundi, S., Siddiqui, R., & Khan, N. A. (2015). Cellulose degradation: a therapeutic strategy in the improved treatment of *Acanthamoeba infections*. *Parasit Vectors, 8*, 23. doi:10.1186/s13071-015-0642-7.

Lam, K. N., Hall, M. W., Engel, K., Vey, G., Cheng, J., Neufeld, J. D., & Charles, T. C. (2014). Evaluation of a pooled strategy for high-throughput sequencing of cosmid clones from metagenomic libraries. *PLoS One, 9*(6), e98968. doi:10.1371/journal.pone.0098968.

Lansky, S., Salama, R., Dann, R., Shner, I., Manjasetty, B. A., Belrhali, H., . . . Shoham, G. (2014). Cloning, purification and preliminary crystallographic analysis of Ara127N, a GH127 beta-L-arabinofuranosidase from *Geobacillus stearothermophilus* T6. *Acta Crystallogr F Struct Biol Commun, 70*(Pt 8), 1038-1045. doi:10.1107/S2053230X14012680.

Lansky, S., Salama, R., Solomon, V. H., Belrhali, H., Shoham, Y., & Shoham, G. (2013). Crystallization and preliminary crystallographic analysis of Abp, a GH27 beta-L-arabinopyranosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun, 69*(Pt 6), 695-699. doi:10.1107/S1744309113013705.

Le, Y., & Wang, H. (2014). High-level soluble expression of a thermostable xylanase from thermophilic fungus *Thermomyces lanuginosus* in *Escherichia coli* via fusion with OsmY protein. *Protein Expr Purif, 99*, 1-5. doi:10.1016/j.pep.2014.03.004.

Lebuhn, M., Derenko, J., Rademacher, A., Helbig, S., Munk, B., Pechtl, A., . . . Klocke, M. (2016). DNA and RNA extraction and Quantitative Real-Time PCR-based assays for biogas biocenoses in an interlaboratory comparison. *Bioengineering, 3*(1). doi:10.3390/bioengineering3010007.

Lee, D. J., Show, K. Y., & Wang, A. (2013). Unconventional approaches to isolation and enrichment of functional microbial consortium--a review. *Bioresour Technol, 136*, 697-706. doi:10.1016/j.biortech.2013.02.075.

Leijdekkers, A. G., Bink, J. P., Geutjes, S., Schols, H. A., & Gruppen, H. (2013). Enzymatic saccharification of sugar beet pulp for the production of galacturonic acid and arabinose; a study on the impact of the formation of recalcitrant oligosaccharides. *Bioresour Technol, 128*, 518-525. doi:10.1016/j.biortech.2012.10.126.

Leis, B., Angelov, A., & Liebl, W. (2013). Screening and expression of genes from metagenomes. *Adv Appl Microbiol, 83*, 1-68. doi:10.1016/B978-0-12-407678-5.00001-5.

Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V. T., Lauinger, B., . . . Liebl, W. (2015). Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front Microbiol, 6*, 275. doi:10.3389/fmicb.2015.00275.

Lenz, L. L., Mohammadi, S., Geissler, A., & Portnoy, D. A. (2003). SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *Proc Natl Acad Sci U S A, 100*(21), 12432-12437. doi:10.1073/pnas.2133653100.

Li, P., Lin, W., Liu, X., Wang, X., & Luo, L. (2016). Environmental factors affecting microbiota dynamics during traditional solid-state fermentation of chinese daqu starter. *Front Microbiol, 7*, 1237. doi:10.3389/fmicb.2016.01237.

Liebl, W., Angelov, A., Juergensen, J., Chow, J., Loeschcke, A., Drepper, T., . . . Jaeger, K. E. (2014). Alternative hosts for functional (meta)genome analysis. *Appl Microbiol Biotechnol, 98*(19), 8099-8109. doi:10.1007/s00253-014-5961-7.

Lin, C., & Stahl, D. A. (1995). Comparative analyses reveal a highly conserved endoglucanase in the cellulolytic genus *Fibrobacter*. *J Bacteriol, 177*(9), 2543-2549. doi:10.1128/jb.177.9.2543-2549.1995.

Liu, L., Lv, A. P., Ming, Y. Z., Jiao, J. Y., Xiao, M., Lin, J. G., . . . Li, W. J. (2020). A xylan-degrading thermophilic and obligate anaerobe *Xylanivirga thermophila* gen. nov., sp. nov., isolated from an anammox dominant wastewater treatment plant, and proposal of *Xylanivirgaceae* fam. nov. *Anaerobe, 61*, 102075. doi:10.1016/j.anaerobe.2019.102075.

Liu, Y., Vanderhaeghen, S., Feiler, W., Angelov, A., Baudrexl, M., Zverlov, V., & Liebl, W. (2021). Characterization of two α-L-arabinofuranosidases from *Acetivibrio mesophilus* and their synergistic effect in degradation of arabinose-containing substrates. *Microorganisms, 9*(7). doi:10.3390/microorganisms9071467.

Liu, Y., Angelov, A., Feiler, W., Baudrexl, M., Zverlov, V., Liebl, W., Vanderhaeghen, S. (2022). Arabinan saccharification by biogas reactor metagenome-derived arabinosyl hydrolases. Manuscript submitted.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res, 42*(Database issue), D490-495. doi:10.1093/nar/gkt1178.

Lu, Y. T., Chen, X. B., Zhou, P., & Li, Z. H. (2005). Screening on oil-decomposing microorganisms and application in organic waste treatment machine. *J Environ Sci (China), 17*(3), 440-444. PMID: 16083120.

Lueking, A., Holz, C., Gotthold, C., Lehrach, H., & Cahill, D. (2000). A system for dual protein expression in *Pichia pastoris* and *Escherichia coli*. *Protein Expr Purif, 20*(3), 372-378. doi:10.1006/prep.2000.1317.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev, 66*(3), 506-577, table of contents. doi:10.1128/MMBR.66.3.506-577.2002.

Ma, R., Bai, Y., Huang, H., Luo, H., Chen, S., Fan, Y., . . . Yao, B. (2017). Utility of thermostable xylanases of *Mycothermus thermophilus* in generating prebiotic xylooligosaccharides. *J Agric Food Chem, 65*(6), 1139-1145. doi:10.1021/acs.jafc.6b05183.

O'Neill, M., Albersheim, P., Darvill, A. (1990). The pectic polysaccharides of primary cell walls. *Methods in plant biochemistry, 2*, 415-441. doi:10.1016/B978-0-12-461012-5.50018-5.

Ralet, M. C., Thibault, J. F., Faulds, C. B., Willamson, G. (1994). Isolation and purification of feruloylated oligosaccharides from cell walls of sugar-beet pulp. *Carbohydrate Research, 263*, 227-241. doi: 10.1016/0008-6215(94)00175-8.

Martinez, A., Kolvek, S. J., Yip, C. L., Hopke, J., Brown, K. A., MacNeil, I. A., & Osburne, M. S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microbiol, 70*(4), 2452-2463. doi:10.1128/AEM.70.4.2452-2463.2004.

McCleary, B. V., McKie, V. A., Draga, A., Rooney, E., Mangan, D., & Larkin, J. (2015). Hydrolysis of wheat flour arabinoxylan, acid-debranched wheat flour arabinoxylan and arabino-xylo-oligosaccharides by beta-xylanase, alpha-L-arabinofuranosidase and beta-xylosidase. *Carbohydr Res, 407*, 79-96. doi:10.1016/j.carres.2015.01.017.

McIlvaine, T. C. (1921). A buffer solution for colorimetric comparison. *Journal of Biological Chemistry, 49*(1), 183-186. doi:10.1016/s0021-9258(18)86000-8.

Mechelke, M., Koeck, D. E., Broeker, J., Roessler, B., Krabichler, F., Schwarz, W. H., . . . Liebl, W. (2017). Characterization of the arabinoxylan-degrading machinery of the thermophilic bacterium *Herbinix hemicellulosilytica*-Six new xylanases, three arabinofuranosidases and one xylosidase. *J Biotechnol, 257*, 122-130. doi:10.1016/j.jbiotec.2017.04.023.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet, 11*(1), 31-46. doi:10.1038/nrg2626.

Mewis, K., Lenfant, N., Lombard, V., & Henrissat, B. (2016). Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl Environ Microbiol, 82*(6), 1686-1692. doi:10.1128/AEM.03453-15.

Madadi, M., Tu, Y., Abbas, A. (2017). Recent status on enzymatic saccharification of lignocellulosic biomass for bioethanol production. *Electronic Journal of Biology, 13*(2), 135-143.

Micard, V., Renard, C. M. G. C., & Thibault, J. F. (1996). Enzymatic saccharification of sugar beet pulp. *Enzyme Microb Technol, 19* (3), 162-170. doi:10.1016/0141-0229(95)00224-3.

Micard, V., Renard, C. M. G. C., & Thibault, J. F. (1997). Influence of pretreatments on enzymic degradation of a cellulose-rich residue from sugar-beet pulp. *LWT- Food Science and Technology, 30*, 284-291. doi: 10.1006/FSTL.1996.0182.

Huisman, M. M., Brül, L. P., Thomas-Oates, J. E., Haverkamp, J., Schols, H. A., Voragen, A. G. (2001). The occurrence of internal 1,5-linked arabinofuranose and arabinopyranose residues in arabinogalactan side chains from soybean pectic substances. *Carbohydrate Research, 2001*(330), 103-114. doi: 10.1016/S0008-6215(00)00269-X.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E. L. L., . . .

Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*(D1), D412-D419. doi:10.1093/nar/gkaa913.

Mitchell, P. (1967). Translocations through natural membranes. *Advances in Enzymology and Related Areas of Molecular Biology, 29*(33-87). doi: 10.1002/9780470122747.ch2.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Curr Opin Plant Biol, 11*(3), 266-277. doi:10.1016/j.pbi.2008.03.006. doi: 10.1016/j.pbi.2008.03.006.

Moneo-Sanchez, M., Vaquero-Rodriguez, A., Hernandez-Nistal, J., Albornos, L., Knox, P., Dopico, B., . . . Martin, I. (2020). Pectic galactan affects cell wall architecture during secondary cell wall deposition. *Planta, 251*(5), 100. doi:10.1007/s00425-020-03394-2.

Montella, S., Amore, A., & Faraco, V. (2016). Metagenomics for the development of new biocatalysts to advance lignocellulose saccharification for bioeconomic development. *Crit Rev Biotechnol, 36*(6), 998-1009. doi:10.3109/07388551.2015.1083939.

Morais, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., . . . Bayer, E. A. (2010). Contribution of a xylan-binding module to the degradation of a complex cellulosic substrate by designer cellulosomes. *Appl Environ Microbiol, 76*(12), 3787-3796. doi:10.1128/AEM.00266-10.

Mroueh, M., Aruanno, M., Borne, R., de Philip, P., Fierobe, H. P., Tardif, C., & Pages, S. (2019). The xyl-doc gene cluster of *Ruminiclostridium cellulolyticum* encodes GH43- and GH62-alpha-L-arabinofuranosidases with complementary modes of action. *Biotechnol Biofuels, 12*, 144. doi:10.1186/s13068-019-1483-y.

Nakamura, A., Ishiwata, D., Visootsat, A., Uchiyama, T., Mizutani, K., Kaneko, S., . . . Iino, R. (2020). Domain architecture divergence leads to functional divergence in binding and catalytic domains of bacterial and fungal cellobiohydrolases. *J Biol Chem, 295*(43), 14606-14617. doi:10.1074/jbc.RA120.014792.

O' Neill, M., Albersheim, P., & Darvill, A. (1990). The pectic polysaccharides of primary cell walls. *Carbohydrates*, 415-441. doi:10.1016/B978-0-12-461012-5.50018-5.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 27*(1), 29-34. doi:10.1093/nar/27.1.29.

Oumer, O. J. (2017). Pectinase: substrate, production and their biotechnological applications. *IJEAB, 2*(3), 1007-1014. doi:10.22161/ijeab/2.3.1.

Biely, P., Vršanská, M.,Tenkanen, M., Kluepfel, D. Tenkanen, D. Kluepfel. (1997a). Endo-beta-1,4-xylanase families: differences in catalytic properties. *Journal of Biotechnology, 57*, 151-166. doi: 10.1016/s0168-1656(97)00096-5.

Biely, P., Vrsanská, M., Kratky, Z. (1981). Mechanism of substrate digestion by endo-1,4-beta-xylanase of *Cryptococcus albidus*. Lysozyme-type pattern of action. *Eur J Biochem, 19*(3), 565-571. doi: 10.1111/j.1432-1033.1981.tb05645.x.

Christakopoulos, P., Katapodis, P., Kalogeris, E., Kekos, D.,…Skaltsa, H. (2003). Antimicrobial activity of acidic xylo-oligosaccharides produced by family 10 and 11 endoxylanases. *Biological Macromolecules, 31*, 171-175. doi: 10.1016/s0141-8130(02)00079-x.

Pantazaki, A. A., Pritsa, A. A., & Kyriakidis, D. A. (2002). Biotechnologically relevant enzymes from *Thermus thermophilus*. *Appl Microbiol Biotechnol, 58*(1), 1-12. doi:10.1007/s00253-001-0843-

1.

Paulsen I T, B. A. M., Saier M H. (1997). Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology, 143*, 2685-2699. doi: 10.1099/00221287-143-8-2685.

Perez, J., Munoz-Dorado, J., de la Rubia, T., & Martinez, J. (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol, 5*(2), 53-63. doi:10.1007/s10123-002-0062-3.

Pérez, R., & Eyzaguirre, J. (2016). *Aspergillus fumigatus* produces two arabinofuranosidases from glycosyl hydrolase family 62: comparative properties of the recombinant enzymes. *Appl Biochem Biotechnol, 179*(1), 143-154. doi:10.1007/s12010-016-1984-0.

Biely, P., Krátký. Z., Vrsanska, M. (1981). Substrate binding site of endo-1,4-beta-xylanase of the yeast *Cryptococcus albidus. Eur J Biochem, 119*, 559-564. doi: 10.1111/j.1432-1033.1981.tb05644.x.

Pierre Beguin, J.-P. A. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews, 13*, 25-58. doi: 10.1111/j.1574-6976.1994.tb00033.x.

Pinkowska, H., Krzywonos, M., Wolak, P., & Zlocinska, A. (2019). Pectin and neutral monosaccharides production during the simultaneous hydrothermal extraction of waste biomass from refining of sugar-optimization with the use of doehlert design. *Molecules, 24*(3). doi:10.3390/molecules24030472.

Poehlein, A., Zverlov, V. V., Daniel, R., Schwarz, W. H., & Liebl, W. (2013). Complete genome sequence of *Clostridium stercorarium* subsp. stercorarium Strain DSM 8532, a thermophilic degrader of plant cell wall fibers. *Genome Announc, 1*(2), e0007313. doi:10.1128/genomeA.00073-13.

Ponder GR, R. G. (1997). Arabinogalactan from western larch, part III: alkaline degradation revisited, with novel conclusions on molecular structure. *Carbohydr Polym, 34*, 261-261. doi:10.1016/S0144-8617(97)00099-4.

Rabausch, U., Juergensen, J., Ilmberger, N., Bohnke, S., Fischer, S., Schubach, B., . . . Streit, W. R. (2013). Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl Environ Microbiol, 79*(15), 4551-4563. doi:10.1128/AEM.01077-13.

Ratnayake, S. j. a. R. (2019). Microbial cellulases: An overview and application. doi: 10.5772/intechopen.84531.

Régis Fauré, C. M. C., Jan A. Delcour,D Claire Dumon,Craig B. Faulds. (2009). A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell. *Research Front, 62*, 533-537. doi: 10.1071/CH08458.

Reitmeier, S., Kiessling, S., Clavel, T., List, M., Almeida, E. L., Ghosh, T. S., . . . Haller, D. (2020a). Arrhythmic gut microbiome signatures predict risk of type 2 diabetes. *Cell Host Microbe, 28*(2), 258-272.e256. doi:10.1016/j.chom.2020.06.004.

Reitmeier, S., Kiessling, S., Neuhaus, K., & Haller, D. (2020b). Comparing circadian rhythmicity in the human gut microbiome. *STAR Protoc, 1*(3), 100148. doi:10.1016/j.xpro.2020.100148.

Rettenmaier, R., Gerbaulet, M., Liebl, W., & Zverlov, V. V. (2019). *Hungateiclostridium mesophilum* sp. nov., a mesophilic, cellulolytic and spore-forming bacterium isolated from a biogas fermenter

fed with maize silage. *Int J Syst Evol Microbiol, 69*(11), 3567-3573. doi:10.1099/ijsem.0.003663.

Rettenmaier, R., Schneider, M., Munk, B., Lebuhn, M., Junemann, S., Sczyrba, A., . . . Liebl, W. (2020). Importance of *Defluviitalea raffinosedens* for hydrolytic biomass degradation in co-culture with *Hungateiclostridium thermocellum*. *Microorganisms, 8*(6). doi:10.3390/microorganisms8060915.

Ridley, B. L., O'Neill, M. A., & Mohnen, D. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry, 57*(6), 929-967. doi:10.1016/s0031-9422(01)00113-3.

Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., . . . Bolam, D. N. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun, 6*, 7481. doi:10.1038/ncomms8481.

Rohman, A., van Oosterwijk, N., Puspaningsih, N. N. T., & Dijkstra, B. W. (2018). Structural basis of product inhibition by arabinose and xylose of the thermostable GH43 beta-1,4-xylosidase from *Geobacillus thermoleovorans* IT-08. *PLoS One, 13*(4), e0196358. doi:10.1371/journal.pone.0196358.

Wirth, R., Kovács, E., Maróti, G., Bagi, Z., Rákhely, G., Kovács, K. L. (2012). Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnology for biofuels, 5*(41). doi: 10.1186/1754-6834-5-41.

Sae-Lee, R., & Boonmee, A. (2014). Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities. *Folia Microbiol (Praha), 59*(5), 409-417. doi:10.1007/s12223-014-0313-7.

Saier, M. H. (1998). Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. In (pp. 81-136). doi: 10.1016/s0065-2911(08)60130-7.

Sakamoto, T., & Ishimaru, M. (2013). Peculiarities and applications of galactanolytic enzymes that act on type I and II arabinogalactans. *Appl Microbiol Biotechnol, 97*(12), 5201-5213. doi:10.1007/s00253-013-4946-2.

Sakka, M., Kunitake, E., Kimura, T., & Sakka, K. (2019). Function of a laminin_G_3 module as a carbohydrate-binding module in an arabinofuranosidase from *Ruminiclostridium josui*. *FEBS Lett, 593*(1), 42-51. doi:10.1002/1873-3468.13283.

Sarai, N. S., Himmel, M. E., Bomble, Y. J., Kahn, A., & Bayer, E. A. (2019). Fundamentals and industrial applicability of multifunctional CAZyme systems. In *Comprehensive Biotechnology* (pp. 14-23). doi: https://doi.org/10.1016/B978-0-12-809633-8.09128-7.

Shallom, D., Belakhov, V., Solomon, D., Gilead-Gropper, S., Baasov, T., Shoham, G., & Shoham, Y. (2002). The identification of the acid-base catalyst of alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *FEBS Lett, 514*(2-3), 163-167. doi:10.1016/s0014-5793(02)02343-8.

Shi, H., Li, X., Gu, H., Zhang, Y., Huang, Y., Wang, L., Wang, F. . (2013). Biochemical properties of a novel thermostable and highly xylose-tolerant beta-xylosidase/alpha-arabinosidase from *Thermotoga thermarum*. *Biotechnology for biofuels, 6*(1), 27. doi: 10.1186/1754-6834-6-27.

Shi, H., Zhang, Y., Xu, B., Tu, M., & Wang, F. (2014). Characterization of a novel GH2 family alpha-L-arabinofuranosidase from hyperthermophilic bacterium T*hermotoga thermarum*. *Biotechnol Lett,*

36(6), 1321-1328. doi:10.1007/s10529-014-1493-6.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Mol Ecol, 21*(8), 1794-1805. doi:10.1111/j.1365-294X.2012.05538.x.

Shrotri, A., Kobayashi, H., & Fukuoka, A. (2017). Catalytic conversion of structural carbohydrates and lignin to chemicals. In C. Song (Ed.), *Advances in Catalysis* (Vol. 60, pp. 59-123): Academic Press. doi:https://doi.org/10.1016/bs.acat.2017.09.002.

Simon, C., & Daniel, R. (2009a). Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol, 85*(2), 265-276. doi:10.1007/s00253-009-2233-z.

Simon, C., & Daniel, R. (2011). Metagenomic analyses: past and future trends. *Appl Environ Microbiol, 77*(4), 1153-1161. doi:10.1128/AEM.02345-10.

Simon, C., Herath, J., Rockstroh, S., & Daniel, R. (2009b). Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol, 75*(9), 2964-2968. doi:10.1128/AEM.02644-08.

Smith, P. J., Wang, H. T., York, W. S., Pena, M. J., & Urbanowicz, B. R. (2017). Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. *Biotechnol Biofuels, 10*, 286. doi:10.1186/s13068-017-0973-z.

Sorensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol, 115*(2), 113-128. doi:10.1016/j.jbiotec.2004.08.004.

Sørensen, H. R., Jørgensen, C. T., Hansen, C. H., Jørgensen, C. I., Pedersen, S., & Meyer, A. S. (2006). A novel GH43 alpha-L-arabinofuranosidase from *Humicola insolens*: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. *Appl Microbiol Biotechnol, 73*(4), 850-861. doi:10.1007/s00253-006-0543-y.

Spagnuolo, M., Crecchio, C., Pizzigallo, M. D. R., & Ruggiero, P. (1997). Synergistic effects of cellulolytic and pectinolytic enzymes in degrading sugar beet pulp. *Bioresource Technology, 60*(3), 215-222. doi:10.1016/S0960-8524(97)00013-8.

Squina, F. M., Santos, C. R., Ribeiro, D. A., Cota, J., de Oliveira, R. R., Ruller, R., . . . Prade, R. A. (2010). Substrate cleavage pattern, biophysical characterization and low-resolution structure of a novel hyperthermostable arabinanase from *Thermotoga petrophila*. *Biochem Biophys Res Commun, 399*(4), 505-511. doi:10.1016/j.bbrc.2010.07.097.

Pao, S. S., Paulsen, I. T., Saier, M. H., Jr. (1998). Major facilitator superfamily. *Microbiology and molecular biology reviews, 62*(1), 1-34. doi: 10.1128/MMBR.62.1.1-34.1998.

Sterner, R., & Hocker, B. (2005). Catalytic versatility, stability, and evolution of the $(beta/alpha)_8$-barrel enzyme fold. *Chem Rev, 105*(11), 4038-4055. doi:10.1021/cr030191z.

Stolze, Y., Zakrzewski, M., Maus, I., Eikmeyer, F., Jaenicke, S., Rottmann, N., . . . Schluter, A. (2015). Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. *Biotechnol Biofuels, 8*, 14. doi:10.1186/s13068-014-0193-8.

Debeche, T., Cummings, N., , Connerton, I., Debeire, P., O'Donohue, M. J. (2000). Genetic and biochemical characterization of a highly thermostable alpha-L-arabinofuranosidase from

Thermobacillus xylanilyticus. *Appl Environ Microbiol, 66*(4), 1734-1736. doi: 10.1128/AEM.66.4.1734-1736.2000.

Kim, T. H., Lee, Y. Y. , Sunwoo, C. , Kim, J. S. (2006). Pretreatment of corn stover by low liquid ammonia recycle percolation process. *Appl Biochem Biotechnol, 133*(1), 41-57. doi: 10.1016/j.biortech.2005.01.015.

Takao, M., Akiyama, K., & Sakai, T. (2002). Purification and characterization of thermostable endo-1,5-alpha-L-arabinase from a strain of *Bacillus thermodenitrificans*. *Appl Environ Microbiol, 68*(4), 1639-1646. doi:10.1128/AEM.68.4.1639-1646.2002.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., . . . Potocki-Veronese, G. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res, 20*(11), 1605-1612. doi:10.1101/gr.108332.110.

Taupp, M., Mewis, K., & Hallam, S. J. (2011). The art and design of functional metagenomic screens. *Curr Opin Biotechnol, 22*(3), 465-472. doi:10.1016/j.copbio.2011.02.010.

Taylor, E. J., Smith, N. L., Turkenburg, J. P., D'Souza, S., Gilbert, H. J., & Davies, G. J. (2006). Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*. *Biochem J, 395*(1), 31-37. doi:10.1042/BJ20051780.

Taylor, P. L., Blakely, K. M., de Leon, G. P., Walker, J. R., McArthur, F., Evdokimova, E., . . . Junop, M. S. (2008). Structure and function of sedoheptulose-7-phosphate isomerase, a critical enzyme for lipopolysaccharide biosynthesis and a target for antibiotic adjuvants. *J Biol Chem, 283*(5), 2835-2845. doi:10.1074/jbc.M706163200.

Teeri, T. T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology, 15*(5), 160-167. doi:10.1016/S0167-7799(97)01032-9

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp, 2*(1), 3. doi:10.1186/2042-5783-2-3.

Torres-Cortes, G., Millan, V., Ramirez-Saad, H. C., Nisa-Martinez, R., Toro, N., & Martinez-Abarca, F. (2011). Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environ Microbiol, 13*(4), 1101-1114. doi:10.1111/j.1462-2920.2010.02422.x.

Van den Broek, L. A. M, Lloyd, R. M., Beldman, G., Verdoes, J. C., McCleary, B. V., & Voragen, A. G. (2005). Cloning and characterization of arabinoxylan arabinofuranohydrolase-D3 (AXHd3) from *Bifidobacterium adolescentis* DSM20083. *Appl Microbiol Biotechnol, 67*(5), 641-647. doi:10.1007/s00253-004-1850-9.

Vandermarliere, E., Bourgois, T. M., Winn, M. D., van Campenhout, S., Volckaert, G., Delcour, J. A., . . . Courtin, C. M. (2009). Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem J, 418*(1), 39-47. doi:10.1042/BJ20081256.

Velasco, J., Oliva, B., Goncalves, A. L., Lima, A. S., Ferreira, G., Franca, B. A., . . . Segato, F. (2020). Functional characterization of a novel thermophilic exo-arabinanase from *Thermothielavioides terrestris*. *Appl Microbiol Biotechnol, 104*(19), 8309-8326. doi:10.1007/s00253-020-10806-6.

Verma, D., Anand, A., & Satyanarayana, T. (2013). Thermostable and alkalistable endoxylanase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1: cloning, expression, characteristics and its applicability in generating xylooligosaccharides and fermentable sugars. *Appl Biochem Biotechnol, 170*(1), 119-130. doi:10.1007/s12010-013-0174-6.

Villegas, A., & Kropinski, A. M. (2008). An analysis of initiation codon utilization in the domain bacteria - concerns about the quality of bacterial genome annotation. *Microbiology, 154*(Pt 9), 2559-2661. doi:10.1099/mic.0.2008/021360-0.

Wagschal, K., Heng, C., Lee, C. C., & Wong, D. W. (2009). Biochemical characterization of a novel dual-function arabinofuranosidase/xylosidase isolated from a compost starter mixture. *Appl Microbiol Biotechnol, 81*(5), 855-863. doi:10.1007/s00253-008-1662-4.

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

Wang, M., Lai, G. L., Nie, Y., Geng, S., Liu, L., Zhu, B., . . . Wu, X. L. (2015). Synergistic function of four novel thermostable glycoside hydrolases from a long-term enriched thermophilic methanogenic digester. *Front Microbiol, 6*, 509. doi:10.3389/fmicb.2015.00509.

Wang, W., Mai-Gisondi, G., Stogios, P. J., Kaur, A., Xu, X., Cui, H., . . . Master, E. R. (2014). Elucidation of the molecular basis for arabinoxylan-debranching activity of a thermostable family GH62 α-L-arabinofuranosidase from *Streptomyces thermoviolaceus*. *Appl Environ Microbiol, 80*(17), 5317-5329. doi:10.1128/aem.00685-14.

Wang, X., Huang, H., Xie, X., Ma, R., Bai, Y., Zheng, F., . . . Luo, H. (2016b). Improvement of the catalytic performance of a hyperthermostable GH10 xylanase from *Talaromyces leycettanus* JCM12802. *Bioresour Technol, 222*, 277-284. doi:10.1016/j.biortech.2016.10.003.

Wang, Y., Hayatsu, M., & Fujii, T. (2012). Extraction of bacterial RNA from soil: challenges and solutions. *Microbes Environ, 27*(2), 111-121. doi:10.1264/jsme2.me11304.

Ward, O. P., & Moo-Young, M. (1988). Thermostable enzymes. *Biotechnol Adv, 6*(1), 39-69. doi:10.1016/0734-9750(88)90573-3.

Warren, R. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S., & Holt, R. A. (2008). Transcription of foreign DNA in *Escherichia coli*. *Genome Res, 18*(11), 1798-1805. doi:10.1101/gr.080358.108.

Wefers, D., & Bunzel, M. (2016). Arabinan and galactan oligosaccharide profiling by High-Performance Anion-Exchange chromatography with pulsed amperometric detection (HPAEC-PAD). *J Agric Food Chem, 64*(22), 4656-4664. doi:10.1021/acs.jafc.6b01121.

Wefers, D., Dong, J., Abdel-Hamid, A. M., Paul, H. M., Pereira, G. V., Han, Y., . . . Cann, I. (2017). Enzymatic mechanism for arabinan degradation and transport in the Thermophilic Bacterium *Caldanaerobius polysaccharolyticus*. *Appl Environ Microbiol, 83*(18). doi:10.1128/AEM.00794-17.

Wefers, D., Tyl, C. E., & Bunzel, M. (2014). Novel arabinan and galactan oligosaccharides from dicotyledonous plants. *Front Chem, 2*, 100. doi:10.3389/fchem.2014.00100.

Willats, W. G. T., McCartney, L., Mackie, W., Knox, J. P.(2001). Pectin: cell biology and prospects for

functional analysis. *Plant Molecular Biology, 47*, 9-27. doi:10.1023/A:1010662911148.

Withers, S. G. (2001). Mechanisms of glycosyl transferases and hydrolases. *Carbohydr Polym, 44*, 325-337. doi:10.1016/S0144-8617(00)00249-6.

Wong, D. W., Chan, V. J., & Batt, S. B. (2008). Cloning and characterization of a novel exo-alpha-1,5-L-arabinanase gene and the enzyme. *Appl Microbiol Biotechnol, 79*(6), 941-949. doi:10.1007/s00253-008-1504-4.

Gong, X., Gruninger, R. J.,, Qi, M., Paterson, L., Forster, R. J., Teather, R. M., McAllister, T. A. (2012). Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. *BMC Res Notes*, 5:566. doi:10.1186/1756-0500-5-566.

Xu, Q., Luo, Y., Ding, S. Y., Himmel, M. E., Bu, L., Lamed, R., & Bayer, E. A. (2011). Multifunctional enzyme systems for plant cell wall degradation. In *Comprehensive Biotechnology* (pp. 15-25). doi:10.1016/B978-0-08-088504-9.00167-7

Yang, W., Bai, Y., Yang, P., Luo, H., Huang, H., Meng, K., . . . Yao, B. (2015). A novel bifunctional GH51 exo-alpha-L-arabinofuranosidase/endo-xylanase from *Alicyclobacillus* sp. A4 with significant biomass-degrading capacity. *Biotechnol Biofuels, 8*, 197. doi:10.1186/s13068-015-0366-0.

Yeoman, C. J., Han, Y., Dodd, D., Schroeder, C. M., Mackie, R. I., & Cann, I. K. O. (2010). Thermostable enzymes as biocatalysts in the biofuel industry. In (pp. 1-55). doi:10.1016/S0065-2164(10)70001-0.

Yourno, J., Kohno, T., & Roth, J. R. (1970). Enzyme evolution: generation of a bifunctional enzyme by fusion of adjacent genes. *Nature, 228*(5274), 820-824. doi:10.1038/228820a0.

Zakrzewski, M., Goesmann, A., Jaenicke, S., Junemann, S., Eikmeyer, F., Szczepanowski, R., . . . Schluter, A. (2012). Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscript ome sequencing. *J Biotechnol, 158*(4), 248-258. doi:10.1016/j.jbiotec.2012.01.020.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., . . . Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res, 46*(W1), W95-W101. doi:10.1093/nar/gky418. doi: 10.1093/nar/gky418.

Zykwinska, A., Thibault, J. F., & Ralet, M. C. (2007). Organization of pectic arabinan and galactan side chains in association with cellulose microfibrils in primary cell walls and related models envisaged. *J Exp Bot, 58*(7), 1795-1802. doi:10.1093/jxb/erm037.

6 Appendix

6.1 Cloning, production and purification of recombinant enzymes.

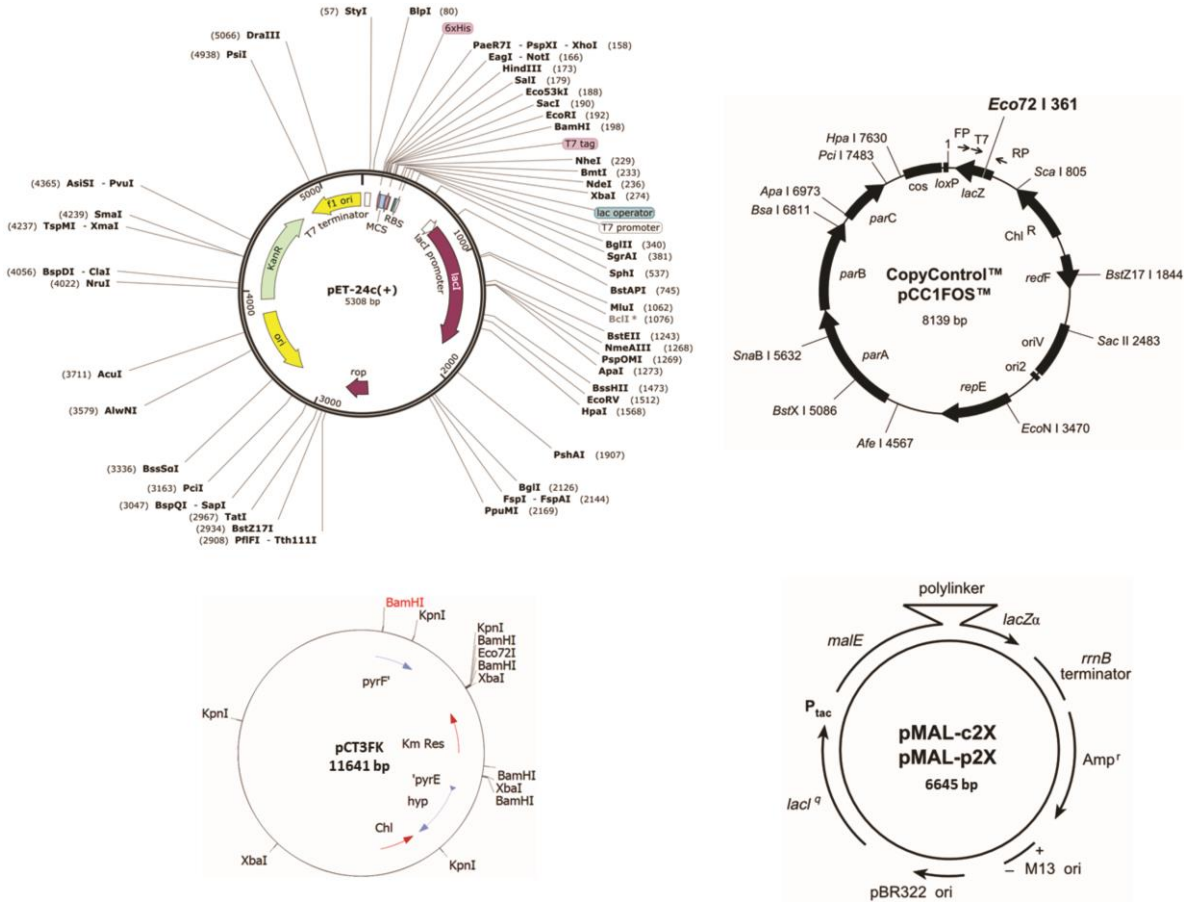
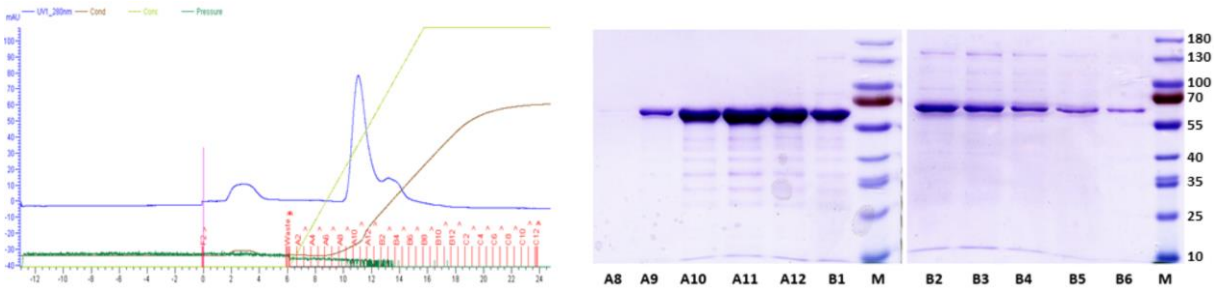
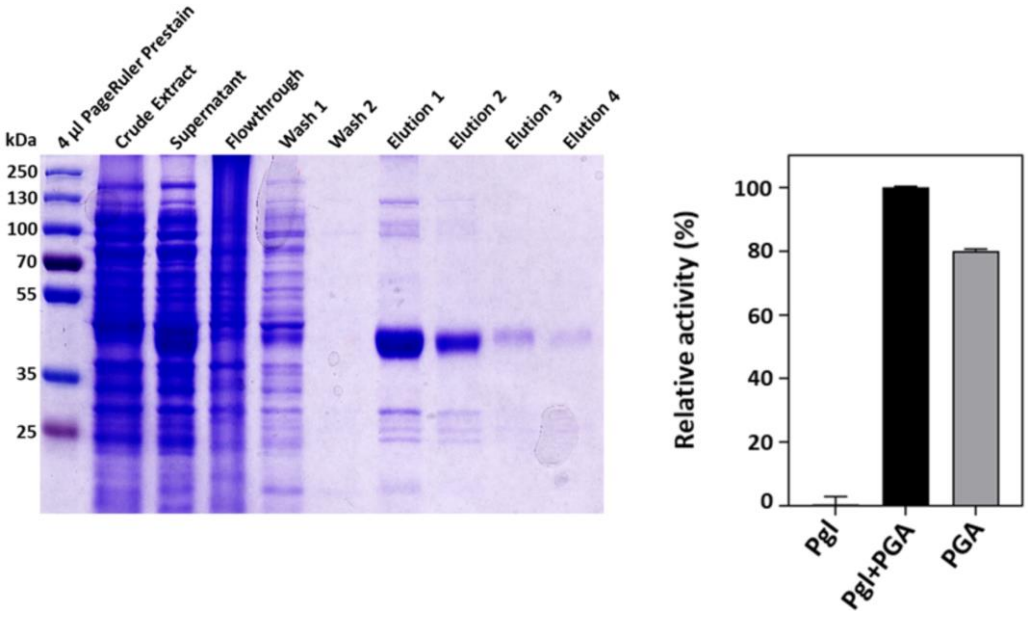


Figure 6.1 Maps of empty vectors/fosmids used in the study. Including the sites of the restriction enzymes. antibiotic resistance gene (*Kan*R in pET-24c, *Chl* in pCC1FOS and pCT3FK, *Amp*^{r} in pMAL-c2X), multipe cloning site (MCS).



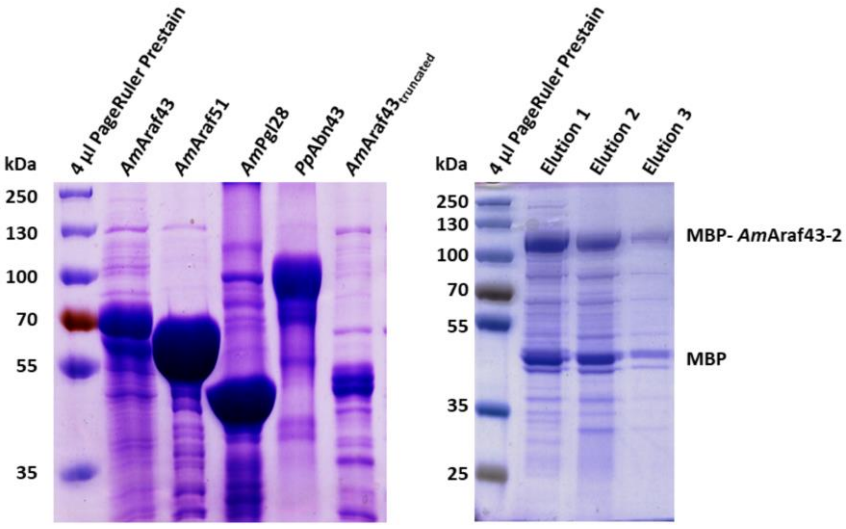
(a) (b)

Figure 6.2 Examples of chromatogram and SDS-PAGE analysis from protein purification. (a) Chromatogram of the fast protein liquid chromatography with anion exchange chromatography (FPLC-AEC) with a 1 mL QFF column. (b) SDS-PAGE of the protein fractions from the FPLC-AEC purification of *Am*Araf51. M: PageRuler^{TM} Prestained Protein Ladder.



(a) (b)

Figure 6.3 SDS-PAGE analysis the purification of *Am*Pgl28 by a Nickel column and DNS assay measure the reducing sugar released from polygalacturonic acid (PGA) by *Am*Pgl28 (from *Acetivibrio mesophilus*). The enzyme assay was performed by incubating 96 µg *Am*Pgl28 with 0.5% of PGA in 25 mM phosphate buffer (pH=7.0) at 37°C overnight.



(a) (b)

Figure 6.4 SDS-PAGE analysis the purification of protein *Am*Araf43, *Am*Araf51, *Am*Pgl28, *Pp*Abn43, *Am*Araf43$_{truncated}$ by immobilized affinity chromatography with Nickel column and purification of MBP-*Am*Araf43-2 fusion protein by amylose resin (Gene bank RXE58509.1; from *Acetivibrio mesophilus*).

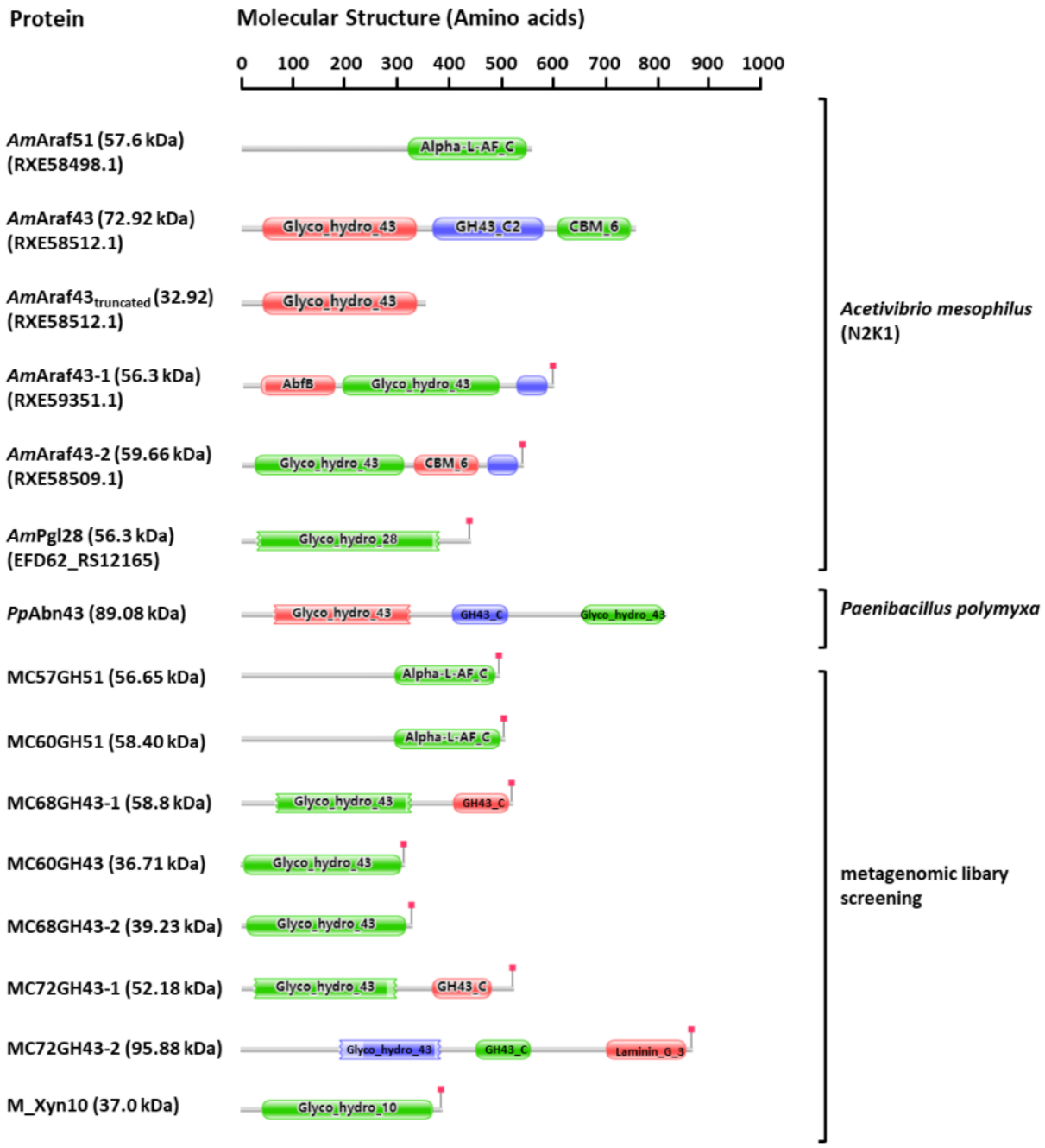


Figure 6.5 Schematic structure of glycoside hydrolases in this study. The molecular mass of the native proteins is indicated in parentheses. The size of enzymes and modules is scaled according to the number of amino acids. Classification of GH and CBM modules was determined using the CAZy database. The module structures were determined using the Pfam database.

Table 6.1 Basic parameters of characterized proteins.

Proteins	Origin	Gene size (bps)	Mass (kDa)	pI	Activity	SP (Yes/No)	Concentration (mg mL^{-1})
*Am*Araf51 (RXE58498.1)	*Acetivibrio mesophilus*	1,530	57.6	6.34	Exo-AFase Exo-xylosidase	No	0.269 (AEX)
*Am*Araf43 (RXE58512.1)	*Acetivibrio mesophilus*	1,951	72.92	5.31	Exo-AFase	Yes	0.274 (AEX)
*Am*Araf43$_{truncated}$ (RXE58512.1)	*Acetivibrio mesophilus*	876	32.92	6.03	No activity	Yes	0.357 (IMAC)
*Am*Araf43-1 (RXE59351.1)	*Acetivibrio mesophilus*	1,779	66.6	6.06	Exo-AFase (Test CE)	Yes	0.545 (IMAC)
*Am*Araf43-2 (RXE58509.1)	*Acetivibrio mesophilus*	1,611	101.224 (MBP-fusion)	4.96	No activity	Yes	0.15 Amylose resin purification
*Am*Pgl28 (EFD62_RS12165)	*Acetivibrio mesophilus*	1,299	46.95	4.91	Polygalacturonase	No	0.645 (IMAC)
*Pp*Abn43	*Paenibacillus polymyxa*		89.08	6.00	Endo-arabinanase	No	0.824 (IMAC)
MC57GH51	Metagenomic	1,500	56.65	5.56	Exo-AFase	No	0.055 (AEX)
MC60GH51	Metagenomic	1,533	58.40	5.10	Exo-AFase	No	0.190 (AEX)
MC68GH43-1	Metagenomic	1,575	58.8	5.06	Endo-arabinanase	No	0.35 (IMAC)
MC60GH43	Metagenomic	960	36.71	5.21	Exo-AFase	No	0.083 (AEX)
MC68GH43-2	Metagenomic	1,002	39.23	6.29	Exo-AFase	No	1.063 (AEX)
MC72GH43-1	Metagenomic	1,407	52.18	4.36	No activity	No	0.246 (AEX)
MC72GH43-2	Metagenomic	2,580	95.88	4.51	Endo-arabinanase	No	0.251 (AEX)
M_Xyn10	Metagenomic	1,078	36.9	5.52	Endo-xylanase Exo-xylosidase Transglycosidase	No	0.692 (IMAC)

AFase: arabinofuranosidase, MBP: maltose-binding protein, AEC: anion-exchange choromatography, IMAC: immobilized metal affinity chromatography, CE: crude extract.

6.2 Reference line

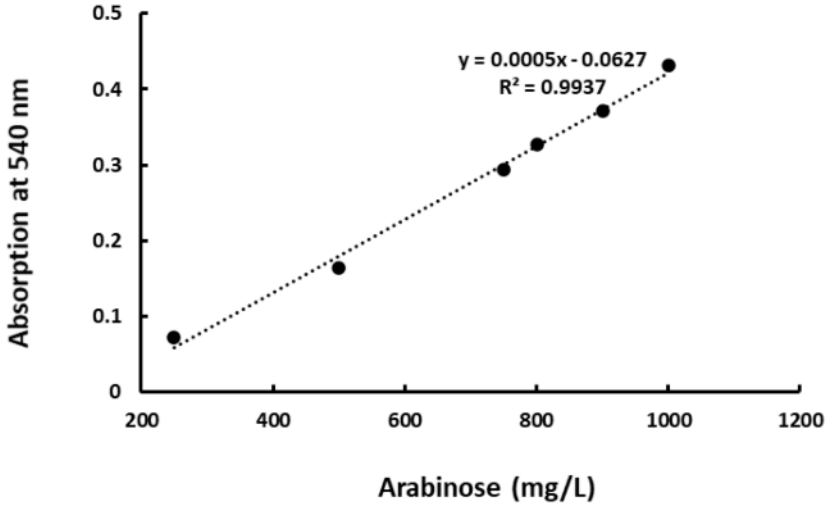


Figure 6.6 Reference line for photometric quantification of arabinose concentration using DNS assay at a wavelength of 540. Six concentrations were measured for the reference line: Arabinose: 1000, 900, 800, 750, 500, 250 mg L^{-1}.

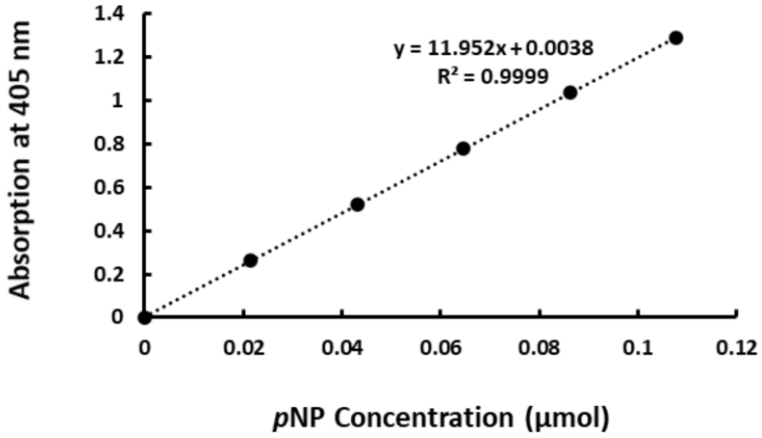


Figure 6.7 Reference line for photometric quantification of *p*NP concentration using *p*NP assay at a wavelength of 405. Six concentrations were measured for the reference line: 4-Nitrophenol: 0, 0.022, 0.043, 0.065, 0.086, 0.108 µmol.

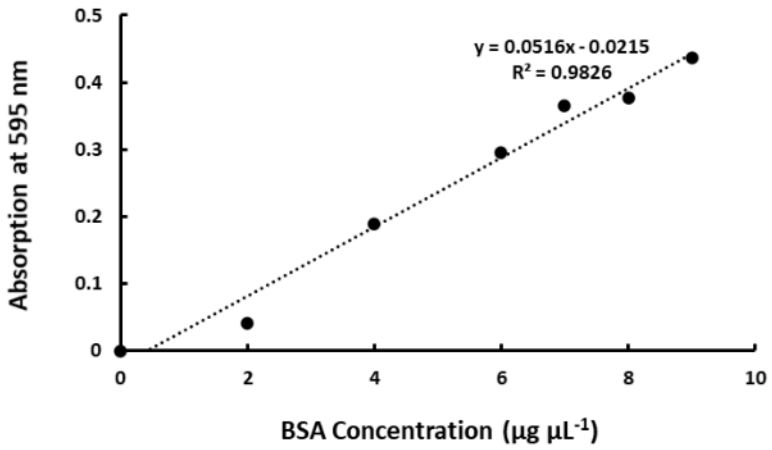


Figure 6.8 Reference line for photometric quantification of protein concentration (BSA) using Bradford assay at a wavelength of 540. Seven concentrations were measured for the reference line: BSA: 0, 2, 4, 6, 7, 8, 9 µg.

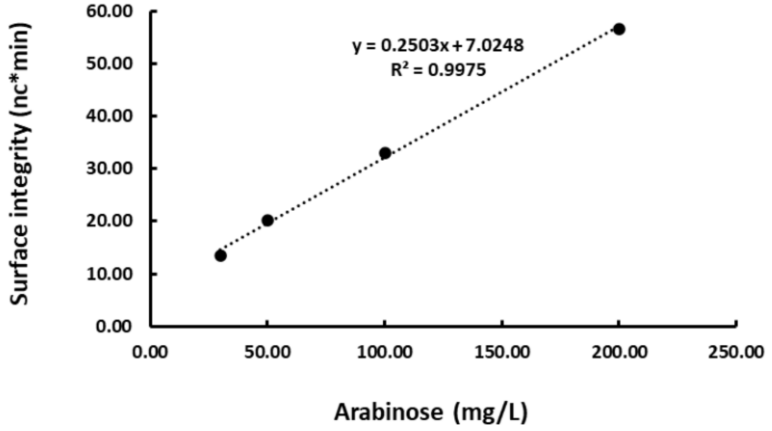


Figure 6.9 Reference line for quantifying the amount of arabinose with HPAEC-PAD. For the reference, four concentrations of arabinose were used for measurement: 25, 50, 100, 200 mg L^{-1}.

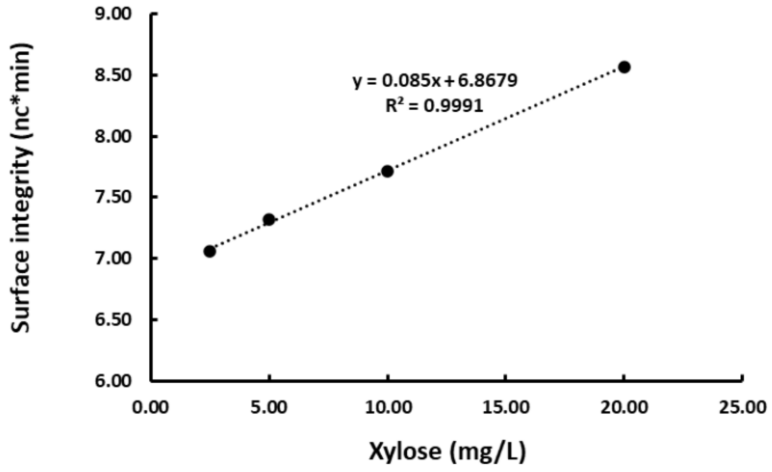


Figure 6.10 Reference line for quantifying the amount of xylose with HPAEC-PAD. For the reference, four concentrations of xylose were used for measurement: 2.5, 5, 10, 20 mg L^{-1}.

6.3 Function based screenings

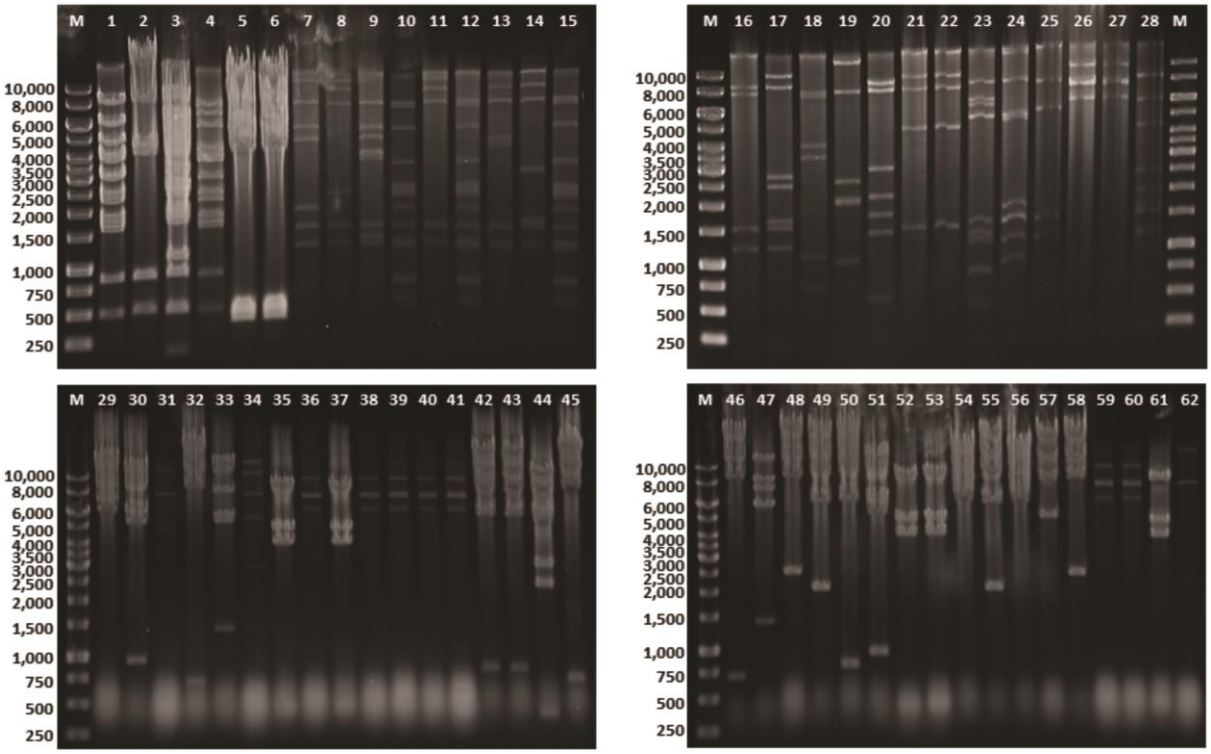


Figure 6.11 BamH I digests of selected positive fosmid clones. M. GeneRuler 1 kB DNA Ladder, 1. MOD18 Ara1, 2. MOD18 Ara2, 3. MOD18 Ara3, 4. MOD18 Ara4, 5. T1T2-33 Ara, 6. T1T2-24 Ara, 7. BM1T-3 D12, 8. BM1T-4 G4, 9. BM1T-9 D12, 10. BM1T-2 G5, 11. BM1T-3 A7, 12. BM1T-2 B12, 13. BM1T-8 A8, 14. BM1T-2 B9, 15. BM1T-2 H7, 16. BM1T-4 B7, 17. BM1T-3 A11, 18. ED-21S B6, 19. ED-27m B11-1, 20. ED-27m B11-2, 21. ED-26m C9-1, 22. ED-26m C9-2, 23. BM1T-10 B11, 24. BM1T-6 D5, 25. BM1T-7-1 A11, 26. T1T2-8, 27. T1T2-31, 28. BM1T-5 H4, 29.T1T2-1, 30. T1T2-2, 31.T1T2-3, 32.T1T2-4, 33.T1T2-5, 34.T1T2-6, 35.T1T2-7, 36.T1T2-8, 37.T1T2-9, 38.T1T2-10, 39.T1T2-11, 40.T1T2-12, 41.T1T2-13, 42.T1T2-14, 43.T1T2-15, 44.T1T2-16, 45.T1T2-17, 46.T1T2-18, 47.T1T2-19, 48.T1T2-20, 49.T1T2-21, 50.T1T2-22, 51.T1T2-23, 52.T1T2-24, 53.T1T2-25, 54.T1T2-26, 55.T1T2-27, 56.T1T2-28, 57.T1T2-29, 58.T1T2-30, 59.T1T2-31, 60.T1T2-32, 61.T1T2-33, 62.T1T2-34.

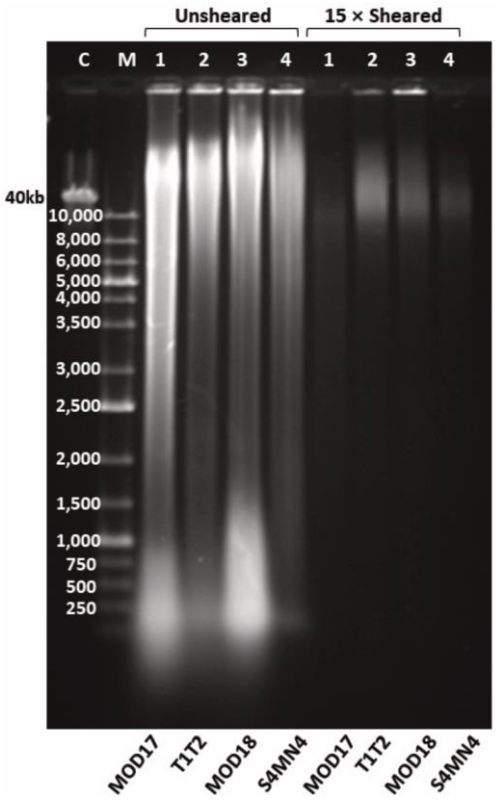
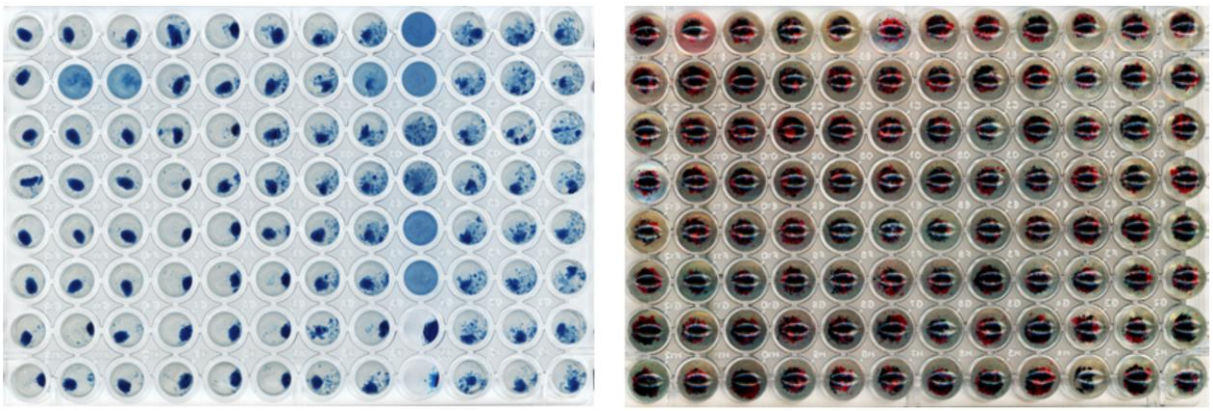


Figure 6.12 Pulsed-field gel electrophoresis (PFGE) of unsheared metagenomic DNA and 15-times sheared metagenomic DNA by using a Hamilton syringe treatment. C: Fosmid control (40 kB); M: Gene Ruler 1 kB DNA Ladder; 1: MOD17, 2: T1T2, 3: MOD18, 4: S4MN4.



(a) (b)

Figure 6.13 Example of metagenomic library screening in 96- well plates. (a) Library ED-27m was screened with CPH-Xylan, (b) Library BM1T-3 was screened with CPH-CMC and CPH-Arabinan. The overnight culture was pretreated by three freeze/thaw cycles to disrupt the cells. The reactions were composed of 150 µL pretreated cell culture, 25 mM citric phosphate buffer pH 7.0, around 30 µL substrate suspension, resulting in a final volume of 200 µL. The plate was sealed using an adhesive PCR membrane and incubated overnight at 37℃ without shaking. Positive clones were identified by the change of color in the supernatant.

Table 6.2 Information about existing metagenomic libraries.

Name of libraries	Origin of Samples	Host strain	Fosmid	Researchers
TN-1 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-2 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-3 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-4 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-5 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-6 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-7 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-8 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-10 m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
TN-9m	Teredo navalis (TN) from W. Streit (Universität Hamburg)	EPI300-T1M	pCC1FOS	Benedikt Leis
M12FOS1	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS2	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS3	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS4	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS5	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS6	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS7	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS8	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS9	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS10	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS16	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS17	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS18	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS19	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOS20	"Kompostmiete" in Göttingen	EPI300-T1M	pCC1FOS	Markus Mientus
M12FOSpCT3FK1	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOSpCT3FK2	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOSpCT3FK4	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOSpCT3FK5	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOSpCT3FK8	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus

Name of libraries	Origin of Samples	Host strain	Fosmid	Researchers
M12FOST	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOST13	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
M12FOST16	"Kompostmiete" in Göttingen	EPI300-T1M	pCT3FK	Markus Mientus
E_rf22(96×V14) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf23(96×V15) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf24(96×V17) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf25(96×V18) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf18(96×V9) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf19(96×V11) KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf13 17.9.12 besteht aus_V5 (52 Klone)	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf13 17.9.12 besteht aus_V3 (44 Klone)	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf15 17.9.12TP besteht aus: V6 (96 klone)	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
E_rf16(96×V7) 24.9.12 KL+TP	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	EP1300-T1M	pCC1FOS	Trang Pham and Katharina Lachmann
T_rf 07	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 08	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 09	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 10	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 11	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 20	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann

Name of libraries	Origin of Samples	Host strain	Fosmid	Researchers
T_rf 21	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
T_rf 22	Mexico (Sedimente aus einer schwefelhaltigen, heißen Quelle in Chignahuapan, Puebla in Mexiko	Thermus	pCT3FK	Trang Pham and Katharina Lachmann
BM1T-2 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-3 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-4 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-6 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-8 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-9 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-5 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-7-1 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-7-2 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
BM1T-10 2015	-	EP1300-T1M	pCT3FK	Dr. Alexander Pechtl
ED-21m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-22m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-23m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-24m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-25m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-26m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-27m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-28m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-29m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ED-30m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ES-1m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ES-6m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit
ES-7m	Elephant feces	EPI300-T1M	pCC1FOS	Prof. Streit

Table 6.3 55 positive fosmid clones for both forward and reverse Sanger end sequencing.

Number	Clone Name	Seq.ID	Functional Screen_Activity	Library Name (Internal)	Vector Backbone	Length （bps）
1	T1T2_2 For	88CC87	CMC	T1T2	pCC1FOS	838
2	T1T2_2 Rev	88CC86	CMC	T1T2	pCC1FOS	762
3	T1T2_1 For	88CC83	CMC	T1T2	pCC1FOS	767
4	T1T2_1 Rev	88CC88	CMC	T1T2	pCC1FOS	885
5	T1T2_6 For	88CC93	CMC	T1T2	pCC1FOS	681
6	T1T2_6 Rev	88CC92	CMC	T1T2	pCC1FOS	574
7	T1T2_4 For	88CC91	CMC	T1T2	pCC1FOS	726
8	T1T2_4 Rev	88CC90	CMC	T1T2	pCC1FOS	963
9	MOD18_17_For	94EE66	CMC	MOD18	pCC1FOS	1,120
10	MOD18_17_Rev	94EE65	CMC	MOD18	pCC1FOS	1,090
11	MOD18_28 For	94EE80	CMC	MOD18	pCC1FOS	718
12	MOD18_28 Rev	94EE85	CMC	MOD18	pCC1FOS	557
13	MOD18_14_For	94EE62	CMC	MOD18	pCC1FOS	762
14	MOD18_14_Rev	94EE67	CMC	MOD18	pCC1FOS	557
15	MOD18_13_For	94EE72	CMC	MOD18	pCC1FOS	718
16	MOD18_13_Rev	94EE71	CMC	MOD18	pCC1FOS	839
17	T1T2_14 For	88CC97	CMC	T1T2	pCC1FOS	808
18	T1T2_14 Rev	88CC96	CMC	T1T2	pCC1FOS	846
19	T1T2_26 For	88CD15	CMC	T1T2	pCC1FOS	804
20	T1T2_26 Rev	88CD14	CMC	T1T2	pCC1FOS	1,171
21	T1T2_23 For	88CD11	CMC	T1T2	pCC1FOS	840
22	T1T2_23 Rev	88CD10	CMC	T1T2	pCC1FOS	891
23	MOD18_26 For	94EE86	CMC	MOD18	pCC1FOS	632
24	MOD18_26 Rev	94EE91	CMC	MOD18	pCC1FOS	853
25	MOD18_18_For	94EE64	CMC	MOD18	pCC1FOS	1,111
26	MOD18_18_Rev	94EE63	CMC	MOD18	pCC1FOS	842
27	MOD18_7-2 For	94EE68	CMC	MOD18	pCC1FOS	1038
28	MOD18_7-2 Rev	94EE73	CMC	MOD18	pCC1FOS	845
29	MOD18_8 For	88CD39	CMC	MOD18	pCC1FOS	679
30	MOD18_8 Rev	88CD38	CMC	MOD18	pCC1FOS	965
31	T1T2_21 For	88CD09	CMC	T1T2	pCC1FOS	879
32	T1T2_21 Rev	88CD08	CMC	T1T2	pCC1FOS	1,205
33	T1T2_20 For	88CD05	CMC	T1T2	pCC1FOS	477
34	T1T2_20 Rev	88CD04	CMC	T1T2	pCC1FOS	1,126
35	T1T2_29 For	88CD13	CMC	T1T2	pCC1FOS	1,168
36	T1T2_29 Rev	88CD18	CMC	T1T2	pCC1FOS	1,165
37	MOD18_6_For	94EE70	CMC	MOD18	pCC1FOS	274
38	MOD18_6_Rev	94EE69	CMC	MOD18	pCC1FOS	434
39	T1T2_30 For	88CD17	CMC	T1T2	pCC1FOS	603
40	T1T2_30 Rev	88CD16	CMC	T1T2	pCC1FOS	1,193
41	T1T2_17 For	88CC99	CMC	T1T2	pCC1FOS	847
42	T1T2_17 Rev	88CC98	CMC	T1T2	pCC1FOS	838
43	MOD18_24 For	88CD25	CMC	MOD18	pCC1FOS	886
44	MOD18_24 Rev	88CD30	CMC	MOD18	pCC1FOS	1,129
45	T1T2_22 For	88CD07	CMC	T1T2	pCC1FOS	854
46	T1T2_22 Rev	88CD12	CMC	T1T2	pCC1FOS	841
47	T1T2_18 For	88CD03	CMC	T1T2	pCC1FOS	888
48	T1T2_18 Rev	88CD02	CMC	T1T2	pCC1FOS	803
49	T1T2_16 For	88CC95	CMC	T1T2	pCC1FOS	598
50	T1T2_16 Rev	88CD00	CMC	T1T2	pCC1FOS	598

Number	Clone Name	Seq.ID	Functional Screen_Activity	Library Name (Internal)	Vector Backbone	Length （bps）
51	MOD18_29_For	94EE60	CMC	MOD18	pCC1FOS	759
52	MOD18_29_Rev	94EE59	CMC	MOD18	pCC1FOS	883
53	MOD18_12 For	CHG836	CMC	MOD18	pCC1FOS	1,317
54	MOD18_12 Rev	CHG835	CMC	MOD18	pCC1FOS	1,625
55	MOD18_7-1 For	94EE78	CMC	MOD18	pCC1FOS	561
56	MOD18_7-1 Rev	94EE77	CMC	MOD18	pCC1FOS	683
57	MOD18_11 For	94EE84	CMC	MOD18	pCC1FOS	633
58	MOD18_11 Rev	94EE83	CMC	MOD18	pCC1FOS	964
59	ED-26m C9-1 For	CHG851	CMC	ED-library	pCC1FOS	1,319
60	ED-26m C9-1 Rev	CHG852	CMC	ED-library	pCC1FOS	900
61	ED-26m C9-2 For	CHG869	CMC	ED-library	pCC1FOS	
62	ED-26m C9-2 Rev		CMC	ED-library	pCC1FOS	
63	BM1T10-B11 For	CRQ151	AX	BM1T	pCT3FK	1,306
64	BM1T10-B11 Rev	CRQ150	AX	BM1T	pCT3FK	1,197
65	BM1T6-D5 For	CRQ161	AX	BM1T	pCT3FK	1,079
66	BM1T6-D5 Rev	CRQ160	AX	BM1T	pCT3FK	1,336
67	BM1T-7-1 A11 For	CRQ159	AX	BM1T	pCT3FK	1,333
68	BM1T-7-1 A11 Rev	CRQ164	AX	BM1T	pCT3FK	1,493
69	MOD18 Ara1_For	94EE52	Arabinan	MOD18	pCC1FOS	975
70	MOD18 Ara1_Rev	94EE51	Arabinan	MOD18	pCC1FOS	961
71	MOD18 Ara2_For	94EE50	Arabinan	MOD18	pCC1FOS	1,037
72	MOD18 Ara2_Rev	94EE55	Arabinan	MOD18	pCC1FOS	758
73	MOD18 Ara3_For	94EE54	Arabinan	MOD18	pCC1FOS	1,176
74	MOD18 Ara3_Rev	94EE53	Arabinan	MOD18	pCC1FOS	679
75	MOD18 Ara4_For	94EE58	Arabinan	MOD18	pCC1FOS	803
76	MOD18 Ara4_Rev	94EE57	Arabinan	MOD18	pCC1FOS	398
77	T1T2_33 For	88CD19	Arabinan	T1T2	pCC1FOS	836
78	T1T2_33 Rev	88CD24	Arabinan	T1T2	pCC1FOS	1,136
79	T1T2_24 For	88CD21	Arabinan	T1T2	pCC1FOS	847
80	T1T2_24 Rev	88CD20	Arabinan	T1T2	pCC1FOS	401
81	BM1T-3 D12 For	CRQ156	Xylan+AX	BM1T	pCT3FK	1,392
82	BM1T-3 D12 Rev	CRQ157	Xylan+AX	BM1T	pCT3FK	1,703
83	BM1T4-G4 For	CRQ163	Xylan+AX	BM1T	pCT3FK	1,281
84	BM1T4-G4 Rev	CRQ162	Xylan+AX	BM1T	pCT3FK	1,408
85	BM1T9-D12 For	CRQ155	Xylan+AX	BM1T	pCT3FK	1,289
86	BM1T9-D12 Rev	CRQ154	Xylan+AX	BM1T	pCT3FK	1,687
87	BM1T2-G5 Rev	CRQ153	Xylan+AX	BM1T	pCT3FK	1,682
88	BM1T2-G5 For	CRQ158	Xylan+AX	BM1T	pCT3FK	1,393
89	BM1T3-A7 For	CRQ152	Xylan+AX	BM1T	pCT3FK	1,682
90	BM1T3-A7 Rev	CRQ149	Xylan+AX	BM1T	pCT3FK	950
91	BM1T2-B12 For	CRQ210	Xylan+AX	BM1T	pCT3FK	1,463
92	BM1T2-B12 Rev	CRQ123	Xylan+AX	BM1T	pCT3FK	1,492
93	BM1T8-A8 For	CRQ211	Xylan+AX	BM1T	pCT3FK	1,795
94	BM1T8-A8 Rev	CRQ212	Xylan+AX	BM1T	pCT3FK	1,766
95	BM1T2-B9 For	CRQ214	Xylan+AX	BM1T	pCT3FK	1,583
96	BM1T2-B9 Rev	CRQ213	Xylan+AX	BM1T	pCT3FK	1,585
97	BM1T2-H7 For	CRQ215	Xylan+AX	BM1T	pCT3FK	1,538
98	BM1T2-H7 Rev	CRQ125	Xylan+AX	BM1T	pCT3FK	1,479
99	T1T2_8 For	88CD23	Xylan	T1T2	pCC1FOS	686
100	T1T2_8 Rev	88CD22	Xylan	T1T2	pCC1FOS	886
101	T1T2_31 For	88CD27	Xylan	T1T2	pCC1FOS	603
102	T1T2_31 Rev	88CD26	Xylan	T1T1	pCC1FOS	643

Number	Clone Name	Seq.ID	Functional Screen_Activity	Library Name (Internal)	Vector Backbone	Length （bps）
103	BM1T5-H4 For	CRQ217	Xylan	BM1T	pCT3FK	1,361
104	BM1T5-H4 Rev	CRQ216	Xylan	BM1T	pCT3FK	1,186
105	BM1T4-B7 Rev	CQR219	Xylan	BM1T	pCT3FK	1,390
106	BM1T4-B7 For	CRQ218	Xylan	BM1T	pCT3FK	1,429
107	BM1T-3 A11 For	CRQ121	Xylan	BM1T	pCT3FK	1,501
108	BM1T-3 A11 Rev	CRQ120	Xylan	BM1T	pCT3FK	1,583
109	ED-21S B6/B1 For	GH838	Xylan	ED-library	pCC1FOS	926
110	ED-21S B6/B1 Rev	GH837	Xylan	ED-library	pCC1FOS	1,721

Table 6.4 Raw data summary of Shotgun sequencing and assemblage of fosmid clones. 40 positive fosmid clones derived from pCC1 fosmid libraries and 15 positive fosmid clones from pCT3FK fosmid library (CAZyme is predicted by dbCAN database).

Summary	Contigs pCC1 Fosmid	Contigs with CAZy genes pCC1 Fosmid	Contigs pCTF3K Fosmid	Contigs with CAZy genes pCTF3K Fosmid
No. of contigs (>=1000 bp)	130	54	128	70
Total length (>=1000 bp)	4,750,665	3,968,230	4,805,647	4,164,758
Average length (bps)	36,543	73,485	37,544	59,496
GC content (%)	50.04	-	49.91	-
N50	92,492	-	87,225	-
Longest contig (bps)	327,147	327,147	327,164	327,164

Contigs exceeding the expected insert size (>40 kb) were determined to be *E. coli* genomic DNA contamination, the presence of which did not interfere with clone sequence retrieval as retrieval was done using clone end seqnences, but this contamination may affect “required-depth of coverage calculations”.

Table 6.5 Putative CAZymes on the non-*E. coli* contigs.

dbCAN		KEGG	blastnr indentity
GH43			
551_Contig 41_27753_30071	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	66.84% MULTISPECIES: family 43 glycosylhydrolase [unclassified *Fibrobacter*]
551_Contig 41_30291_32525	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	80.66% MULTISPECIES: carbohydrate-binding protein [unclassified *Fibrobacter*]
551_Contig 60_8873_9811	lacZ	beta-galactosidase [EC:3.2.1.23]	64.69% family 43 glycosylhydrolase [*Gorillibacterium massiliense*]
551_Contig 68_3297_4850	abnA	arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	81.35% glycoside hydrolase family 43 protein [*Thermoclostridium stercorarium*]
551_Contig 68_4869_5849	lacZ	beta-galactosidase [EC:3.2.1.23]	84.31% family 43 glycosylhydrolase [*Thermoclostridium stercorarium*]
551_Contig 72_1_1561	abnA	arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	65.27% hypothetical protein BK120_06600 [*Paenibacillus* sp. FSL A5-0031]
551_Contig 72_2960_5554	abnA	arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	75.21% S-layer domain protein [*Thermoanaerobacter mathranii* subsp. mathranii str. A3]
552_Contig 40_25320_26750	abnA	arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	99.79% TPA: endo-alpha-(1->5)-L-arabinanase [*Firmicutes* bacterium]
552_Contig 40_34791_35702	abnA	arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	80.47% TPA: endo-alpha-(1->5)-L-arabinanase [*Firmicutes* bacterium]
552_Contig 53_19346_20782	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	83.86% family 43 glycosylhydrolase [*Thermoclostridium stercorarium]*
552_Contig 53_20832_22727	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	61.72% Select seq ref|WP_137182487.1| family 43 glycosylhydrolase [*Paenibacillus* sp. SY21-1]
552_Contig 53_22839_24615		xylosidase	64.18% Select seq gb|ABD48561.1| putative exo-xylanase [*Geobacillus thermoleovorans*]
552_Contig 68_1489_3672	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	100% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
552_Contig 68_3698_5116	xynD	arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	99.79% family 43 glycosylhydrolase [*Hungateiclostridium saccincola*]
552_Contig 83_3191_5158		arabinoxylan arabinofuranohydrolase [EC:3.2.1.55]	100% family 43 glycosyl hydrolase [*Hungateiclostridium saccincola*]
GH51			
551_Contig 57_20222_21700	abfA	alpha-L-arabinofuranosidase [EC:3.2.1.55]	99.59% alpha-N-arabinofuranosidase [*Xylanivirga thermophila*]

dbCAN		KEGG	blastnr indentity
551_Contig 60_10084_11595	abfA	alpha-L-arabinofuranosidase [EC:3.2.1.55]	97.42% alpha-N-arabinofuranosidase [*Xylanivirga thermophila*]
GH10			
551_Contig 41_32705_34759		endo-1,4-beta-xylanase [EC:3.2.1.8]	71.2% Select seq ref|WP_088661344.1| carbohydrate-binding protein [*Fibrobacter* sp. UWB2]
551_Contig 41_34983_36488		endo-1,4-beta-xylanase [EC:3.2.1.8]	73.72% carbohydrate-binding protein [*Fibrobacter* sp. UWR3]
551_Contig 58_2578_5235		endo-1,4-beta-xylanase [EC:3.2.1.8]	59.75% TPA: hypothetical protein [*Bacteroidales* bacterium]
552_Contig 40_20987_22144		endo-1,4-beta-xylanase [EC:3.2.1.8]	99.73% endo-1,4-xylanase [bacterium enrichment culture]
552_Contig 53_6674_9940		endo-1,4-beta-xylanase [EC:3.2.1.8]	100% glycoside hydrolase [*Hungateiclostridium saccincola*]
552_Contig 55_4567_6363		endo-1,4-beta-xylanase [EC:3.2.1.8]	99.83% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
552_Contig 57_16610_17836		endo-1,4-beta-xylanase [EC:3.2.1.8]	66.92% 1,4-beta-xylanase [*Halanaerobium saccharolyticum*]
552_Contig 58_1_1195		endo-1,4-beta-xylanase [EC:3.2.1.8]	92.87% glycoside hydrolase family 11 protein [*Hungateiclostridium saccincola*]
552_Contig 61_10525_11592		endo-1,4-beta-xylanase [EC:3.2.1.8]	58.43% endo-1,4-beta-xylanase [*Paenibacillus nanensis*]
552_Contig 61_8034_9023		endo-1,4-beta-xylanase [EC:3.2.1.8]	59.08% endo-1,4-beta-xylanase [*Paenibacillus bovis*]
552_Contig 61_9250_10428		endo-1,4-beta-xylanase [EC:3.2.1.8]	46.15% Select seq ref|WP_088078005.1| endo-1,4-beta-xylanase [*Bacillus alkalitelluris*]
552_Contig 82_1700_4123		endo-1,4-beta-xylanase [EC:3.2.1.8]	100% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
552_Contig 94_1095_1781		endo-1,4-beta-xylanase [EC:3.2.1.8]	70.93% carbohydrate-binding protein [*Gracilibacillus thailandensis*]
GH2			
551_Contig 56_21841_22494		beta-mannosidase [EC:3.2.1.25]	100% Select seq ref|WP_144403490.1| glycoside hydrolase family 2 protein [*Defluviitoga tunisiensis*]
551_Contig 94_1_1849		beta-mannosidase [EC:3.2.1.25]	99.35% glycoside hydrolase family 2 protein [*Defluviitoga tunisiensis*]
552_Contig 33_32489_35470	lacZ	beta-galactosidase [EC:3.2.1.23]	100% TPA: beta-galactosidase [*Shigella* sp.]
552_Contig 102_1_632		beta-mannosidase [EC:3.2.1.25]	99.52% glycoside hydrolase family 2 protein [*Defluviitoga tunisiensis*]
			99.70% glycoside hydrolase family 2 protein [*Defluviitoga tunisiensis*]
552_Contig 128_1_1012		beta-mannosidase [EC:3.2.1.25]	

dbCAN		KEGG	blastnr indentity
GH11			
551_Contig 45_18784_20319		endo-1,4-beta-xylanase [EC:3.2.1.8]	99.22% Select seq sp|P33558.2| RecName: Full=Endo-1,4-beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; Flags: Precursor [*Thermoclostridium stercorarium*]
552_Contig 58_1_1195		endo-1,4-beta-xylanase [EC:3.2.1.8]	92.87% glycoside hydrolase family 11 protein [*Hungateiclostridium saccincola*]
552_Contig 82_1700_4123		endo-1,4-beta-xylanase [EC:3.2.1.8]	100% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
552_Contig 94_1095_1781		endo-1,4-beta-xylanase [EC:3.2.1.8]	70.93% carbohydrate-binding protein [*Gracilibacillus thailandensis*]
GH3			
551_Contig 93_754_2180	bglX	beta-glucosidase [EC:3.2.1.21]	98.53% Select seq ref|WP_144403489.1| glycosyl hydrolase [*Defluviitoga tunisiensis*]
552_Contig 40_13964_16135	bglX	beta-glucosidase [EC:3.2.1.21]	99.86% Select seq gb|AJO67863.1| beta-xylosidase [bacterium enrichment culture]
552_Contig 102_635_2230	bglX	beta-glucosidase [EC:3.2.1.21]	98.87% glycosyl hydrolase [*Defluviitoga tunisiensis*]
GH9			
552_Contig 63_677_2287		endo-glucanase [EC:3.2.1.4]	100% Select seq gb|AXR85444.1| glycoside hydrolase family 9 endoglucanase [uncultured bacterium]
552_Contig 70_1716_3551		endo-glucanase [EC:3.2.1.4]	100% glycoside hydrolase [*Hungateiclostridium saccincola*]
552_Contig 75_820_4731		endo-glucanase [EC:3.2.1.4]	99.77% hypothetical protein [*Hungateiclostridium saccincola*]
GH13			
seq83_831_2270		starch synthase (maltosyl-transferring) [EC:2.4.99.16] (Second best)	100% maltodextrin glycosyltransferase [*Defluviitoga tunisiensis*]
seq111_1_570		starch synthase (maltosyl-transferring) [EC:2.4.99.16] (Second best)	99.47% maltodextrin glycosyltransferase [*Defluviitoga tunisiensis*]
GH4			
551_Contig 56_1592_3049	melA	alpha-galactosidase [EC:3.2.1.22]	100% Select seq ref|WP_045087944.1| alpha-glucosidase/alpha-galactosidase [*Defluviitoga tunisiensis*]
GH106			

dbCAN		KEGG	blastnr indentity
551_Contig 56_4383_7496			99.23% hypothetical protein [*Defluviitoga tunisiensis*]
GH73			
551_Contig 57_518_1327	amiABC	N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28] (second best)	100% glucosaminidase domain-containing protein [*Xylanivirga thermophila*]
GH6			
551_Contig 66_2282_4735	clpC	ATP-dependent Clp protease ATP-binding subunit ClpC	100% ATP-dependent Clp protease ATP-binding subunit [*Xylanivirga thermophila*]
GH127			
551_Contig 72_5640_7598		uncharacterized protein	64.89% glycoside hydrolase family 127 protein [*Petroclostridium xylanilyticum*]
551_Contig 72_7624_7782			78% Select seq gb|PYE64535.1| hypothetical protein C7489_12345 [*Paenibacillus* sp. OV191]
551_Contig 72_7745_8221		alpha-galactosidase [EC:3.2.1.22]	62.33% glycoside hydrolase family 27 protein [*Paenibacillus prosopidis*]
GH94			
551_Contig 83_2629_3328		cellobiose phosphorylase [EC:2.4.1.20]	99.57% glycosyl transferase [*Defluviitoga tunisiensis*]
GH1			
551_Contig 93_36_536	bglB	beta-glucosidase [EC:3.2.1.21]	99.4% beta-glucosidase [*Defluviitoga tunisiensis*]
GH24			
552_Contig 33_18118_18612		lysozyme [EC:3.2.1.17]	100% Select seq ref|WP_001070143.1| MULTISPECIES: lysozyme [Bacteria]
GH36			
552_Contig 33_31184_32437	lacY	MFS transporter, OHS family, lactose permease	100% Select seq ref|WP_000291549.1| MULTISPECIES: lactose permease [*Enterobacteriaceae*]
GH67			
552_Contig 40_11877_13895	aguA	alpha-glucuronidase [EC:3.2.1.139]	99.55% TPA: alpha-glucuronidase [*Firmicutes bacterium*]
CE1			
552_Contig 53_6674_9940		endo-1,4-beta-xylanase [EC:3.2.1.8]	100% glycoside hydrolase [*Hungateiclostridium saccincola*]
552_Contig 68_646_1449			100% esterase family protein [*Hungateiclostridium saccincola*]
552_Contig 82_1_1429	xynA	endo-1,4-beta-xylanase [EC:3.2.1.8] (second best)	100% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
CE4			

dbCAN		KEGG	blastnr indentity
552_Contig 58_1_1195		endo-1,4-beta-xylanase [EC:3.2.1.8]	92.87% glycoside hydrolase family 11 protein [*Hungateiclostridium saccincola*]
552_Contig 82_1700_4123		endo-1,4-beta-xylanase [EC:3.2.1.8]	100% carbohydrate-binding protein [*Hungateiclostridium saccincola*]
552_Contig 94_1095_1781 GH16		endo-1,4-beta-xylanase [EC:3.2.1.8]	70.93% Select seq ref|WP_153836669.1| carbohydrate-binding protein [*Gracilibacillus thailandensis*]
552_Contig 68_5429_6430 GH30	exoK	endo-1,3-1,4-beta-glycanase ExoK [EC:3.2.1.-]	100% Select seq ref|WP_101301458.1| family 16 glycosylhydrolase [*Hungateiclostridium saccincola*]
552_Contig 68_7788_9389 GH8			99.81% hypothetical protein [*Hungateiclostridium saccincola*]
552_Contig 70_114_1295 GH18	bcsZ	endoglucanase [EC:3.2.1.4]	99.74% hypothetical protein [*Hungateiclostridium saccincola*]
552_Contig 70_3623_5536 CE6		yaaH: spore germination protein (second best)	100% Select seq ref|WP_105367824.1| LysM peptidoglycan-binding domain-containing protein [*Hungateiclostridium saccincola*]
552_Contig 70_7432_8478 CE11		pnbA: para-nitrobenzyl esterase [EC:3.1.1.-] (second best)	99.71% Select seq ref|WP_101302863.1| sialate O-acetylesterase [*Hungateiclostridium saccincola*]
552_Contig 81_1_297		fabZ; 3-hydroxyacyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.59]	100% 3-hydroxyacyl-ACP dehydratase FabZ [*Hungateiclostridium saccincola*]

Table 6.6 Putative functions of all the encoded proteins on the assembled DNA fragment.

QUERY	KO	Definition (Blast KAOLA)	Homologous proteins (highest % identity) identified by BLASTP-Organism		Highest nucleotide sequence identity with genes from *X. thermophila*
Contig57ORF1	K00615	E2.2.1.1, tktA, tktB; transketolase [EC:2.2.1.1] (Pentose phosphate pathway)	WP_129721524.1 transketolase	*X. thermophila*	100% identity with RXHQ01000022.1_gene_31
Contig57ORF2	K00853	araB; L-ribulokinase [EC:2.7.1.16]	WP_129721528.1 ribulokinase	*X. thermophila*	99.88% identity with RXHQ01000002.1_gene_61
Contig57ORF3	K10441	rbsA; ribose transport system ATP-binding protein [EC:7.5.2.7]	WP_129721534.1sugar ABC transporter ATP-binding protein	*X. thermophila*	100% identity withRXHQ01000002.1_gene_66
Contig57ORF4	K10548	ABC.GGU.A, gguA; putative multiple sugar transport system ATP-binding protein [EC:7.5.2.-]	WP_129721531.1sugar ABC transporter ATP-binding protein	*X. thermophila*	99.74% identity withRXHQ01000002.1_gene_64
Contig57ORF5	K07484	K07484; transposase	IS66 family transposase	*X. thermophila*	100% identity withRXHQ01000073.1_gene_1
Contig57ORF6	K01804	araA; L-arabinose isomerase [EC:5.3.1.4]	L-arabinose isomerase	*X. thermophila*	100% identity with RXHQ01000002.1_gene_60
Contig57ORF7 (MC57GH51)	K01209	abfA; alpha-L-arabinofuranosidase [EC:3.2.1.55]	alpha-N-arabinofuranosidase	*X. thermophila*	99.73% identity with RXHQ01000002.1_gene_69
Contig57ORF8	K08151	tetA; MFS transporter, DHA1 family, tetracycline resistance protein	WP_129721535.1 MFS transporter	*X. thermophila*	99.84% identity with RXHQ01000002.1_gene_68
Contig57ORF9	K10547	ABC.GGU.P, gguB; putative multiple sugar transport system permease protein	sugar ABC transporter permease	*X. thermophila*	99.91% identity with RXHQ01000002.1_gene_63
Contig57ORF10	K10546	ABC.GGU.S, chvE; putative multiple sugar transport system substrate-binding protein	WP_129721532.1 substrate-binding domain-containing protein	*X. thermophila*	99.82% identity with RXHQ01000002.1_gene_65
Contig57ORF11	K02103	araR; GntR family transcriptional regulator, arabinose operon transcriptional repressor	WP_129721529.1 GntR family transcriptional regulator	*X. thermophila*	99.81% identity with RXHQ01000002.1_gene_62

QUERY	KO	Definition (Blast KAOLA)	Homologous proteins (highest % identity) identified by BLASTP-Organism		Highest nucleotide sequence identity with genes from *X. thermophila*
Contig57ORF12			DeoR/GlpR transcriptional regulator	*X. thermophila*	100% identity with RXHQ01000002.1_gene_66
Contig57ORF13	amiABC	N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	glucosaminidase domain-containing protein	*X. thermophila*	100% identity with RXHQ01000002.1_gene_56
Contig57ORF14	K03077	araD, ulaF, sgaE, sgbE; L-ribulose-5-phosphate 4-epimerase [EC:5.1.3.4]	WP_129721526.1 L-ribulose-5-phosphate 4-epimerase	*X. thermophila*	99.71% identity with RXHQ01000002.1_gene_59
Contig57ORF15			RAO16843.1 hypothetical protein LUPAC07_03011	*Micromonospora noduli*	100% identity withRXHQ01000002.1_gene_64
Contig57ORF16	K00616	E2.2.1.2, talA, talB; transaldolase [EC:2.2.1.2] (Pentose phosphate pathway)	WP_129721525.1 fructose-6-phosphate aldolase	*X. thermophila*	100% identity with RXHQ01000002.1_gene_58
Contig57ORF17			no significant similarity		100% identity with RXHQ01000002.1_gene_57
Contig57ORF18			phosphate transport system permease protein pstA	*Clostridium tetani* E88	100% identity with RXHQ01000002.1_gene_57
Contig57ORF19	K01338	lon; ATP-dependent Lon protease [EC:3.4.21.53]	KJZ86634.1 hypothetical protein ClosIBUN125C_CONTIG40g02382	*Clostridium sp.* IBUN125C	100% identity with RXHQ01000002.1_gene_58
Contig60ORF1	K17318	K17318, lplA; putative aldouronate transport system substrate-binding protein	ABC transporter substrate-binding protein	*Caldicellulosiruptor sp.* YA01	72.83% identity with RXHQ01000016.1_gene_40
Contig60ORF2	K02005	ABC.CD.TX; HlyD family secretion protein	HlyD family efflux transporter periplasmic adaptor subunit	*X. thermophila*	99.20% identity with RXHQ01000002.1_gene_81
Contig60ORF3 (MC60GH51)	K01209	abfA; alpha-L-arabinofuranosidase [EC:3.2.1.55]	WP_129721540.1 alpha-N-arabinofuranosidase	*X. thermophila*	97.21% identity with RXHQ01000002.1_gene_77
Contig60ORF4	K02004	ABC.CD.P; putative ABC transport system permease protein	ABC transporter permease	*X. thermophila*	98.72% identity with RXHQ01000002.1_gene_79

QUERY	KO	Definition (Blast KAOLA)	Homologous proteins (highest % identity) identified by BLASTP- Organism		Highest nucleotide sequence identity with genes from *X. thermophila*
Contig60ORF5	K14982	ciaH; two-component system, OmpR family, sensor histidine kinase CiaH [EC:2.7.13.3]	HAMP domain-containing histidine kinase	*X. thermophila*	99.91% identity with RXHQ01000002.1_gene_83
Contig60ORF6	K00096	araM, egsA; glycerol-1-phosphate dehydrogenase [NAD(P)+] [EC:1.1.1.261]	WP_165000677.1 iron-containing alcohol dehydrogenase	*X. thermophila*	97.54% identity with RXHQ01000002.1_gene_78
Contig60ORF7 (MC60GH43)			amily 43 glycosylhydrolase	*Gorillibacterium massiliense*	No significant similarity found
Contig60ORF8	K02003	ABC.CD.A; putative ABC transport system ATP-binding protein	WP_129721724.1 ABC transporter ATP-binding protein	*X. thermophila*	99.89% identity with RXHQ01000002.1_gene_84
Contig60ORF9	K17320	lplC; putative aldouronate transport system permease protein	carbohydrate ABC transporter permease	*Caldicellulosiruptor changbaiensis*	67.51% identity with RXHQ01000028.1_gene_9
Contig60ORF10	K17319	lplB; putative aldouronate transport system permease protein	PZN08535.1 sugar ABC transporter permease	*Caldicoprobacter oshimai*	86.11% identity with RXHQ01000025.1_gene_22
Contig60ORF11	K18941	arlR; two-component system, OmpR family, response regulator ArlR	WP_129721545.1 response regulator transcription factor	*X. thermophila*	99.70% identity with RXHQ01000002.1_gene_82
Contig60ORF12	K02003	ABC.CD.A; putative ABC transport system ATP-binding protein	ABC transporter ATP-binding protein	*X. thermophila*	100% identity with RXHQ01000013.1_gene_20
Contig60ORF13	K02003	ABC.CD.A; putative ABC transport system ATP-binding protein	WP_129721547.1 ABC transporter ATP-binding protein	*X. thermophila*	99.33% identity with RXHQ01000002.1_gene_85
Contig60ORF14		uncharacterized protein	CRX29690.1hypothetical protein PAERUG_P54_1_London_24_VIM_2_04_13_0616 8	*Pseudomonas aeruginosa*	98.52% identity with RXHQ01000002.1_gene_79
Contig60ORF15			no significant similarity		100% identity with RXHQ01000002.1_gene_84

QUERY	KO	Definition (Blast KAOLA)	Homologous proteins (highest % identity) identified by BLASTP- Organism		Highest nucleotide sequence identity with genes from *X. thermophila*
Contig68ORF1			NLM10898.1 extracellular solute-binding protein	*Clostridiaceae bacterium*	No significant similarity found
Contig68ORF2	K03657	uvrD, pcrA; DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.4.12]	ATP-dependent helicase	*Pseudoclostridium thermosuccinogenes*	No significant similarity found
Contig68ORF3 (MC68GH43-1)	K06113	abnA; arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	NLM10889.1 glycoside hydrolase family 43 protein	*Clostridiaceae bacterium*	No significant similarity found
Contig68ORF4 (MC68GH43-2)	K01190	lacZ; beta-galactosidase [EC:3.2.1.23]	WP_015358799.1 family 43 glycosylhydrolase	*Thermoclostridium stercorarium*	No significant similarity found
Contig68ORF5			no significant similarity		91.67% identity with RXHQ01000003.1_gene_155
Contig68ORF6	K03431	glmM; phosphoglucosamine mutase [EC:5.4.2.10]	WP_065820691.1 phosphoglucosamine mutase	*Thermoclostridium stercorarium*	No significant similarity found
Contig68ORF7			N-acetyltransferase	*Thermoclostridium stercorarium*	95.45% identity with RXHQ01000009.1_gene_57
Contig68ORF8			no significant similarity		No significant similarity found
Contig68ORF9			no significant similarity		No significant similarity found
Contig72ORF1 (MC72GH43-2)	K06113	abnA; arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	PZN11428.1| arabinanase	*Caldicoprobacter oshimai*	No significant similarity found
Contig72ORF2	K09955	K09955; uncharacterized protein	TPA: glycoside hydrolase family 127 protein	*Clostridiales bacterium*	90% identity with RXHQ01000002.1_gene_76
Contig72ORF3 (MC72GH43-1)	K06113	abnA; arabinan endo-1,5-alpha-L-arabinosidase [EC:3.2.1.99]	family 43 glycosylhydrolase	*Paenibacillus castaneae*	No significant similarity found
Contig72ORF4	K03657	uvrD, pcrA; DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.4.12]	no significant similarity		No significant similarity found
Contig72ORF5	K07407	E3.2.1.22B, galA, rafA; alpha-galactosidase [EC:3.2.1.22]	TPA: glycoside hydrolase family 27 protein	*Clostridiales bacterium*	No significant similarity found

Table 6.7 Common sugar mass for MALDI-TOF mass spectrometry.

Product code	Substrates	Molecular Formula	Molecular Weight
Xyloglucan Oligosaccharides			
O-IPRM	Isoprimeverose (Xyloglucan Derived)	$C_{11}H_{20}O_{10}$	312.3
O-XCBIR	Xylosyl-cellobiose (borohydride reduced)	$C_{17}H_{32}O_{15}$	476.4
O-X3G4	Heptasaccharide (X_3Glc_4)	$C_{39}H_{66}O_{33}$	1,062.9
O-X3G4R	Heptasaccharide (X_3Glc_4 - borohydride reduced)	$C_{39}H_{68}O_{33}$	1,064.9
O-XGHON	Xyloglucan (hepta+octa+nona saccharides)		1,064.94, 1,225.06, 1,387.2
O-XGHDP	Higher DP Xyloglucan Oligosaccharides		3,000-4,000
1,4-β-D-Xylooligosaccharides			
O-XBI	Xylobiose	$C_{10}H_{18}O_9$	282.2
O-XTR	Xylotriose	$C_{15}H_{26}O_{13}$	414.4
O-XTE	Xylotetraose	$C_{20}H_{34}O_{17}$	546.5
O-XPE	Xylopentaose	$C_{25}H_{42}O_{21}$	678.6
O-XHE	Xylohexaose	$C_{30}H_{50}O_{25}$	810.7
O-XBIRD	1,4-β-D-Xylobiitol (borohydride reduced)	$C_{10}H_{20}O_9$	284.3
O-XTRRD	1,4-β-D-Xylotriitol (borohydride reduced)	$C_{15}H_{28}O_{13}$	416.4
Glucurono-xylo oligosaccharides/Aldouronic acids			
O-UX	2^2-(4-*O*-Methyl-α-D-Glucuronyl)-xylobiose	$C_{17}H_{28}O_{15}$	472.4
O-UXX	2^3-(4-*O*-Methyl-α-D-Glucuronyl)-xylotriose	$C_{22}H_{36}O_{19}$	604.5
O-UXXR	Aldotetrauronic acid (terminal substitution, borohydride reduced)	$C_{22}H_{38}O_{19}$	606.5
O-XUX	2^2-(4-*O*-Methyl-α-D-Glucuronyl)-xylotriose	$C_{22}H_{36}O_{19}$	604.5
O-XUXX	2^3-(4-*O*-Methyl-α-D-Glucuronyl)-xylotetraose	$C_{27}H_{44}O_{23}$	736.6
O-XUXXR	Aldopentauronic acid (internal substitution, borohydride reduced)	$C_{27}H_{46}O_{23}$	738.6
O-AMX	Aldouronic Acids Mixture		
O-XA3XX	3^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX)	$C_{25}H_{42}O_{21}$	678.6
O-XAXXMIX	3^3-α-L- plus 2^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX/XA^2XX) mixture	$C_{25}H_{42}O_{21}$	678.6
O-A2XX	2^3-α-L-Arabinofuranosyl-xylotriose (A^2XX)	$C_{20}H_{34}O_{17}$	546.5
O-A3X	3^2-α-L-Arabinofuranosyl-xylobiose (A^3X)	$C_{15}H_{26}O_{13}$	414.4
O-A23XX	2^3,3^3-di-α-L-Arabinofuranosyl-xylotriose ($A^{2,3}XX$)	$C_{25}H_{42}O_{21}$	678.6
O-XUXX	2^3-(4-*O*-Methyl-α-D-Glucuronyl)-xylotetraose	$C_{27}H_{44}O_{23}$	736.6
O-XA23XX		$C_{25}H_{42}O_{21}$	678.6

6.4 Characterization of enzymes from either the genome of isolated strain or from metagenomic library screening

Table 6.8 Specific activity of characterized enzymes against different substrates.

	MC57 GH51	MC60 GH43	MC68 GH43-2	MC72 GH43-1	MC72 GH43-2	MC68 GH43-1	MC60 GH51	M_Xyn10	*Am*Araf51	*Am*Araf43	*Am*Pgl28	*Pp*Abn43
BWX	-	-	-	-	-	-	-	+	-	ND	-	-
WAX	-	-	-	-	-	-	-	+	+ (weak)	+	-	-
SBA	+	-	-	-	+	+	+	-	+	-	-	+
DA	+ (weak)	+	+	-	+	+	+	-	+	-	-	+
PGP	-	+	+	-	+	-	+	-	-	-	-	ND
PGL	-	-	+	-	+	-	+ (weak)	-	-	-	-	ND
GL	-	-	+ (weak)	-	-	-	-	-	-	-	-	ND
GP	-	-	-	-	-	-	-	-	-	-	-	ND
LAG	-	-	-	-	-	-	-	-	-	-	-	ND
WAX-RS	-	-	-	-	-	-	-	+	+ (weak)	+	-	ND
PGA	-	-	-	-	-	-	-	-	-	-	+	ND
MBG	-	-	-	-	-	-	-	-	-	-	-	ND
Rham	-	-	-	-	-	-	-	-	-	-	-	ND
HMP	-	-	-	-	-	-	-	-	-	-	-	+
LMP	-	-	-	-	-	-	-	-	-	-	-	ND
CMC	-	-	-	-	-	-	-	-	-	-	-	ND

BWX: Beechwood xylan, WAX: Wheat arabinoxylan, SBA: Sugar beet arabinan, DA: Debranched arabinan, PGP: Pectic galactan (potato), PGL: Pectic galactan (Lupin), GL: Galactan (ex. Lupin), GP: Galactan (ex. potato), LAG: Larch arabinogalactan, WAX-RS: Wheat arabinoxylan for reducing sugar assay, PGA: Polygalacturonic acid, MBG: Mixed linkage β-glucan, Rham: Rhamnogalacturonan, HMP: HM pectin, LMP: LM pectin, CMC: Carboxymethyl cellulose, ND: Not detected.

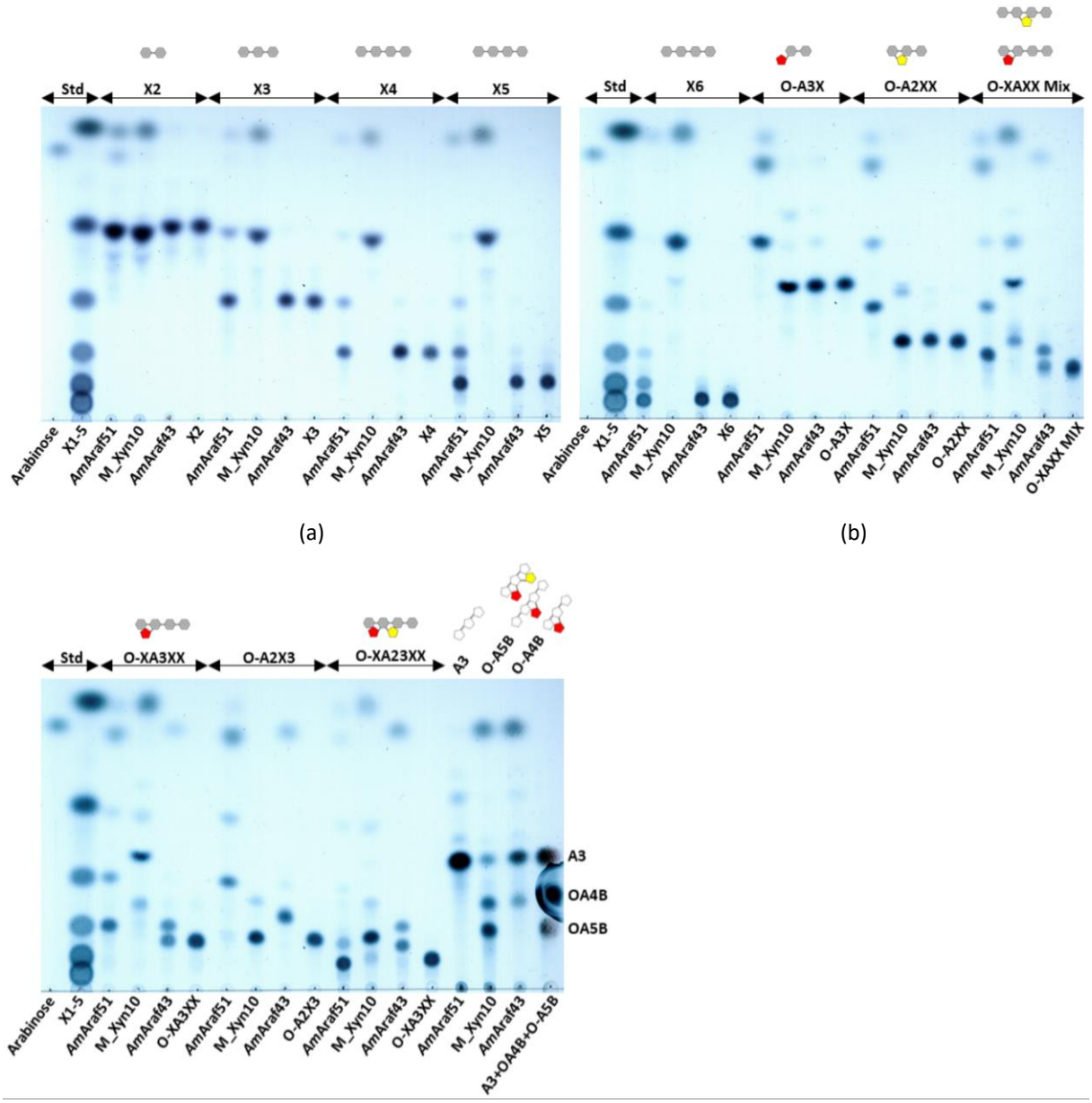
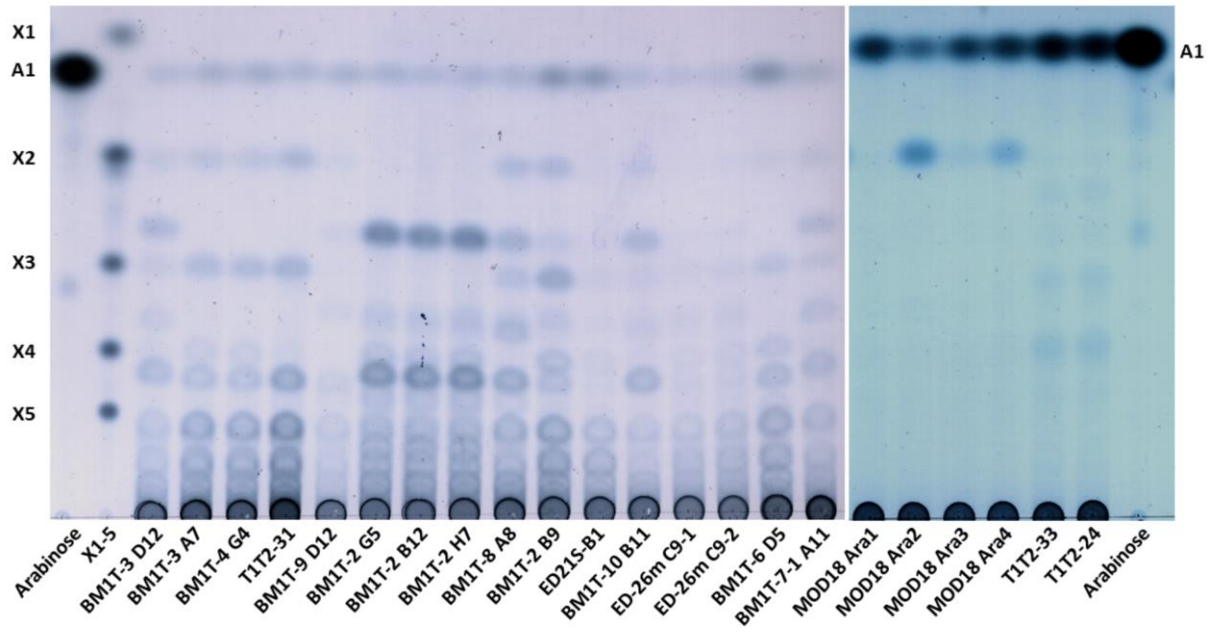
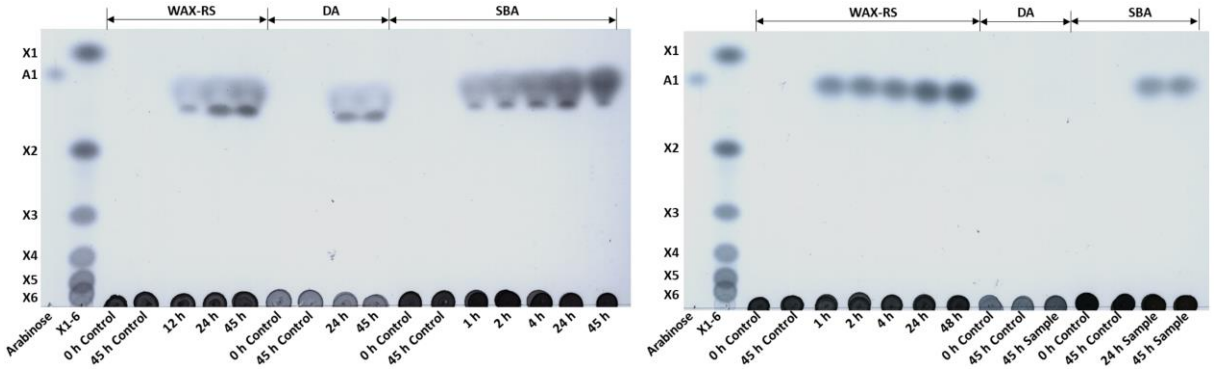


Figure 6.14 TLC analysis of the hydrolysis products of XOS, AXOS and AOS by *Am*Araf51, *Am*Araf43 and M_Xyn10, respectively. Standard reactions were performed by incubating 85 nM *Am*Araf51, 82 nM *Am*Araf43 or 168.78 nM M_Xyn10 with 0.25% of different XOS at 60°C and pH 6.0 for *Am*Araf51, at 40°C and pH 5.0 for *Am*Araf43, at 50°C and pH 6.0 for M_Xyn10 for 24 h.



(a) (b)

Figure 6.15 TLC analysis of the hydrolysis products released from WAX and DA by crude extract of positive fosmid clones. (a) WAX as a substrate, (b) SBA as a substrate. 40 µL cell crude extract was incubated with 0.5% BWX/WAX for 24 h in a 200 µL of enzyme reaction.



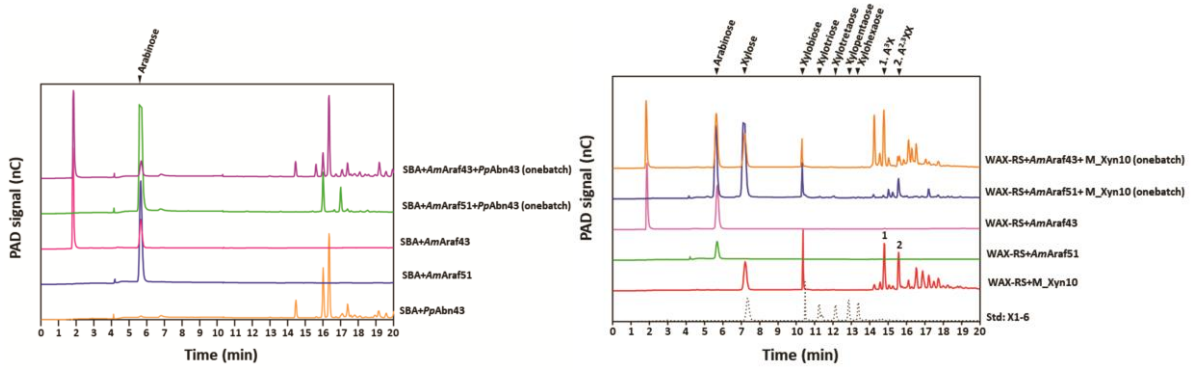
(a) (b)

Figure 6.16 TLC analysis of the hydrolysis products released from WAX-RS, DA and SBA by *Am*Araf51 and *Am*Araf43. (a) *Am*Araf51 (b) *Am*Araf43. Standard reactions were performed by incubating each substrate with 2.15 µM of *Am*Araf51 and 2.01 µM of *Am*Araf43 at pH 6.5, 50°C for different periods.

Table 6.9 Arabino-oligosaccharides (AOS) and arabinoxylo-oligosaccharides (AXOS) used in this study.

Product code	Name	Structure
O-A3X	3^3-α-L-Arabinofuranosyl-xylobiose (A^3X)	
O-A2XX	2^3-α-L-Arabinofuranosyl-xylotriose (A^2XX)	
O-XA3XX	3^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX)	
O-A2X3	$2^2,3^2$-di-α-L-Arabinofuranosyl-xylotriose ($A^{23}XX$)	
O-XA23XX	$2^2,3^2$-di-α-L-Arabinofuranosyl-xylotretraose ($XA^{23}XX$)	
O-XAXX MIX	3^3-α-L- plus 2^3-α-L-Arabinofuranosyl-xylotetraose (XA^3XX/XA^2XX) mixture	
O-A4B	3^2-α-L-Arabinofuranosyl-(1,5)-α-L-arabinotriose (AA^3A)	
O-A5BMIX	$2^2,3^2$-di-α-L-Arabinofuranosyl-(1,5)-α-L-arabinotriose plus 3^2-α-L-arabinofuranosyl-(1,5)-α-L-arabinotetraose ($AA^{23}A$+AAA^3A)	

The oligosaccharides were named following the Megazyme Nomenclature.



(a) (b)

Figure 6.17 HPAEC-PAD analysis of the hydrolysis products of SBA or WAX-RS by *Am*Araf51 and *Am*Araf43 involved mini-enzyme cocktails. 0.5% (w/v) SBA (a) or WAX-RS (b) was incubated with endo-active enzymes (0.90 µm *Pp*Abn43 or 2.9 µm M_Xyn10) and exo-active enzymes (0.3 µm *Am*Araf51 or 0.3 µm *Am*Araf43) at 50°C and pH 5.5 for 24 h incubation.

Table 6.10 Quantification of arabinose and xylose released from arabinan and arabinoxylan.

Reactions	Arabinose (mg L^{-1})								
Arabinose from WAX-RS	**0 h**	**1 h**	**2 h**	**8 h**	**12 h**	**13 h**	**14 h**	**18 h**	**24 h**
WAX-RS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
M_Xyn10	0.00	0.00	0.00	0.00	0.00			0.00	0.00
*Am*Araf51	0.00	0.00	0.00	0.00	24.48	—	—	38.27	61.38
*Am*Araf43	0.00	383.83	410.45	443.82	455.29	—	—	437.96	463.57
Onebatch_M_Xyn10+*Am*Araf51	0.00	209.74	324.02	461.03	—	—	—	505.83	845.06
Onebatch_M_Xyn10+*Am*Araf43	0.00	337.59	340.82	405.67	376.88	—	—	379.30	564.45
Stepwise_M_Xyn10+*Am*Araf51	0.00	—	—	0.00	0.00	0.00	160.14	206.43	378.65
Stepwise_M_Xyn10+*Am*Araf43	0.00	—	—	0.00	0.00	0.00	355.09	274.07	308.48
Xylose from WAX-RS									
M_Xyn10	0.00	113.25	125.14	121.64	161.91	—	—	177.48	213.51
*Am*Araf51	0.00	0.00	0.00	0.00	0.00	—	—	0.00	0.00
*Am*Araf43	0.00	0.00	0.00	0.00	0.00	—	—	0.00	0.00
Onebatch_M_Xyn10+*Am*Araf51	0.00	196.58	308.27	480.26	—	—	—	641.76	1059.03
Onebatch_M_Xyn10+*Am*Araf43	0.00	18.24	102.71	227.46	—	—	—	291.84	306.50
Stepwise_M_Xyn10+*Am*Araf51	0.00	—	—	0.00	0.00	179.72	117.81	131.68	170.43
Stepwise_M_Xyn10+*Am*Araf43	0.00	—	—	200.08	218.28	219.44	246.57	185.90	196.22
Arabinose from SBA									
SBA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
*Am*Araf51	0.00	247.82	378.20	462.20	629.53	—	—	682.40	800.39
*Am*Araf43	0.00	0.00	0.00	34.78	69.70	—	—	126.74	130.49
*Pp*Abn43	0.00	0.00	0.00	0.00	0.00	—	—	0.00	0.00
Onebatch_SBA_*Pp*Abn43+*Am*Araf51	0.00	886.99	1,138.85	1,571.07	—	—	—	1,512.56	1,656.00
Onebatch_SBA_*Pp*Abn43+*Am*Araf43	0.00	0.00	0.00	0.00	—	—	—	0.00	163.46
Stepwise_*Pp*Abn43+*Am*Araf51	0.00	0.00	0.00	0.00	0.00	563.06	748.58	1,364.97	1,440.58
Stepwise_*Pp*Abn43+*Am*Araf43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	228.97

VIII Statutory declaration

I declare in lieu of an oath that I have written this thesis independently and that I have only used the aids offered in accordance with § 6 Ab. 6 and 7 Satz 2 of the doctoral regulations. Only the sources and aids expressly named in the thesis have been used. I have marked as such any ideas taken over verbatim or in spirit.

Date, Place Signature

IX Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give a particular mention here.

My first round of thanks must go to my supervisor, Prof. Dr. Wolfgang Liebl, who gave me the possibility to conduct my dissertation at the Chair of Microbiology of TUM. He is always very patient and pleasant. He provides valuable insights into the microbiology field and encouraged me to open my eyes to learn about the research progress in other areas by attending the international conference. He gave me a lot of good advice when I had a problem with my experiment but never pressure me. He has worked tirelessly with the reading and correction process of my manuscripts and my thesis. Many things make me think he is a great person.

Besides, I would like to thank my group leader Dr. Sonja Vanderhaeghen and my mentor Dr. Angel Angelov for the opportunity to work in their fantastic research group on an exciting and current research topic. They are very motivational in science and very hardworking, which set a perfect example for me in this direction. They are also very patient and kind, always provide me a lot of help when I need it. I also want to thank Dr. Vladimir Zverlov, who kindly offer me the enzymes and substrates used in my study.

Specifically, I would like to express my sincere thanks to my colleague Maria Überlecker, who provided expert technical support and always shared her experience with me unreservedly. I want to say thanks to my friend Rui Han, she is always full of sunshine and bring our group a lot of happiness. I would like to thank my student Werner Feiler, my colleagues Luis Serrano, Dr. Nils Thieme, Melanie Baudrexl, Dr. Maria Cecilia Rasuk, Enzo Joaquín Torasso Kasem, Martin Bimmer, my technician Kornelia Garus, Sibylle Schadhauser. All of you made my doctoral experience wonderful, I will never forget.

I would like to express my gratitude to my parent, who always support the decision I made in my life and provide me the safety during my growth. I want to acknowledge the China Scholarship Council's financial support, which allows me to study in Germany and finish my degree. I also want to acknowledge the financial support of the Technical University of Munich for the international conference in Brazil.

X Resume

Education

11.2021 – Until now: Technical University of Munich, Chair of Chemistry of Biogenic Resources, **Postdoc**.

10.2017 – Until now: Technical University of Munich, Chair of Microbiology, **PhD study**, Microbiology.

09.2014 – 06.2017: Southwest University, College of Animal Science and Technology, **Master**, Microbiology.

09.2010 – 06.2014: Nanyang Normal University, College of Life Science, **Bachelor**, Biological Engineering.