# Novel bacterial enzymes for plant biomass degradation discovered by meta-omics approach

The research reported in this thesis was carried out at the Microbial Ecology group, which is part of the Genomics Research in Ecology and Evolution in Nature (GREEN) research group, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, The Netherlands.

The research was supported by the BE-BASIC foundation (http://www.be-basic.org/).

Layout design: Mukil Maruthamuthu
Cover design: Mukil Maruthamuthu

Printed by: Ridderprint B.V (https://www.ridderprint.nl/)

ISBN: 978-94-6299-682-3 (printed version)
ISBN: 978-94-6299-683-0 (electronic version)

rijksuniversiteit
groningen

# Novel bacterial enzymes for plant biomass degradation discovered by meta-omics approach

**PhD thesis**

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 11 September 2017 om 16:15 hours

by

**Mukil Marutha Muthu**

born on 14 April 1981
in Pollachi, India

**Promotor**
Prof. J.D. van Elsas

**Assessment Committee**
Prof. J.T.M. Elzenga
Prof. D.B. Janssen
Prof. H. van Veen

*Ultimately, education in its real sense is the pursuit of truth.*
*It is an endless journey through knowledge and enlightenment*
*A. P. J. Abdul Kalam*

# *Contents*

# Chapter 1

# General Introduction

Mukil Maruthamuthu

## Introduction

Prokaryotes constitute the first organisms on Earth; they have dominated the biosphere, playing central roles in biological systems [1]. Estimates have consistently suggested that the immense majority of prokaryotic microorganisms in natural environments is difficult to cultivate [2]. "Uncultivable" indicates that current bench methods are unable to grow such microorganisms in the laboratory. This implies a lack of knowledge of their biological traits and mechanisms that allow growth. Currently, there are ample opportunities – using direct molecular techniques - to assess the molecular traits that reside in such cryptic organisms, allowing us to gain access to previously hidden metabolic diversity. Thus, by applying such novel methods, new natural products can be obtained and the factors that contribute to both ecological balance and health revealed. The genetic information (DNA) present in ecosystems is responsible for programming the physical characteristics of the organisms living therein. Differences in DNA sequences between organisms create diversity. Such diversity manifests itself as biological diversity through the structure, organization, regulation and expression of DNA. There are ample data that indicate that diversity is extremely high in the Prokaryotes [3]. In contrast, the diversity in the Eukarya is much lower, although still considerable and concentrated in the microbial members. The biochemical properties carried in this genetic diversity are thought to support the stability (resilience) of natural systems. To unlock the genetic diversity contained within microbiomes, including both "cultivable" and "uncultivable" microorganisms, the use of metagenomics is very useful [4]. This approach can reveal how the members of microbial communities function and relate to their niches, enabling inferences about processes and their rates in the systems studied.

## Importance and potential of biological catalysts

In biotechnological processes, there is an increasing demand for biological catalysts (enzymes) with novel properties, to be used in all aspects of the biochemical reactions that are pursued (such as increased reaction rate, faster kinetics, better specificity and being non-hazardous to nature) [5]. Biocatalyst are widely used in the detergent, textile, cosmetics, leather, pharmaceutical, biotechnology, chemical, food processing and brewing industries, to improve the productivity and efficiency of industrial processes [6]. In addition, the environmental burden of industrial processes needs to be reduced and - in this process - enzymes are also important [7]. Recent research has explored the potential of enzymes (known or unknown ones) to mediate new catalytic functions that are difficult to achieve through chemical catalysis [5]. Moreover, improvement of existing enzymes is receiving great interest [8]. In particular, *in silico* protein design tools are playing vital roles in efforts that aim to foster the characterization of biocatalysts [9]. Thus, enzymes are considered that have specific properties, e.g binding to target positions of substrates with better affinity

and specificity, allowing optimized catalytic activity. For example, the efficiency of production of sugar monomers from lignocellulosic (plant biomass) material poses quite a challenge in the light of the demand by Industry [10]. Thus, carbohydrate-active enzymes (CAE; cellulases, lipases/esterases and xylanases) are required. There is a huge diversity of CAE, which have been grouped into more than 276 families, among which glycosidases (135), transferases (101), lyases (24) and esterases (16) (CAZy: Carbohydrate-Active Enzyme, (http:// www.cazy.org/) [11]. Only a small part of this diversity has been unlocked so far. Thus, accessing the enzyme diversity in natural or derived systems will allow to explore the wide catalytic potential of enzymes as related to relevant reaction contexts.

**From microbial ecology to biotechnology**

Recent assessments estimate that about $10^{30}$ bacterial cells inhabit Earth, with 2.6 x $10^{29}$ cells in terrestrial and 1.2 x $10^{29}$ in marine systems [12]. Only part of this huge biomass is accessible by cultivation; studies in marine systems indicated extremely low levels, i.e. 0.001–0.1%, of culturability and those in terrestrial systems about 1% culturability [13]. Given this general lack of culturability of extant microorganisms, the use of metagenomics is indicated to make progress with explorations for biotechnology. Moreover, such an approach will also foster fundamental ecological studies on microbial behavior and survival. Thus, if evidence for the growth of particular microorganisms in particular conditions in an ecosystem can be found, this will enable the discovery of novel enzymes that are required for growth and survival under those conditions [14]. For example, recent studies have shown that the addition of chitin to soil enhances the bacterial abundance along with increased chitinase activities, allowing the unlocking of novel chitinases [15]. Thus, an interplay exists between our attempts to understand the functioning of ecosystems and to foster biotechnology, in terms of the discovery of novel enzymes or antibiotics. Historically, the discovery of streptomycin and other antibiotics sprang from very basic studies of the taxonomy and ecology of actinomycetes in soil, as conducted by soil microbial ecologists [16]. Similarly, turbomycin, one of the first antibiotics discovered by a metagenomics approach, was identified in a soil-derived library [17], in a clone that had hemolytic activity. Thus, to optimize or maximize biotechnological applications, it is essential that both basic biology assessments and utility screens are pursued in metagenomics [6].

**Metagenomics**

Metagenomic gene discovery studies need fine tuning to allow the retrieval of enzymes with potential useful applications [18]. Here, the approaches, in terms of the organisms being accessed and mined, are complex [19]. First, suitable microbiome DNA needs to be produced. Second, suitable genetic fragments need to be cloned into suitable hosts and then genes should be expressed to allow the

detection of useful functions (Figure 1). In effect, many metagenomics approaches can be seen as being analogous to the classical genome library construction and screening method, with the difference that the 'genome' cloned is not from a single organism, but rather from the entire microbial community present in an environmental sample. It is, thus, a reflection of the 'community genome'. 'Genome' coverage - in this case - is an ephemeral notion, since different community members will be present in different numbers in the sample, and their genomes will be extracted with different efficiencies. Hence, genes of different organisms will be present in very different, and often completely serendipitous or idiosyncratic prevalences in the DNA used to construct metagenomic libraries [20].

Different strategies can be taken in metagenomics, according to the primary goal. Small (plasmid-based) or large (fosmid-, cosmid- and/or bacterial artificial chromosome (BAC)-based) insert libraries can be constructed for archiving and sequence homology screening purposes (Figure 1). Some genes in such libraries may be expressed in the cloning host and may thus be found in activity screens, however there will be a bias towards genes from organisms related to the host [21]. The reason may be that there is greater biochemical compatibility among closely-related that across distantly related organisms. Moreover, selection of the cloning vector is often guided by the need for insertion of the adequate size of DNA inserts. Here, as inserts between 20 to 40 Kb were required, fosmids were chosen, as these vectors work efficiently and accommodate these insert sizes [22].

**Metagenomics: method, importance and challenges**

Two basic approaches have been used for the recovery of novel biocatalysts from different environmental sources by metagenomics. In both cases, shotgun libraries are produced in suitable vectors. Then, the difference between the two approaches are determined by the library screening method applied: function- and sequence - based screening of metagenomic libraries (Figure 1).

***Sequence-based metagenomics*** - In sequence-based metagenomics screening, libraries are screening by sequencing, and comparisons of sequences with databases will give information about the presumed types of genes that are present. Coupled with sequencing and exploration of DNA from the sample source, this method will allow us to infer potential function, next to assemble operons, genomes and genes (Figure 1) [23]. Moreover, this application encompasses the design of DNA probes and/or primers that are derived from already-known genes or protein families. In this way, only novel variants of identified functional group of proteins can be screened [24]. In contrast, ***function-based metagenomics*** is based on the screening for a specific function or activity as the basis for clone selection (Figure 1) [23]. Function-based screening plays a key role in identifying new products for biotechnological applications. Bottlenecks are

the often-low expression of genes in metagenomic libraries, the screening for specific genes without a clear phenotype and (substrate) induced genes. However, function-based screening has been successfully used to recognize novel enzymes (For example: agarases [25], amylases [26], esterase and lipases [25], glycosyl hydrolase [27])and other products [28]. It is the only approach that enables to identify genes that effectively encode novel functions [4,18].

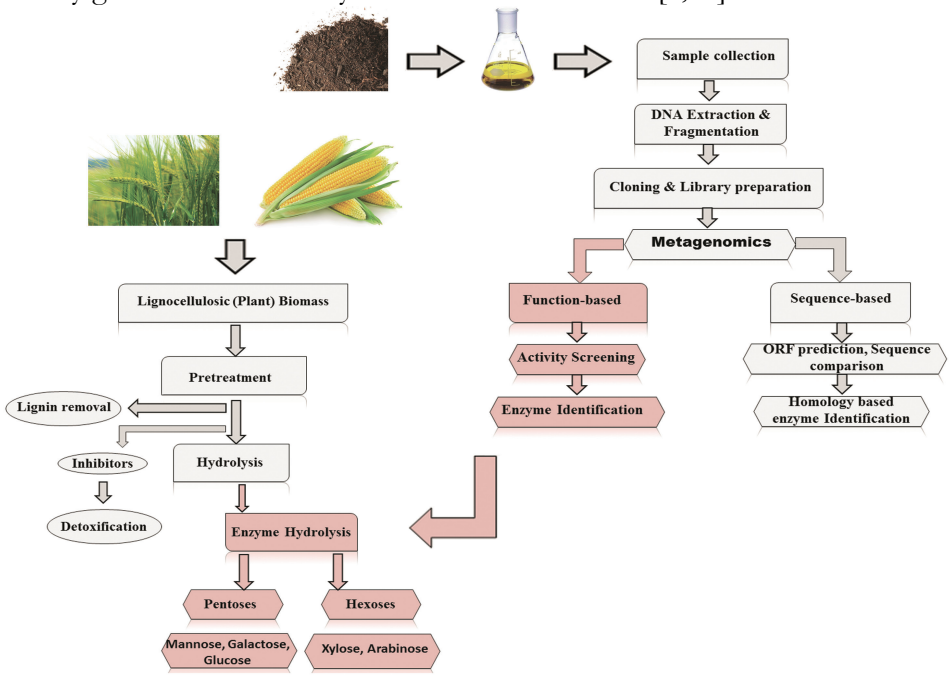



**Figure 1.** Flow chart of our work

## Plant biomass

Plant secondary cell walls are formed by specific plant cell type. In those cells with structural support, secondary cell walls are dynamic in different aspects, e.g. water transport [29]. The secondary cell wall is composed of the main polymers cellulose, hemicellulose, lignin and proteins (Figure 2). Building the complex cell wall structure encompasses a harmonized execution of different biosynthetic pathways (Figure 2). These include the biosynthesis and inter-conversion of precursor sugars, the polymerization of precursors into large polymers, the transport of precursors or polymers to the cell wall and the final assembly of component polymers into the cell wall structure (Figure 2). The secondary structure that is formed is often rigid and highly stable and does not easily decay. Upon decay, however, the major polymers used for energy production are, next to proteins, cellulose and hemicellulose. The latter are considered to be the most energy-rich polysaccharides. Evolution has brought these biopolymers together in very prominent architectures, foremost to boundless biodiversity from the molecular to

the macroscopic level. They are hierarchically organized (Figure 2); cellulose is embedded in a complex matrix composed of hemicellulose and lignin.

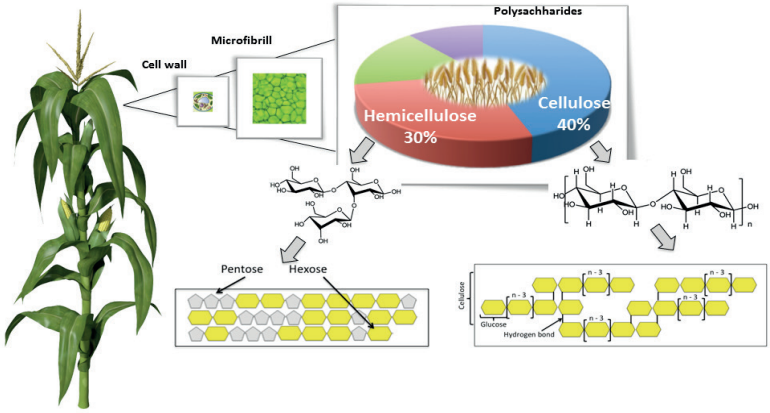



**Figure 2:** Structure of lignocellulose

***Classification of lignocellulosic biomass*** - The presence of polysaccharides as well as the proportion of the different sugars of the hemicelluloses differs between different plants (Table 1). In contrast, ***cellulose*** is considered as an abundant homo-polysaccharide of the plant cell wall across the board. It is composed of chains with high numbers of D-glucose linked together by β-1, 4-glucosidic bonds (Figure 3). The linear structure of the cellulose chain (formed with hydrogen bonds) leads to crystalline fibrils. In these, the hydrogen bonds of the cellulose make the structure flexible, insoluble in most solvents and quite resistant to microbial degradation [30,31]. The surface of cellulose is hydrophobic, which results in the formation of a repelled layer of water that may hinder diffusion of enzymes and degradation products near the cellulose surface [32].

***Hemicelluloses*** are complex heterogeneous polysaccharides composed of D-glucose, D-galactose, D-mannose, D-xylose, L-arabinose, D-glucuronic acid and 4-O-methyl-D-glucuronic acid (Figure 3). Hemicelluloses have a degree of polymerization below 200, and side chains can be acetylated [33]. They are classified according to the main sugar of the backbone of the polymer, e.g. xylan (β-1,4-linked xylose) or mannan (β-1,4-linked mannose) [34,35]. Plants belonging to the grass family (Poaceae), e.g. rice, wheat, oat and switch grass, have hemicelluloses that are composed of mainly glucurono arabinoxylans [36]. Due to differences in hemicellulose composition, agricultural waste products like wheat straw and corn stover, as well as hardwood materials, are rich in the pentose sugar xylose, whereas softwoods are rich in the hexose sugar mannose (Table 1) [34,35,37–40].

***Lignin*** is a complex network formed by polymerized phenyl propane units. It constitutes the most abundant non-polysaccharide fraction in lignocellulose. The

three monomers in lignin are p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol that are joined through alkyl–aryl, alkyl–alkyl and aryl–aryl ether bonds. Lignin embeds the cellulose, thereby offering protection against microbial and chemical degradation. Furthermore, it is able to form covalent bonds with some hemicelluloses, e.g. benzyl ester bonds with the carboxyl group of 4-O-methyl-D-glucuronic acid in xylan. More stable ether bonds, also known as lignin carbohydrate complexes (LCC), can be formed between lignin and arabinose or galactose side groups in xylans and mannans [41]. In general, herbaceous plants, such as grasses, have the lowest content of lignin, whereas softwoods have the highest lignin content (Table 1).

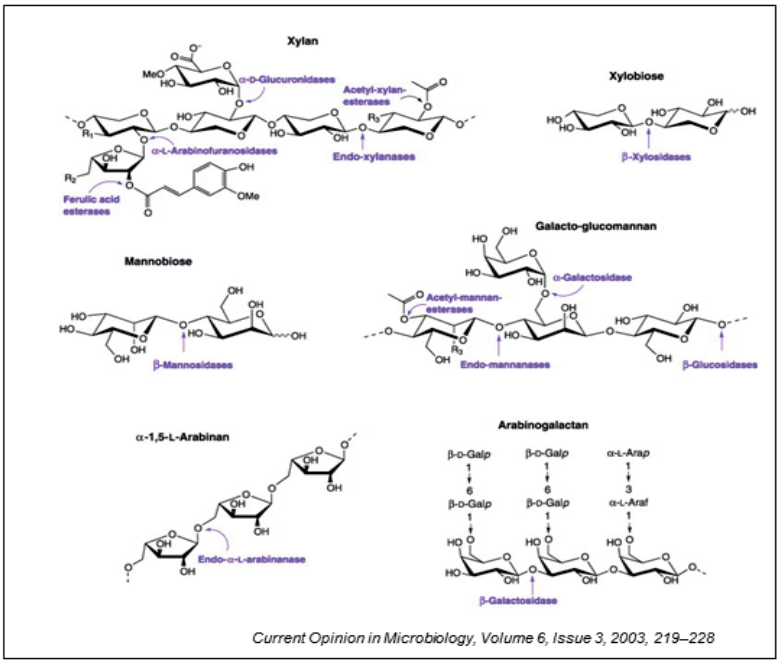



**Figure3:** Structural components of hemicellulose

## Pretreatment of lignocellulosic materials

The saccharification of (hemi)cellulose from lignocellulose is hindered by different factors of physico-chemical or structural nature. A mechanical or physical (pre) treatment may aid in the opening up of the structures, so as to obtain the polysaccharide fibers (Figure 1). Several pretreatment techniques are currently playing roles in this process, including ammonia fiber explosion, steam explosion (SE), chemical and biological treatments. Steam explosion appears as a method of choice (Garrote *et al.*, (1999), as it uses no chemicals, yields high amounts of hemicelluloses with low degraded byproducts, does not cause equipment corrosion, is user-friendly (no acids) and allows the disruption of the solid residues

from bundles to individual fibers. However, it may produce inhibitory compounds such as furfural and hydroxyl methyl furfural (HMF). In effect, steam-explosion was an effective process for increasing methane yield from wheat straw (upto 30%), as compared to untreated wheat straw [43]. Although (energy) expensive, SE has great potential for the industrial scale [44]. This process has to meet several requirements: (1) increase the production of sugars or the enzyme convertibility, (2) reduce the loss of sugars and avoid the degradation of carbohydrate, (3) avoid the formation of inhibitors for further processes, and (4) be scalable to industrial level. However, in practice it is very difficult to consider all issues. Clearly, loss of sugars and cost effectivity are points of concern in scaling up. In fact, the combination of steam explosion and enzyme hydrolysis has been recommended as a standard technology that best meet current industrial demands [45]. During steam explosion, the lignocellulosic material pore size can be increased, however, the method does not affect the microfibrils [46]. Thus, enzymes can have easy access to the microfibrils to increase hydrolyses process.

**Screening microbiomes for glycosyl hydrolases**

The enzymes that break glycosidic bonds between sugar residues or a sugar and a non-sugar moiety within carbohydrates or oligosaccharide are known as glycoside hydrolases. The glycoside hydrolase family enzymes are scattered throughout the prokaryotic, eukaryotic and archaeal domains [47]; they have remarkable functional diversity. Till now, 135 glycoside hydrolase (GH) families were identified. The classification of GH enzymes based only on substrate specificities is inappropriate, since the same protein fold often harbors several types of specificities. Therefore, one key study proposed a better classification pattern based on the consequence of protein folding, as dictated by the amino acid sequence [47]. The 135 different GH families provide insight into the comparative structural features within a family, their evolutionary relationships with other family members and their mechanisms of action. Through functional metagenomics, several carbohydrate-active enzymes (glycosyl hydrolase family) have been identified using p-nitrophenyl tagged substrates. For example, p-nitrophenyl-β-D-glucopyranoside and p-nitrophenyl-α-L-arabinopyranoside were used to screen for (hemi)cellulose hydrolytic enzymes in *E. coli* clones harboring metagenomic fosmid libraries derived from (hemi)cellulose-depleting microbiomes of a fresh cast of earthworms. This study revealed two glycosyl hydrolases that had no resemblance (63% and 75% of identity with hypothetical proteins) to any identified glycosyl hydrolase family enzymes; two novel families (G03-3 and G04-9) of β-galactosidases/α-arabinopyranosidases were thus defined [48]. The purified G03-3 and G04-9 enzymes were tested against *pNP*-β-D-galactoside (*pNP*Gal) and *pNP*Ara. Remarkably, these two enzymes showed approximately 90- and 2-fold higher catalytic efficiencies with β-galactoside than with α-arabinopyranoside. *In-silico* structural alignment and 3D modelling were performed, and these methods

predicted (based on the crystal structure) similarity to the UDP-glucose epimerase of *Thermus thermophilus*. Thus the analyses suggested that the two enzymes belong to two new families of functional GHs with β-galactosidase (EC 3.2.1.23) and α-arabino pyranosidase activities

Table 1: Presence of polysaccharides in lignocellulosic materials

| Samples | Cellulose (%) | Hmicellulose (%) | Lignin (%) | Ref |
|---|---|---|---|---|
| **Grasses** | | | | |
| Wheat straw | 38,2 | 24 | 23,4 | [58] |
| Rice straw | 34,2 | 24,5 | 11,9 | [58] |
| Corn stover | 35,6 | 22,1 | 12,3 | [59] |
| **Hardwood** | | | | |
| Birch | 38,2 | 19,7 | 22,8 | [59] |
| Willow | 43 | 29,3 | 24,2 | [60] |
| **Softwood** | | | | |
| Spruce | 43,4 | 18 | 28,1 | [61] |
| Pine | 46,4 | 22,9 | 29,5 | [58] |

**Importance of enzymes for hydrolysis of lignocellulose**

Efficient hydrolysis of lignocellulose requires a number of efficient enzymes that work on the cellulose, hemicellulose and lignin moieties of the substrate. Concerning the required enzymes for cellulose, cellulolytic enzymes are divided into three major classes; *i*) cellobiohydrolases (CBH) (EC 3.2.1.91), which break the cellulose chain into cellobiose units from the ends; *ii*) endo-1,4-β-D-glucanases (EG) (EC 3.2.1.4), which hydrolyse internal β-1,4-glucosidic bonds; *iii*) 1,4-β-D-glucosidases (EC 3.2.1.21), which hydrolyse cellobiose to glucose and also cleave glucose units from cello oligosaccharides. All these enzymes work synergistically to hydrolyse cellulose by creating new accessible sites for each other, removing obstacles and relieving product inhibition [49]. The degradation of hemicellulose requires several tasks from the hemicellulases (Figure 3). The main structural part in hemicellulose is xylan, which is composed of d-xylopyranosyl units linked by β-1,4-glycosidic bonds. The hemicellulase system includes as enzymes, among others, endo-1,4-β-D-xylanases (EC 3.2.1.8), which hydrolyse internal bonds in the xylan chain; 1,4-β-D-xylosidases (EC 3.2.1.37), which attack xylo oligosaccharides from the non-reducing end and liberate (EC 3.2.1.22), α-D-arabinofuranosidases (EC 3.2.1.55), α-glucuronidases (EC 3.2.1.139), acetyl xylan esterases (EC 3.1.1.72) and feruloyl and p-cumaric acid esterases (EC 3.1.1.73) [50,51]. Although several of these glycosyl hydrolase family enzymes have been characterized and crystallized to better understand the structural functions, there is a paucity of knowledge about the full diversity across these (Table 2). While enzymes from organisms in natural systems exert habitually selective catalytic activities, these are often not appropriate for the burdens of industrial processes in which greater xylose; endo-1,4-β-D-mannanases (EC 3.2.1.78), which cleave internal bonds in

Table 2: Important (hemi)cellulases compiled from CAZy* and PDB* databases

| Enzyme function | EC Number | GH Family | Number of PDB entries | | | Number of Genes |
|---|---|---|---|---|---|---|
| | | | Bacteria | Eukaryota | Archaea | |
| **Cellulases** | | | | | | |
| cellobiohydrolases (CBH) | EC 3.2.1.91 | 5, 6, 9 | 30 | 87 | ND | 28 |
| endo-1,4-β-D-glucanases (EG) | EC 3.2.1.4 | 5, 6, 7, 8, 9, 10, 12,26, 44, 45, 48, 51, 74, 124 | 188 | 51 | 17 | 85 |
| 1,4-β-D-glucosidases | EC 3.2.1.21 | 1, 2, 3, 5, 9, 30, 116 | 63 | 85 | 3 | 38 |
| **Hemicellulases** | | | | | | |
| endo-1,4-β-D-xylanases | EC 3.2.1.8 | 3, 5, 8, 9, 10, 11, 12, 16, 26, 30. 43. 44. 51. 62. 98. | 241 | 81 | ND | 56 |
| 1,4-β-D-xylosidases | EC 3.2.1.37 | 1, 3, 5, 30, 39, 43, 51, 52, 54, 116, 120 | 19 | 2 | ND | 11 |
| endo-1,4-β-D-mannanases | EC 3.2.1.78 | 5, 9, 26, 44, 113, 134 | 33 | 17 | ND | 17 |
| 1,4-β-D-mannosidases | EC 3.2.1.25 | 1, 2, 5 | - | - | - | - |
| α-D-galactosidases | EC 3.2.1.22 | 4, 27, 31, 36, 57, 97, 110 | 22 | 24 | ND | 14 |
| α-D-arabinofuranosidases | EC 3.2.1.55 | 2, 3, 10, 43, 51, 54, 62 | 37 | 19 | ND | 24 |
| α-glucuronidases | EC 3.2.1.139 | 4, 67 | - | - | - | - |
| acetyl xylan esterases | EC 3.1.1.72 | 5, 11 | 31 | 4 | ND | 19 |
| feruloyl and p-cumaric acid esterases | EC 3.1.1.73 | 10, 78 | ND | 9 | ND | 4 |

* Carbohydrate-Active enzymes (CAZy); Protein data bank (PDB)

ND: Not determiner; EC: Enzyme Commission number; -: absent from database

mannan and 1,4-β-D-mannosidases (EC 3.2.1.25), which cleave manno oligosaccharides to mannose. The side groups are removed by a number of

enzymes; α-D-galactosidases stability and productivity are required [52]. Recent developments and industrial applications of a variety of biocatalysts including hydrolases, reductases, transaminase, oxidases and others, have been reviewed in detail [53]. To boost degradative processes, enzyme mixtures are of increasing importance in bio-refinery industries. Ideally, the function of each enzyme in such a mix should be known, next to its interactivity with other enzymes [54]. In particular, enzyme characteristics such as stability, degradation, tolerance to inhibitors, allosteric hindrance and mainly competition and/or collaboration between enzymes, needs to be understood. Moreover, the effects of factors like temperature and pH need attention, as these are important in enzyme reactions and may affect the nature of protein structure, such as folding and posttranslational modification (PTMs) [55]. Some commercial enzyme cocktails are already in use in lignocellulose degradation processes for the production of biofuels. At present, Celluclast, Cellic® CTec2 and Cellic® CTec3 are available as rather efficient tailor-made cellulolytic enzyme cocktails [56,57]. However, Celluclast mostly exhibits cellulase activity, and low hemicellulolytic activity. Thus hemicellulolytic enzymes added to Celluclast may promote the enzyme hydrolysis activity to obtain more reducing sugars lignocellulosic feedstocks.

**Overall aim of this thesis**

An improved understanding of how natural microbiome address the problem of degrading recalcitrant plant biomass will provide information that will ultimately permit expansion of bio-refining processes. Thus fuels, plastics and value-added chemicals can be more efficiently produced from biomass. I here propose to exploit the power of metagenomics to access the wealth of microbial genetic diversity in natural microbiomes. A thorough analysis of these will enable effective enzyme cocktails to be produced and modulated for enhanced lignocellulose biodegradation.

**Hypotheses and research questions**

1. Microbial consortia developed from soil on the proper lignocellulose substrate will harbor genes for novel carbohydrate-active enzymes. Functional metagenomics will enable to analyze the presence of glycosyl hydrolase family genes.such as (hemi)cellulases.
2. Specific novel glycosyl hydrolase family genes on fosmids can be successfully expressed, enabling the discovery of novel catalytic protein space.
3. From consortial metagenomes, predicted GH genes can be codon-optimized and expressed. Such targeted synthetic metagenomics is useful for the selection and characterization of GHs with novel properties.
4. The commercial enzyme mix Celluclast can be strongly improved in efficiency by the addition of novel enzymes produced by targeted metagenomics.

**Research questions**

- What are the key carbohydrate-active genes that play roles in lignocellulosic microbial consortia? Can functional metagenomics be used to find novel glycoside hydrolase family enzymes?
- What are the characteristics of novel identified enzymes revealed by metagenomics?
- What are the key enzymes that enhance the catalytic activity of commercial enzyme cocktails? Does plant biomass pre-treatment influence the enzyme catalytic activity? To what extent it can be useful to the bio-refinery industries?

**Outline of the thesis**

**Chapter 1** introduces the topic of this thesis. It describes the importance of microbiomes and the molecular functions they carry in plant biomass degradation. Moreover, it touches on key methods that allow to discover the novel biocatalysts beneficial to the biotechnological applications.

**Chapter 2** gives an overview of function-based metagenomics library screening for the identification of (hemi) cellulolytic enzymes from lignocellulolytic microbial consortia. In this chapter, I, together with co-authors, show the impact of using multi-substrate methodology (based on six chromogenic substrates) to enhance the metagenomics hit rate. Here, we pinpoint clones where three putative thermoalkaliphilic enzymes were covered (a xylanase, a galactosidase, and a glucosidase).

**Chapter 3** develops the data of chapter 2, with respect to the molecular characterizations of the selected genes. Here, eight glycosyl hydrolase (GHase) family candidategenes were selected using molecular cloning, gene expression and purification studies in *Escherichia coli*. Four of the enzymes showed significant activities on *pNP*-β-D-galactopyranoside, *p*NP-β-D-xylopyranoside, *p*NP-α-L-arabinopyranosideand/or *p*NP-α-D-glucopyranoside. Two of the enzymes showed thermo-tolerance and were found to encode a 'new family' of α-glucosidase activity.

**Chapter 4** enhances the strategy of targeted and synthetic metagenomics, enabling the production of glycosyl hydrolases for lignocellulose degradation. Here, we address the practical difficulties and limitations in codon optimization, gene expression and protein purification. Specifically, we characterized a novel dual-function hemicellulolytic enzyme with high furan aldehyde tolerance and optimal catalytic activity at slightly alkaline pH. The features of this enzyme will be useful in the efficient degradation of pretreated plant biomass.

**Chapter 5** describes the power of enzymatic hydrolysis in plant biomass degradation. However, practical limitations are huge while using single enzymes. In order to enhance the efficiency of hydrolysis in plant biomass pretreatment, we developed a steam explosion procedure that helps to expose the microfibrils to enzymes. Remarkably, four novel (hemi)cellulases were identified that efficiently enhance the activity of commercially available cellulase enzyme cocktails.

**Chapter 6** highlights the overall findings of this thesis, with a focus on the methodologies and novelties. Furthermore, the prospects for lignocellulolytic enzymes, enzyme cocktails, pretreatments and metagenomic library screening are examined. In addition, avenues to future work are discussed to improve the biotechnological applications.

## References


1. Cooper GM. The Origin and Evolution of Cells. Sinauer Associates; 2000.
2. Torsvik V, Øvreås L, Thingstad TF. Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science. 2002;296:1064–1066.
3. Torsvik V, Øvreås L. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol. 2002;5:240–245.
4. Ferrer M, Beloqui A, Timmis KN, Golyshin PN. Metagenomics for mining new genetic resources of microbial communities. J Mol Microbiol Biotechnol. 2009;16:109–123.
5. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. Nature. 2012;485:185–194.
6. Coughlan LM, Cotter PD, Hill C, Alvarez-Ordóñez A. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front Microbiol. 2015;6:672.
7. Timmis K, de Lorenzo V, Verstraete W, Garcia JL, Ramos JL, Santos H, *et al.* Pipelines for new chemicals : a strategy to create new value chains and stimulate innovation-based economic revival in Southern European countries. Environ Microbiol. 2014;16:9–18.
8. Reetz MT. Biocatalysis in organic chemistry and biotechnology: past, present, and future. J Am Chem Soc. 2013;135:12480–12496.
9. Swiderek K, Tuñón I, Moliner V, Bertran J. Computational strategies for the design of new enzymatic functions. Arch Biochem Biophys 2015;582:68–79.
10. Chen H, Fu X. Industrial technologies for bioethanol production from lignocellulosic biomass. Renew Sustain Energy Rev. 2016;468–478.
11. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:490-495.
12. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci. 1998;95:6578–6583.
13. Kennedy J, Marchesi JR, Dobson AD. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. Microb Cell Fact. 2008;7:27-35.
14. Vester JK, Glaring MA, Stougaard P. Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. Extremophiles. 2015;17–29.
15. Cretoiu MS, Kielak AM, Abu Al-Soud W, Sørensen SJ, Van Elsas JD. Mining of unexplored habitats for novel chitinases--chiA as a helper gene proxy in metagenomics. Appl Microbiol Biotechnol. 2012;94:1347–1358.
16. Waksman SA, Reilly HC, Harris DA. *Streptomyces griseus* (Krainsky) Waksman and Henrici. J

Bacteriol. 1948;56:259–269.
17. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, *et al.* Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol. 2002;68:4301–4306.
18. Daniel R. The metagenomics of soil. Nature Rev Microbiol. 2005;3:470–478.
19. Zhao C, Chu Y, Li Y, Yang C, Chen Y, Wang X, *et al.* High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. Biotechnol Lett. 2017;39:123–131.
20. Ekkers DM, Cretoiu MS, Kielak AM, Van Elsas JD. The great screen anomaly--a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol. 2012;93:1005–1020.
21. Villamizar GAC, Nacke H, Daniel R. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase Activity. Methods Mol Biol. 2017;1539:249–260.
22. Van Elsas JD, Costa R, Jansson J, Sjöling S, Bailey M, Nalin R, *et al.* The metagenomics of disease-suppressive soils – experiences from the METACONTROL project. Trends Biotechnol. 2008;26:591–601.
23. Simon C, Daniel R. Metagenomic analyses: Past and future trends. Appl Environ Microbiol. 2011;1153–1161.
24. Itoh N, Isotani K, Makino Y, Kato M, Kitayama K, Ishimota T. PCR-based amplification and heterologous expression of *Pseudomonas* alcohol dehydrogenase genes from the soil metagenome for biocatalysis. Enzyme Microb Technol. 2014;55:140–150.
25. Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger K-E, Streit WR. Prospecting for novel biocatalysts in a soil metagenome. Appl Environ Microbiol. 2003;69:6235–6242.
26. Gabor EM, De Vries EJ, Janssen DB. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environ Microbiol. 2004;6:948–958.
27. Mori T, Kamei I, Hirai H, Kondo R. Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. Springerplus. 2014;3:365.
28. Garmendia L, Hernandez A, Sanchez MB, Martinez JL. Metagenomics and antibiotics. Clin. Microbiol. Infect. 2012;18:27–31.
29. Pauly M, Keegstra K. Plant cell wall polymers as precursors for biofuels. Curr Opin Plant Biol. 2010;305–312.
30. De Souza WR. Microbial degradation of lignocellulosic biomass sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization. InTech; 2013.
31. Sweeney MD, Xu F. Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments. Catalysts. 2012;2:244–263.
32. Ogeda TL, Silva IB, Fidale LC, El Seoud OA, Petri DFS. Effect of cellulose physical characteristics, especially the water sorption value, on the efficiency of its hydrolysis catalyzed by free or immobilized cellulase. J Biotechnol. 2012;157:246–252.
33. Ward OP, Moo-young M, Venkat K. Enzymatic degradation of cell wall and related plant polysaccharides. Crit Rev Biotechnol. 1989;8:237–274.
34. Scheller HV, Ulvskov P. Hemicelluloses. Annu Rev Plant Biol. 2010;61:263–289.
35. Decker SR, Siika-Aho M, Viikari L. Enzymatic depolymerization of plant cell wall hemicelluloses. biomass recalcitrance. Oxford, UK: Blackwell Publishing Ltd. 2009;352–373.
36. Carpita NC. Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol. 1996;47:445–476.
37. Kanokratana P, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Eurwilaichitr L, Pootanakit K, *et al.* Phylogenetic analysis and metabolic potential of microbial communities in an industrial bagasse collection site. Microb Ecol. 2013;66:322–334.

38. McKendry P. Energy production from biomass (part 1): overview of biomass. Bioresour Technol. 2002;83:37–46.
39. Decker SR, Siika-Aho M, Viikari L. Enzymatic depolymerization of plant cell wall hemicelluloses. biomass recalcitrance. Oxford, UK: Blackwell Publishing Ltd. 2009;352–373.
40. Peng F, Ren J-L, Xu F, Bian J, Peng P, Sun R-C. Comparative study of hemicelluloses obtained by graded ethanol precipitation from sugarcane bagasse. J Agric Food Chem. 2009;57:6305–6317.
41. Kuhad R, Singh A, Eriksson K. Microorganisms and enzymes involved in the degradation of plant fiber cell walls. Adv Biochem Eng Biotechnol. 1997;57:45-125.
42. Garrote G, Domínguez H, Parajó JC. Hydrothermal processing of lignocellulosic materials. Holz als Roh- und Werkst Springer-Verlag; 1999;57:191–202.
43. Bauer A, Leonhartsberger C, Bösch P, Amon B, Friedl A, Amon T. Analysis of methane yields from energy crops and agricultural by-products and estimation of energy potential from sustainable crop rotation systems in EU-27. Clean Technol Environ Policy Springer-Verlag; 2010;12:153–161.
44. Rocha GJM, Gonçalves AR, Oliveira BR, Olivares EG, Rossell CEV. Steam explosion pretreatment reproduction and alkaline delignification reactions performed on a pilot scale with sugarcane bagasse for bioethanol production. Ind Crops Prod. 2012;35:274–279.
45. Sindhu R, Binod P, Pandey A. Biological pretreatment of lignocellulosic biomass - An overview. Bioresour Technol. 2016;199:76–82.
46. Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, *et al.* Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol. 2005;96:673–686.
47. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009;37:233–238.
48. Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni M-E, Polaina J, *et al.* Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl Environ Microbiol. 2010;76:5934–5946.
49. Väljamäe P, Kipper K, Pettersson G, Johansson G. Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics. Biotechnol Bioeng. 2003;84:254–257.
50. Beg QK, Kapoor M, Mahajan L, Hoondal GS. Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol. 2001;56:326–338.
51. Shallom D, Shoham Y. Microbial hemicellulases. Curr Opin Microbiol. 2003:219–228.
52. Vasconcellos VM, Tardioli PW, Giordano RLC, Farinas CS. Addition of metal ions to a (hemi)cellulolytic enzymatic cocktail produced in-house improves its activity, thermostability, and efficiency in the saccharification of pretreated sugarcane bagasse. Nature Biotechnol. 2016;33:331–337.
53. Huisman GW, Collier SJ. On the development of new biocatalytic processes for practical pharmaceutical synthesis. Curr Opin Chem Biol. 2013;17:284–292.
54. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels. 2010;3:22.
55. Jung S-K, Parisutham V, Jeong SH, Lee SK. Heterologous expression of plant cell wall degrading enzymes for effective production of cellulosic biofuels. J Biomed Biotechnol. 2012:1–10.
56. Haven M, Jørgensen H. Adsorption of β-glucosidases in two commercial preparations onto pretreated biomass and lignin. Biotechnol Biofuels. 2013;6:165.
57. Rodrigues AC, Haven MØ, Lindedam J, Felby C, Gama M. Celluclast and Cellic® CTec2: Saccharification/fermentation of wheat straw, solid–liquid partition and potential of enzyme recycling by alkaline washing. Enzyme Microb Technol. 2015;79–80:70–77.
58. Wiselogel A, Tyson J JD. Biomass feedstock resources and composition, in handbook on

bioethanol: production and utilization. ed. by Wyman CE. Taylor & Francis. Washington. 1996;105–118.
59. Hayn M, Steiner W, Klinger R, Steinmüller H, Sinner MEH. Basic research and pilot studies on the enzymatic conversion of lignocellulosics, in bioconversion of forest and agricultural plant residues. ed. by Saddler JN CAB Int Wallingford. 1993;33–37.
60. Sassner P, Galbe M, Zacchi G. Bioethanol production based on simultaneous saccharification and fermentation of steam-pretreated Salix at high dry-matter content. Enzyme Microb Technol. 2006;39:756–762.
61. Tengborg C, Stenberg K, Galbe M, Zacchi G, Larsson S, Palmqvist E, *et al.* Comparison of $SO_2$ and $H_2SO_4$ impregnation of softwood prior to steam pretreatment on ethanol production. Appl Biochem Biotechnol. 1998;70–72:3–15.

# Chapter 2

# A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes.

**Mukil Maruthamuthu*, Diego Javier Jiménez*, Patricia Stevens and Jan Dirk van Elsas.**

**Equal contribution*

*BMC Genomics*. 2016 Jan 28;17:86. doi: 10.1186/s12864-016-2404-0.

**Abstract**

Functional metagenomics is a promising strategy for the exploration of the biocatalytic potential of microbiomes in order to uncover novel enzymes for industrial processes (e.g. biorefining or bleaching pulp). Most current methodologies used to screen for enzymes involved in plant biomass degradation are based on the use of single substrates. Moreover, highly diverse environments are used as metagenomic sources. However, such methods suffer from low hit rates of positive clones and hence the discovery of novel enzymatic activities from metagenomes has been hampered.

Here, we constructed fosmid libraries from two wheat straw-degrading microbial consortia, denoted RWS (bred on untreated wheat straw) and TWS (bred on heat-treated wheat straw). Approximately 22,000 clones from each library were screened for (hemi)cellulose-degrading enzymes using a multi-chromogenic substrate approach. The screens yielded 71 positive clones for both libraries, giving hit rates of 1:440 and 1:1,047 for RWS and TWS, respectively. Seven clones (NT2-2, T5-5, NT18-17, T4-1, 10BT, NT18-21 and T17-2) were selected for sequence analyses. Their inserts revealed the presence of 18 genes encoding enzymes belonging to twelve different glycosyl hydrolase families (GH2, GH3, GH13, GH17, GH20, GH27, GH32, GH39, GH53, GH58, GH65 and GH109). These encompassed several carbohydrate-active gene clusters traceable mainly to *Klebsiella* related species. Detailed further functional analyses showed that clone NT2-2 (containing a beta-galactosidase of ~116 kDa) had highest enzymatic activity at 55°C and pH 9.0. Additionally, clone T5-5 (containing a beta-xylosidase of ~86 kDa) showed > 90% of enzymatic activity at 55°C and pH 10.0.

This study employed a high-throughput method for rapid screening of fosmid metagenomic libraries for (hemi)cellulose-degrading enzymes. The approach, consisting of screens on multi-substrates coupled to further analyses, revealed high hit rates, compared with recent other studies. Two clones, 10BT and T4-1, required the presence of multiple substrates for detectable activity, indicating a new avenue in library activity screening. Finally, clones NT2-2, T5-5 and NT18-17 were found to encode putative novel thermo-alkaline enzymes, which could represent a starting point for further biotechnological applications.

## Introduction

Lignocellulose constitutes an abundant organic material that is recalcitrant to degradation. Across different plant species, it contains cellulose (~35-50%) and hemicellulose (~25-35%) moieties that are complexed with lignin [1]. The cellulose moiety is a glucose polymer, whereas the hemicellulose part is composed of various pentose and hexose sugars (e.g. xylose, arabinose, mannose and galactose) linked by beta/alpha-glycosidic bonds [2-4]. All of these sugars have great value for the production of bioethanol, biodiesel and/or plastics [5, 6], and so there have been many efforts to release them from the plant matrix. However, current physicochemical methodologies for the degradation of plant biomass and subsequent production of sugars are imperfect [7] and so there is a great interest in the development of alternative and efficient processes, based on enzymes and/or lignocellulolytic microbes [8, 9].

The conversion of plant biomass to sugars requires the concerted action of different proteins, such as carbohydrate-binding modules (CBMs), polysaccharide monooxygenases, pectin lyases, hemicellulases, endoglucanases and beta-glucosidases [10-12]. Among the hemicellulases, xylosidases that can work efficiently at high temperatures in alkaline conditions are highly valued with respect to their usefulness in the pulp bleaching process [13, 14]. Actually, hemicellulases, which have previously been regarded as "accessory enzymes" of cellulases, may themselves exert vital roles in plant biomass hydrolysis [15, 16]. Given the complexity of the enzymes required for efficient lignocellulose breakdown, multi-species microbial consortia offer interesting perspectives [17-20]. To unlock the biocatalytic potential present in lignocellulolytic microbial consortia, metagenomics-based approaches have been proposed [21-23]. Two different strategies can be used: *i*) the unleashing of high-throughput DNA sequencing on degradative consortia, and/or *ii*) the selection of enzymes via functional/genetic screening of metagenomic libraries produced from these consortia [9].

Functional metagenomic screening includes the detection of "positive clones" on the basis of phenotype (e.g. enzymatic activity), heterologous complementation and modulated detection by reporter genes [24]. As such, the approach does not depend on the availability of prior sequence information to detect enzymes and it therefore offers great potential to discover genetic novelty. Using this approach, searches for (hemi)cellulases have already been made in microbiomes from decaying wood, compost, rumen and soil [25-28]. However, only few studies have explored the enzymatic potential of microbial enrichments [29-30]. It is important to notice that functional screenings come with a possible caveat, which relates to the fact that the expression conditions in the heterologous host used need to match the requirements of the insert. Due to this and other caveats (e.g. improper codon usage and/or promoter recognition, inclusion body formation, toxicity of the gene product or inability of the host to induce the gene

expression), the frequency with which positive clones are uncovered may be very low [31]. In attempts to overcome such low hit rates, some studies have applied "biased" (e.g. substrate-enriched environment) samples, coupled to the use of highly sensitive chromogenic substrates (e.g. 5-bromo-3-indolyl-beta-D-xylopyranoside) [32]. Other studies have used plasmid vectors with dual-orientation promoters to obtain more positives clones [33]. The commonly-used substrates for screening for (hemi)cellulose-degrading enzymes include azo-dyed and azurine cross-linked polysaccharides (e.g. AZCL-HE-cellulose, AZCL-xylan or AZCL-beta-glucan), para-nitrophenyl glycosides (e.g. pNP-beta-D-cellobioside, pNP-alpha-galactopyranoside or pNP-alpha-L-arabinofuranoside), carboxymethylcellulose and rimazol brilliant blue dyed-xylan. However, multiple chromogenic substrates as proxies for functional screening of (hemi)cellulases have been underexplored, although, recently, these types of approaches were catalogued as highly interesting [34].

In this study, phylogenetically stable wheat straw-degrading microbial consortia [19-35] served as the sources for two fosmid-based metagenomics libraries. These libraries were subjected to expanded functional screens by a multi-substrate approach using six chromogenic compounds (indolyl-monosaccharides). Sequencing of the inserts of seven selected positive clones indicated the presence of a suite of novel genes encoding proteins of distinct glycosyl hydrolase (GH) families, which were flanked mostly by CBMs and ABC transporters. Thus, we present an effective strategy for exploration of fosmid libraries for (hemi)cellulases, revealing hit rates higher than those reported in previous studies. Subsequent functional analyses unveiled genes encoding putative novel thermo-alkaline-tolerant enzymes, which opens the way for future biotechnological applications (e.g. biorefining or bleaching pulp).

## Materials and methods

### Metagenomic DNA extraction from lignocellulolytic microbial consortia

The lignocellulolytic microbial consortia examined here were primed with a forest soil derived microbial source. Briefly, the microbial cells were introduced into triplicate flasks containing 25 ml of mineral salt medium (MSM) with 1% of *i*) "raw" wheat straw (RWS) and *ii*) heat-treated (240 °C, 1 h) wheat straw (TWS), after which flasks were incubated at 25 °C with shaking at 100 rpm. A sequential-batch approach was followed up to the 10$^{th}$ transfer [19-35]. Then, DNA was extracted from the 10$^{th}$ transfer microbial consortia, by using the UltraClean® Microbial DNA Isolation Kit (MoBio Laboratories Inc., USA). Three replicate crude DNA extracts were pooled for each fosmid library and concentrated up to 250 ng/µl using a Speedvac concentrator. (Eppendorf, Hamburg, Germany).

### Construction of metagenomic libraries in fosmids

Metagenomic libraries were constructed using the CopyControl™ HTP Fosmid

Library Production Kit (Epicentre Biotechnologies, Madison, USA). Briefly, the metagenomic DNA was partially sheared by pipetting, to yield DNA fragments between 30 to 50 kb, after which it was 5'phosphorylated/blunt-ended. The DNA was then analyzed in 1% low-melting-point agarose using a CHIEF-DR III pulsed field gel electrophoresis system (BioRad, Hercules, USA) at 14ºC with the following parameters: gradient 6 V/cm, included angle 120º, initial switch time 0.5 sec, final switch time 8.5 sec, linear ramping factor, 18 h. DNA fragments of approximately 30-40 Kb were excised from the gel and recovered using Zymoclean™ Large Fragment DNA Recovery Kit (Zymo Research, Irvine, USA). The DNA was then ligated into vector pCC2FOS, packaged in phage particles and competent *E. coli* EPI300-T1$^R$ cells were transformed with it. The *E. coli* cells were diluted 1:$10^3$ and plated onto 1% LB agar supplemented with 12.5 μg/ml chloramphenicol (LBA+Cm). Plates were incubated overnight at 37ºC, to produce 500 to 600 colonies per plate. The colonies of each plate were pooled in 1 ml of LB broth with 20% of glycerol and stored as fosmid pools at -80°C for further analysis.

**Screening for fosmid clones expressing (hemi)cellulolityc enzymes**

Screening was done in three steps. First, fosmid pools stored at -80ºC were recovered in 100 μl LB broth at 37ºC for 1 h (shaking at 250 rpm) and serially diluted up to 1:$10^5$. Then, each suspension (100 μl) was plated on LBA+Cm supplemented with a mix of each of the six chromogenic substrates (at 40 μg/ml) (Table 1). After incubation (48 h, 37ºC), dark blue colonies (due to hydrolysis of the chromogenic substrate) were selected, purified to obtain single colonies and retested. Secondly, selected clones were plated onto LBA+Cm supplemented with each of the specific hemicellulose-mimicking substrates, i.e. X-Fuc, X-Gal, X-Man and X-Xyl, in single, double, triple and quadruple combinations. Thirdly, clones that were positive in the first screening (on six substrates) and negative in the second screening (four substrates in different combinations) were further tested on X-Cel and X-Glu (single and double combinations) (Figure 1).

**Extraction of DNA from selected fosmid clones**

Selected positive clones were cultured in 4 ml of LB supplemented with 12.5 μl/ml chloramphenicol (LB+Cm) and incubated at 250 rpm for 8 h at 37 ºC. After incubation, 25 μl was used to inoculate 25 ml of fresh LB+Cm. To increase the fosmid copy numbers, 50 μl of autoinduction solution (500X) (Epicentre Biotechnologies, Madison, USA) were added and flasks incubated (37 ºC, shaking at 250 rpm). At $OD_{600}$ of about 2-2.5, fosmid DNA was extracted from these cultures using the Gene Jet Plasmid Midi Preparation Kit (Thermo Scientific, Waltham, USA). DNA size and integrity were verified by running aliquots of the DNA on 1% agarose gels and DNA concentration was measured by spectrophotometry (Nanodrop 2000; Thermo Scientific). The resulting fosmid

DNA was digested with *Eco*R1 and the restriction patterns was analyzed on 0.8% agarose gel. Band sizes were estimated by comparison to a standard DNA marker (GeneRuler™ 1 Kb ladder, Thermo Scientific). The size of the insert of each fosmid was estimated by calculating the sum of the sizes of the individual *Eco*R1 generated bands minus 8,181 bp (fosmid backbone).

**Sequencing and gene annotation of fosmid inserts**

The selected positive fosmid clones were sequenced using the Illumina HiSeq platform (2 X 100bp). Sequences from the *E. coli* EPI300 genome as well as vector backbone sequences were removed. The resulting set of raw sequence data was quality-checked and further processed, with normalization and Velvet-based *de novo* assembly to generate contigs (Beckman Coulter Genomics, Danvers, USA). Contigs were selected from the data if the average coverage exceeded 200-fold, and final contigs were considered to be representative of the whole insert. ORFs were assigned to each of the contigs using the Rapid Annotation Subsystems Technology (RAST) server [46]. Subsequently, the ORFs were annotated in the CAZymes Analysis Toolkit (CAT) platform [47], using default parameters. Finally, all genes predicted in each fosmid insert were re-annotated and verified (in-house) using BLASTX searches against the NCBI database. The results from this analysis were loaded into MEGAN v5 software [48], after which they were classified taxonomically using the suggested parameters for the LCA algorithm (maximum number of matches per read 10; minimal support 5; minimal score 35; max expected 0.01; minimal complexity 0.3; and top percent 10). The nucleotide sequences of the contigs retrieved from the clones NT2-2, T5-5, NT18-17, T4-1, 10BT, NT18-21 and T17-2 were deposited in the GenBank database under the accession numbers KU505133-KU505147.

**Functional analyses–beta-galactosidase, beta-xylanase and alpha-glucosidase activity assays**

Positives clones were grown (37°C, shaking at 250 rpm) in 5 ml LB+Cm containing 10 µl of auto-induction solution (500X). At $OD_{600}$ of 0.5, cells were harvested by centrifugation (10 min, 10,000 g). Proteins were extracted by adding 2 ml of lysis buffer (20 mM Tris-HCL pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.1% Triton, 5 mM CHAPS and a tablet of protease inhibitor -Roche- to 50 ml) to the pellet. Subsequently, the mixtures were sonicated on ice (6 sec on, 15 sec off, 30 cycles with amplitude of 10-15 microns). Protein concentration was determined by the Bradford method using bovine serum albumin as standard. The total protein fractions were recovered and tested for activity using pNPGal, pNPXyl and pNPGlu. The reaction mixture consisted of 0.3 ml of 10 mM pNPGal, pNPXyl or pNPGlu (diluted in 50mM of Tris-HCL pH 7.0) and 0.3 mL of each clone protein fraction. The mixtures were incubated at 40°C for 30 min to 2 h (depending on the quantity of proteins and activity) and the reactions were stopped on ice. Two

negative controls were used for all assays: *i*) reaction mixture without substrate and *ii*) reaction mixture using the total protein fraction from the fosmid-less host *E. coli* EPI300. Enzymatic activities were determined from the measured absorbance units using a standard calibration curve. The amount of para-nitrophenyl (pNP) liberated was measured by absorbance at 410 nm. One unit (U) of enzyme activity was defined as the activity required for the formation of 1 μM of pNP per min under the above conditions (in this case mg of total protein). Optimum temperature was determined in the range of 4–80 °C using pNPGal, pNPXyl and pNPGlu (at pH 7.0 and 9.0). The pH optimum was determined in a pH range from 3.0 to 10.0 (at 40°C and 55°C) under standard conditions using the following buffers: 50 mM sodium citrate (pH 3.0 to 6.0), 50 mM Tris–HCl (pH 7.0–9.0) and 50 mM glycine-NaOH (pH 10.0).

**Protein electrophoresis and zymographic analyses**

Zymograms were used to detect beta-galactosidase and beta-xylosidase activities on native polyacrylamide gels (4% stacking, 10% resolving gels) using 40 μg of total protein per sample. After running the gels at 4°C, they were washed with water and then incubated with 5 mM of each substrate (MUFGal and MUFXyl) diluted in 0.1 M of Tris-HCl (pH 8.0) at 25°C for 1 h. Following this, bands were visualized under UV light.

## Results

**Construction and functional screening of two fosmid metagenomic libraries**

Two metagenomic libraries were produced in fosmids, one from pooled raw wheat straw (RWS) consortial DNA (~70,000 clones) and another one from torrified wheat straw (TWS; ~70,000 clones). Each library contained clones with inserts of ~35 kb average sizes, yielding approximately 2.4 Gb of total cloned genomic DNA per library. In order to screen for (hemi)cellulose-degrading enzymes, about 22,000 clones per library were subjected to activity screens on LB agar supplemented with mixtures of six chromogenic substrates (Table 1; Figure 1). These screens yielded a total of 71 positive hits, being 50 from RWS and 21 from TWS. This corresponded to, respectively, 1 hit per 440 screened clones (RWS), and 1 hit per 1,047 screened clones (TWS).

The 71 positive fosmid clones were restreaked to purity and then retested for activity to confirm the initial screening result. They were then subjected to additional plate screens with the hemicellulose-mimicking substrates X-Fuc, X-Gal, X-Xyl and X-Man, including single-, double- and mixed-substrate plates. The 71 positive clones showed activity in the latter assay (mixed-substrate plate), confirming the initial data. Clones that showed consistent activities on the mixtures of six substrates but no activity on the aforementioned hemicellulose-mimicking substrates were then tested for their abilities to degrade cellulose or starch using X-Cel and X-Glu (single and double) (Figure 1). After removal of

clones with questionable activity (e.g. faint-blue colonies with white background), the remaining 52 clones showed consistent activities on at least one of the substrates tested. Specifically, 20 clones showed activity on X-Gal (18 from RWS - hit rate 1:1,157; 2 from TWS - hit rate 1:11,000), 9 on X-Xyl (all from TWS – hit rate 1:2,444) and 15 on X-Glu (all from RWS – hit rate 1:1,466) (as singletons). Remarkably, eight were positive only on the mixture of six substrates.

**Table 1.** Chromogenic substrates used in this study

| Activity on * | Type of enzymes detected | Substrate (indolyl-monosaccharide) | Abbreviation | Concentration into the plate | Supplier |
|---|---|---|---|---|---|
| Hemicellulose | alpha-fucosidases | 5-bromo-4-chloro-3-indolyl-α-L-fucopyranoside | X-Fuc | Each 40 µg/ml | Biosynth AG, Switzerland |
| | beta-galactosidases | 5-bromo-4-chloro-indolyl-β-D-galactopyranoside | X-Gal | | |
| | beta-xylosidases | 5-bromo-4-chloro-3-indolyl-β-D-xylopyranoside | X-Xyl | | |
| | alpha-mannosidases | 5-bromo-4-chloro-3-indolyl α-D-mannopyranoside | X-Man | | |
| Cellulose and Starch | beta-glucanases | 5-bromo-4-chloro-3-indolyl-β-D-cellobioside | X-Cel | | |
| | alpha-glucosidases | 5-bromo-4-chloro-3-indolyl-α-D-glucopyranoside | X-Glu | | |

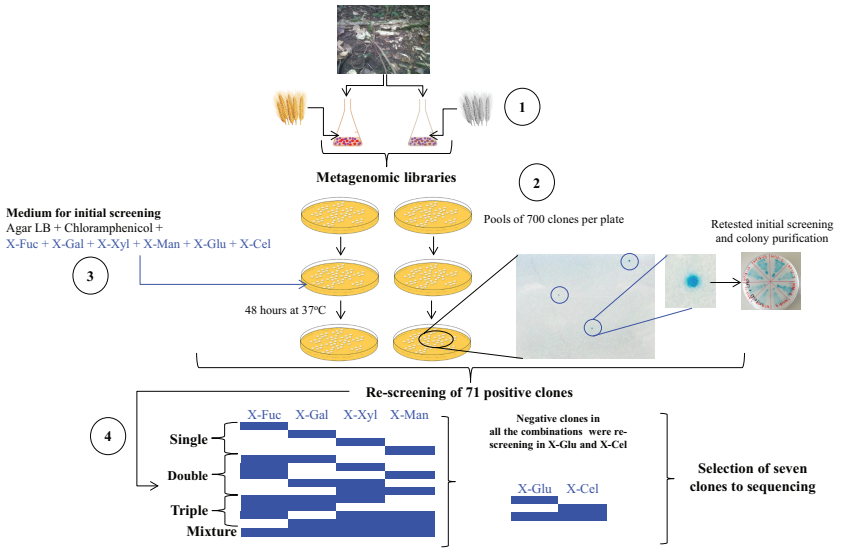



**Figure 1**. Schematic representation of the methodology used to screen for (hemi) cellulases in fosmid-based metagenomic libraries. 1) Biased communities (e.g. soil-derived microbial consortia bred on wheat straw) are at the basis of enhanced hit rates. 2) Screening of fosmid pools (700 per pool) allow high-throughput analyses. 3) Screening on substrate mixtures accelerate the analyses. 4) Re-screening of positive clones in single, double and further mixed combinations of substrates enable the detection of specific activities.

For each fosmid clone type in our library, as determined by activity, 2-4 clones were selected for genetic analysis. This selection was also guided by the

clones, i.e. RWS and TWS. Restriction with *Eco*R1 revealed, for all tested fosmids, the presence of insert sizes of approximately 28 to 35 kb. It also allowed the detection of duplicates for dereplication of the fosmid set. The, thus selected, final set of (seven) fosmids consisted of two clones that were positive for X-Gal (T17-2 and NT2-2), one for X-Xyl (T5-5), two for X-Glu (NT18-17 and NT18-21), and two with activity on multiple mixed substrates (10BT and T4-1) (Table 2).

**Analysis of fosmid insert sequences and detection of carbohydrate-active enzymes**

The seven selected clones (NT2-2, T5-5, NT18-17, T4-1, 10BT, NT18-21 and T17-2), were subjected to full insert sequencing. Final assembly of the inserts revealed a total of 15 contigs, of sizes between 3.0 and 35 kb. Thus, some inserts had more than one contig, indicating the existence of regions with too low coverage. Per clone, the contigs were considered to be of sufficient representation to allow further analyses (Figure 2). All contigs were then screened for the presence of open reading frames (ORFs) on the basis of the presence of start and stop codons (automatic annotation from the RAST server, followed by manual validation). In addition, the identified genes were, gene-by-gene, subjected to BLAST-based analyses, comparing against the NCBI and carbohydrate-active enzyme (CAZy) databases. Overall, we detected 18 promising ORFs encoding proteins from 12 different GH families amongst a total of 211 ORFs. The G+C contents of the inserts ranged from 54.6% to 63.5%. The complete annotation of each of the seven fosmid inserts is presented in the supplementary files. A brief description of each insert is listed below (Table 2 and Table 3).

Table 2. Selected fosmids to insert sequencing and annotation of genes detected in each metagenomics fragment

| | | | | Number of genes based on RAST platform and CAZy database annotation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Fosmid ID** | **Positive on** | **# Contigs (total length)** | **% GC** | **GHs** | **CBMs** | **AAs** | **GTs** | **CEs** | **ABC** | **H/U** | **Others** | **Total** |
| NT2-2 | X-Gal | 2 (31.21 Kb) | 60.1 | 2 | 2 | 1 | 2 | 0 | 3 | 2 | 14 | 26 |
| T5-5 | X-Xyl | 5 (31.63 Kb) | 51.9 | 4 | 2 | 0 | 0 | 0 | 3 | 8 | 13 | 30 |
| NT18-17 | X-Glu | 2 (33.7 Kb) | 63.5 | 4 | 0 | 1 | 1 | 1 | 10 | 4 | 14 | 35 |
| T4-1 | Mixed* | 2 (34.84 Kb) | 57.2 | 2 | 3 | 1 | 0 | 0 | 3 | 5 | 17 | 31 |
| 10BT | Mixed* | 2 (31.87 Kb) | 54.6 | 2 | 0 | 2 | 1 | 0 | 1 | 12 | 25 | 43 |
| NT18-21 | X-Glu | 1 (29.89 Kb) | 55.0 | 0 | 1 | 0 | 0 | 0 | 3 | 2 | 21 | 27 |
| T17-2 | X-Gal | 1 (23.57 Kb) | 54.8 | 4 | 1 | 0 | 2 | 0 | 0 | 3 | 9 | 19 |

Asterisk: positives on the mixture of six chromogenic substrates.
Glycosyl hydrolases (GHs); Carbohydrate binding modules (CBMs); Auxiliary activities (AAs); Glycosyl transferases (GT); predicted ABC transporters (ABC); Hypothetical and unknown genes (H/U).

**Fosmid NT2-2**- Two contigs represented the total 31.2 kb insert, encompassing 26 predicted ORFs. The sizes of the identified ORFs ranged from 123 to 3,309 bp. Gene annotation revealed the presence of two predicted genes encoding proteins of GH families GH2 and GH53. The GH2-encoding gene was annotated as a beta-galactosidase (~116 kDa) and could be correlated with the activity on X-gal. Flanking this gene, genes predicted to encode two CBMs (CBM48 and CBM26), two glycosyl transferases (GTs) (GT2 and GT90) and an operon containing three

methionine ABC transporter genes were found (Additional file 1).

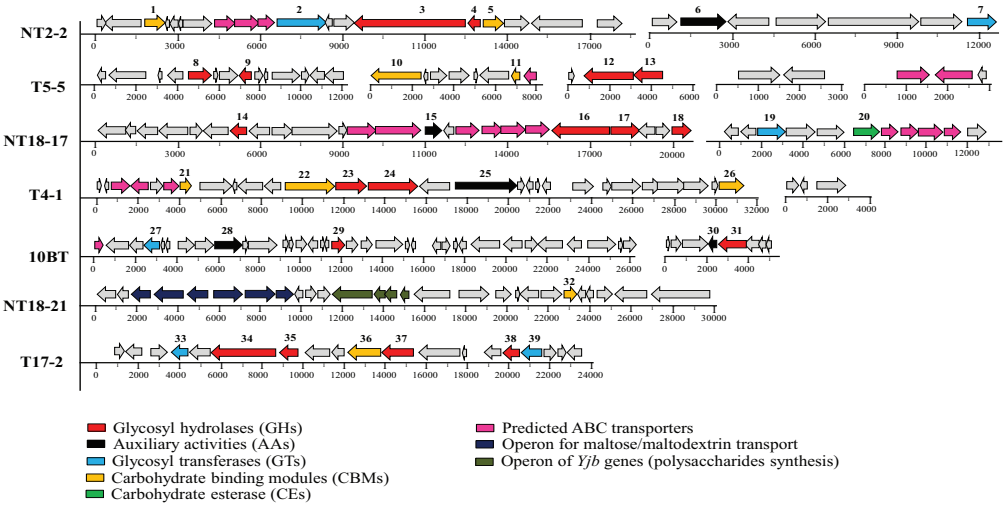



**Figure 2.** Graphical representation of the genes detected and annotated in seven fosmid inserts selected by the multi-substrate screening approach.

**Fosmid T5-5**- Thirty predicted ORFs, with sizes ranging from 138 to 2,379 bp, were present in the 31.6 kb insert (composed of five contigs). A gene predicted to encode a 789-amino acid protein (~86 kDa) was annotated as a gene for beta-xylosidase. This protein could be involved in the detected enzymatic activity on X-Xyl. Flanking this gene, three ABC transporters, a xyloside transporter (*Xyn*T – annotated as GH17 by CAZy) and a CBM50 gene (annotated as a lipoprotein by RAST) were found. In addition, predicted genes for carbohydrate-active enzymes of families CBM20, GH13 and GH32 were detected (Additional file 2)

**Fosmid NT18-17**- Thirty-five ORFs were identified in the 33.7 kb insert (two contigs). The insert showed a high G+C content, i.e. 63.5%. Four GH-encoding genes were detected, which fell in the GH27, GH20, GH58 and GH109 families. Of these, two predicted proteins might be linked to the activity on X-Glu (GH58-hypothetical protein or GH27-aquaporin). In addition, ten predicted ABC transporter genes were detected (Additional file 3).

**Fosmid T4-1**- This contiguous insert (31.4 kb) consisted of 31 ORFs, of sizes between 189 and 3,003 bp. Two genes with predicted GH activity were found. These might be involved in the enzymatic activities detected on mixed substrates, i.e. a GH3 family gene (encoding a beta-xylosidase) and a GH17 family one (annotated as a xyloside transporter; *Xyn*T). In addition, three genes predicted to encode CBMs were found in this insert (two CBM50 and one CBM20). The remainder encompassed either hypothetical and/or uncharacterized genes (five ORFs) or genes encoding different functions (seventeen ORFs) (Additional file 4).

**Fosmid 10BT**- Fourty-three ORFs were identified in the 31.8 kb insert of fosmid 10BT (two contigs). Consistent with the annotation, genes predicted to encode two GHs (i.e. GH39 and GH53), two auxiliary activities (AAs; AA5 and AA3), one GT (GT4) and one ABC transporter were identified. Of these, the newly discovered GH39 (beta-xylosidase) and GH53 (endo-beta-1,4-galactanase) genes could be related to the activities measured with the mixture of substrates. Interestingly, these genes were annotated -by RAST- as transcriptional regulator (AraC family) and an inner membrane protein, *Yft*N, respectively (Additional file 5).

**Fosmid NT18-21**- Fosmid NT18-21 contained a 29.8 kb insert within a single contig. Although this insert contained a total of 27 predicted ORFs, CAZy annotation predicted only one gene with carbohydrate activity, i.e. one encoding CBM50 (identified by RAST as a "shikimate 5-dehydrogenase I gamma"). Furthermore, two operons were detected that might relate to the activity, one of them encompassing predicted genes for maltose/maltodextrin transporters and the second one presumed polysaccharide synthesis genes (*Yjb*H-*Yjb*G-*Yjb*F-*Yjb*E). The latter operon was flanked by a gene for glucose-6-phosphate isomerase and one for an aspartokinase, which are both involved in sugar metabolism (Additional file 6).

**Fosmid T17-2**- Nineteen ORFs were identified within the 23.5 kb insert of fosmid T17-2 (one contig), which had a G+C content of 54.8%. The ORFs had a size range between 117 and 3,084 bp. One ORF encoding a 1,027-amino acid protein was identified as a gene for beta-galactosidase (GH2 family), suggesting it was responsible for the activity of the fosmid on X-gal. In addition, we identified genes for the transcriptional repressor of the lac operon and "PTS system sucrose-specific IIB component" that were identified by CAZy as a GH53 and a GH32 family proteins, respectively. Two genes encoding GTs (GT4 and GT8), and one gene for CBM51 were also identified (Additional file 7).

**Tracking the microbial sources of the fosmid inserts**

To identify the potential microbial source of each metagenomic insert, the predicted amino acid sequences per gene per contig were BLAST-compared to the NCBI database. In addition, such BLAST results were analyzed by the Lowest Common Ancestor (LCA) algorithm in MEGAN v5 (Additional files). Thus 14 predicted protein sequences from the NT2-2 insert were affiliated to proteins of members of the Enterobacteriaceae, notably *Klebsiella oxytoca* and *Enterobacter* sp. However, another 11 predicted proteins from this insert were affiliated, based on the 50 "best" BLAST hits, to those from *Pseudomonas putida*. In the fosmid T5-5 insert, 27 predicted proteins were mainly related to *Klebsiella oxytoca* –derived proteins. The insert of fosmid NT18-17 showed a complexity of genes that were affiliated to different genera (e.g. *Pelagibacterium*, *Rhizobium* and *Mesorhizobium*).

**Table 3. Carbohydrate-active genes detected in each fosmid insert**

| Fosmid ID | ID gene Figure 2 | CAZy Family [a] | Size in amino acids (kDa [b]) | Annotation based on RAST platform (EC number) | Probably protein from (identity / coverage ) |
|---|---|---|---|---|---|
| NT2-2 | 1 | CBM26 | 236 | GCN5-related N-acetyltransferase | *Klebsiella oxytoca* (73% / 83% ) |
| | 2 | GT2 | 521 | Uncharacterized protein ylaB | *Enterobacter mori* (58% / 74% ) |
| | 3* | GH2 | 1028 (116) | Beta-galactosidase (EC 3.2.1.23) | *Enterobacter hormaechei* (74% / 81%) |
| | 4 | GH53 | 119 | Transcriptional repressor of the lac operon | *Enterobacter hormaechei* (59% / 75%) |
| | 5 | CBM48 | 301 | Carboxyl-terminal protease (EC 3.4.21.102) | *Pseudomonas putida* (99% /100%) |
| | 6 | AA3 | 565 | Choline dehydrogenase (EC 1.1.99.1) | *Pseudomonas putida* (99% / 99%) |
| | 7 | GT90 | 144 | Thioredoxin | *Pseudomonas putida* (95% / 99%) |
| T5-5 | 8 | GH32 | 310 | 6-phosphofructokinase class II (EC 2.7.1.11) | *Klebsiella oxytoca* (92% / 97%) |
| | 9 | GH13 | 241 | Ferric siderophore transport system, periplasmic binding protein TonB | *Klebsiella pneumonia* (75% / 84%) |
| | 10 | CBM20 | 792 | Phosphoenolpyruvate synthase (EC 2.7.9.2) | *Klebsiella oxytoca* (98% / 99%) |
| | 11 | CBM50 | 154 | Probable lipoprotein nlpC precursor | *Klebsiella oxytoca* (98% /99%) |
| | 12* | GH3 | 789 (86) | Beta-xylosidase (EC 3.2.1.37) | *Enterobacter mori* (84% /91%) |
| | 13 | GH17 | 465 | Xyloside transporter XynT | *Enterobacter cloacae* (82% / 90%) |
| NT18-17 | 14* | GH27 | 218 (22) | Aquaporin Z | *Hyphomonas neptunium* (65% / 80%) |
| | 15 | AA6 | 195 | NAD(P)H oxidoreductase | *Rhizobium etli* (48% / 64%) |
| | 16 | GH20 | 642 | Beta-hexosaminidase (EC 3.2.1.52) | *Rhizobium leguminosarum* 3841 (50% / 64%) |
| | 17* | GH58 | 356 (38) | hypothetical protein | *Hyphomicrobium denitrificans* (56% / 71%) |
| | 18 | GH109 | 231 | Dehydrogenase | *Rhizobium leguminosarum* (58% / 76%) |
| | 19 | GT2 | 442 | Omega-amino acid-pyruvate aminotransferase (EC 2.6.1.18) | *Mesorhizobium opportunistum* WSM207 (82% / 90%) |
| | 20 | CE9 | 483 | Dihydropyrimidinase (EC 3.5.2.2) | *Rhizobium leguminosarum* (80% / 90%) |
| T4-1 | 21 | CBM50 | 154 | Probable lipoprotein nlpC precursor | *Klebsiella oxytoca* (92% / 99%) |
| | 22 | CBM20 | 792 | Phosphoenolpyruvate synthase (EC 2.7.9.2) | *Klebsiella oxytoca* (98% / 99%) |
| | 23* | GH17 | 465 (52) | Xyloside transporter XynT | *Enterobacter cloacae* (82% / 90%) |
| | 24* | GH3 | 789 (86) | Beta-xylosidase (EC 3.2.1.37) | *Enterobacter mori* (83% / 90%) |
| | 25 | AA4/AA7 | 1000 | Glycolate dehydrogenase (EC 1.1.99.14), subunit GlcD | *Klebsiella oxytoca* (95% / 98%) |
| | 26 | CBM50 | 332 | L,D-transpeptidase YnhG | *Klebsiella oxytoca* (85% / 93%) |
| 10BT | 27 | GT4 | 239 | Thioesterase involved in non-ribosomal peptide biosynthesis | *Pseudomonas putida* (90% / 93%) |
| | 28 | AA3 | 413 | Sarcosine oxidase beta subunit (EC 1.5.3.1) | *Pseudomonas putida* (98% / 99%) |
| | 29* | GH39 | 192 (21) | Transcriptional regulator, AraC family | *Klebsiella oxytoca* (76% / 87%) |
| | 30 | AA5 | 127 | Integral membrane protein YfiB | *Enterobacter cancerogenus* (90% / 94%) |
| | 31* | GH53 | 406 (46) | Inner membrane protein YfiN | *Enterobacter cloacae* (79% / 88%) |
| NT18-21 | 32 | CBM50 | 396 | Shikimate 5-dehydrogenase I gamma (EC 1.1.1.25) | *Klebsiella oxytoca* (83% / 90%) |
| T17-2 | 33 | GT8 | 207 | Galactoside O-acetyltransferase (EC 2.3.1.18) | *Escherichia coli* (81% / 93%) |
| | 34* | GH2 | 1027 (116) | Beta-galactosidase (EC 3.2.1.23) | *Citrobacter freundii* (87% / 92%) |
| | 35 | GH53 | 360 | Transcriptional repressor of the lac operon | *Citrobacter koseri* (89% / 94%) |
| | 36 | CBM51 | 581 | Choline-sulfatase (EC 3.1.6.6) | *Klebsiella oxytoca* (94% / 97%) |
| | 37 | GH32 | 471 | PTS system, sucrose-specific IIB component | *Klebsiella oxytoca* (94% / 97%) |
| | 38 | GH65/GT5 | 266 | Cof, detected in genetic screen for thiamin metabolic genes | *Klebsiella oxytoca* (97% / 98%) |
| | 39 | GT4 | 330 | Lysophospholipase L2 (EC 3.1.1.5) | *Klebsiella oxytoca* (94% / 97%) |

Asterisks correspond to genes predicted to be involved in the detected enzymatic activities

[a] Annotation using the CAZymes Analysis Toolkit (CAT) platform. [b] Predictive molecular size in kDa

These genera all belong to the Rhizobiales, suggesting an organism from this group as the most likely source. In both fosmids T4-1 and NT18-21, virtually all predicted proteins (approximately 96%) were affiliated with proteins from members of the Enterobacteriaceae. Closer (manual) screening of the data indicated that insert T4-1 might come from a *Klebsiella oxytoca* -like organism, whereas insert NT18-21 might originate from an organism affiliated with either *Citrobacter/Klebsiella/Salmonella.* A similar observation was made for the fosmid T17-2 insert. In the case of fosmid 10BT, eleven ORFs yielded predicted proteins that resembled those of *Pseudomonas putida* –like organisms (coverage and identity of > 90%), whereas the remainder of the predicted proteins were more related to those from enteric species (e.g. mostly *Klebsiella oxytoca* –like). This was similar to what was shown for the NT2-2 insert.

**Functional analyses: beta-galactosidase, beta-xylanase and alpha-glucosidase activities**

Based on the initially-detected activities of the fosmid clones, we selected three commercially available substrates, i.e. para-nitrophenyl-beta-D-galactopyranoside (pNPGal), para-nitrophenyl-beta-D -xylanopyranoside (pNPXyl) and para-nitrophenyl-alpha-D-glucopyranoside (pNPGlu), in order to quantify the activities (using total protein extracts) at different temperatures and pH values. Clones NT2-2 and T17-2 were positive on pNPGal, confirming the initial screening data, while clones T5-5, T4-1 and 10BT were positive on pNPXyl. In addition, clone NT18-17 showed activity on pNPGlu (Figure 3a). Clones NT2-2 and T5-5 showed elevated levels of enzymatic activity and were therefore chosen for further assays (Figure 3b,d). Total protein extracts produced from the fosmid-less host (*E. coli* EPI 300) did not show any activity on the selected pNP substrates, confirming that the activities came from the metagenomic inserts. For clone NT2-2, activity on pNPGal was maximal at 55°C and pH 9.0 (between 3,516 U/mg ± 219.02 and 3,377 U/mg ± 47.19), with 34% and 44% of activity remaining at 80°C and pH 10.0, respectively. Activity was not detected below pH 6.0 (Figure 4a,b). The activity of clone T5-5 on pNPXyl was maximal at 55°C, with values of around 93, 21 and 12% of this maximum at 37, 70 and 22°C, respectively. The highest xylanolytic activity for clone T5-5 was obtained at 55°C and pH 8.0 (3.96 U/mg ± 0.04), with ~91% of activity remaining at pH 10.0. The enzyme produced by T5-5 was likely alkaliphilic. Similar to clone NT2-2, T5-5 activity was not detected at pH values below 6.0 (Figure 4c,d). Finally, clone NT18-17 showed maximum activity at pH 7.0 and 55°C (1.87 U/mg ± 0.06), and about 0.94 U/mg at 40°C and pH 10.0. This suggested this enzyme was quite tolerant at alkaline conditions (Figure 4e,f).

**Zymograms**

Native polyacrylamide gels showed that crude protein extracts from the seven fosmid clones had band patterns different from those of the *E. coli* EPI300 host.

Moreover, none of the bands produced from the *E. coli* EPI300 host were positive with MUFGal (MUF-beta-D-galactopyranoside) and MUFXyl (MUF-beta-D-

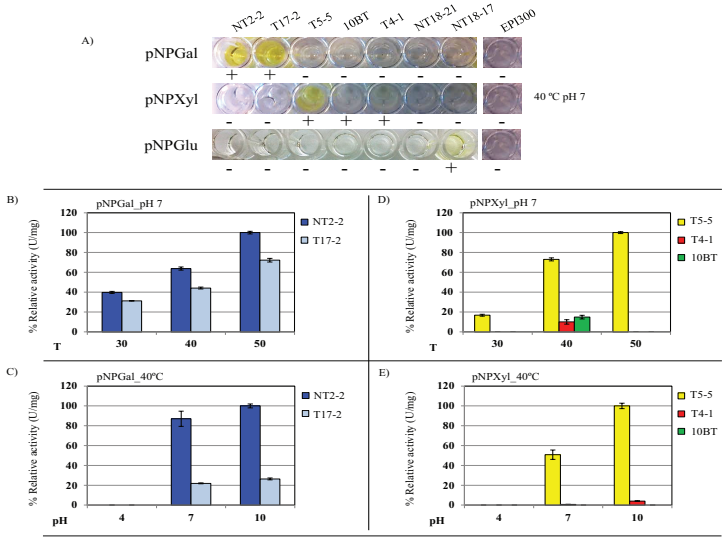



**Figure 3.** Functional analyses (relative enzymatic activities, expressed in U/mg) based on total proteins extracted from selected positive fosmid clones. A) Characterization of the fosmid clones using three pNP-substrates (pNPGal, pNPXyl and pNPGlu). B and C) Effect of temperature and pH values on the activity of the clones NT2-2 and T17-2 with pNPGal. D and E) Effect of temperatures and pH values on the activity of the clones T5-5, T4-1 and 10BT with pNPXyl

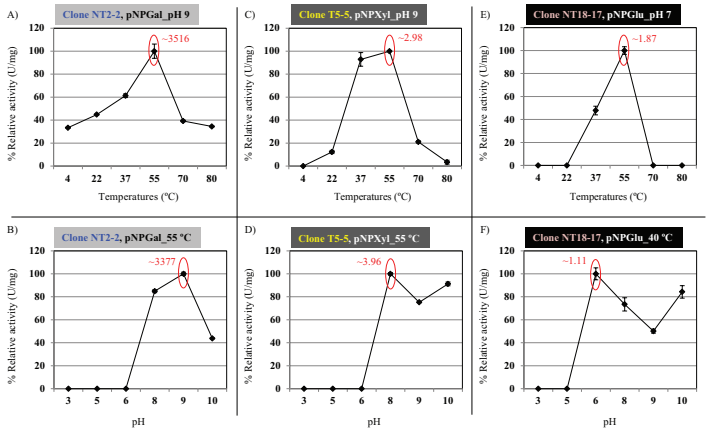



**Figure 4.** Relative enzymatic activities of the total proteins from the fosmids clones NT2-2, T5-5 and NT18-17 using pNPGal, pNPXyl and pNPGlu, respectively, at different temperatures and pH values. A and B) Effect of temperatures and pH values on the activity of the fosmid clone NT2-2. C and D) Effect of temperatures and pH values on the activity of the fosmid clone T5-5. E and F) Effect of temperatures and pH values on the activity of the fosmid clone NT18-17.

xylopyranoside) (used as substrates), confirming that any activities measured came from the metagenomic inserts. Clones N2T-2 and T17-2 both showed a band of

>100 kDa with high activity on MUFGal (Figure 5). Given their estimated sizes, these bands likely represented proteins encoded by genes 3 and 34 (Figure 2; both beta-galactosidase encoding genes). Clone T4-1 showed a band of 75-100 kDa size, with xylosidase activity, which is consistent with the initial finding of activity on pNPXyl. This band likely corresponded to a protein encoded by gene 24 (Table 3), predicted to be a beta-xylosidase of ~86 kDa. Clones T5-5 and 10BT, positive with pNPXyl, did not show any bands with activity on zymograms using MUFXyl.

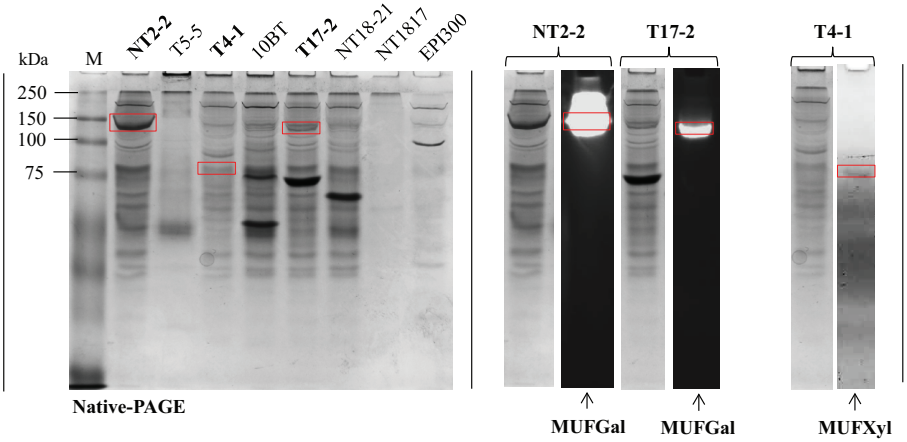



**Figure 5.** Native polyacrylamide gel electrophoresis (NATIVE-PAGE) and zymogram analysis. Left: Native polyacrylamide gel of the total proteins from the positive fosmid clones. Right: Positives fosmids clones that showed bands (red square) with activity on the zymogram assay using MUFGal (MUF-beta-D-galactopyranoside) and MUFXyl (MUF-beta-D-xylopyranoside) as substrates.

## Discussion

In this study, two wheat straw-degrading microbial consortia, RWS and TWS, were successfully subjected to metagenomics library constructions using fosmids, yielding 2.4 Gb of genomic information per library. Taking an estimated average bacterial genome size in our microbial consortia of 4 Mb and considering these were strongly dominated by bacteria, we thus cloned the equivalent of roughly 600 bacterial genomes. Previous data on the two consortia [35] revealed the presence of ~100 (RWS) and ~50 (TWS) dominated bacterial types, in relative abundances within one log unit, giving a coverage of around six-fold for RWS and twelve-fold for TWS. Hence, a back-of-the-envelope calculation revealed us that we basically covered the genes from most of the dominant bacterial members in the degrader consortia.

To detect (hemi)cellulolytic enzymes by functional screenings, two alternative strategies can be followed: 1) high-throughput detection in agar plates (mostly secreted enzymes) using hydrolysis of a chromogenic substrate as the criterion and 2) detection of enzyme activity in crude extracts after cell lysis. In either methodology, additional factors should be taken into account. These are the

vector copy number, the need for induction of gene expression, secretion of the enzyme and recovery of the vector plasmid after expression [31, 36]. Here, we tested our fosmids by functional screenings for (hemi)cellulases initially using a mixture of (six) chromogenic compounds in agar plates. These substrates (indolyl-monosaccharides) can be internalized by *E. coli* and are thus readily available for hydrolysis by intracellularly-expressed exo-glycosidases. This is not the case for substrates such as oat spelt xylan or CMC, the hydrolysis of which relies on the release of fosmid-expressed enzymes, which probably only occurs after cell death and lysis [32]. The substrates were organic compounds, each consisting of a monosaccharide linked to a substituted indole moiety. The substrates yield insoluble blue compounds as a result of enzyme-catalyzed hydrolysis. For example, X-Xyl, when cleaved by beta-xylosidase, yields xylose and 5-bromo-4-chloro-3-hydroxyindole. The latter compound can spontaneously dimerize and is oxidized into 5,5'-dibromo-4,4'-dichloro-indigo, an intensely blue product which is insoluble. Taking into account the structure of these substrates, we hypothesized that the approach allows the screening for debranching enzymes that act in the external chains of sugars (in this case fucose, xylose, galactose, mannose and glucose) and that are linked to the backbone of the (hemi)cellulose structures. Indeed, our multi-substrate approach is also applicable in screens of metagenomic libraries for other classes of enzymes, as long as chromogenic substrates are available for that purpose. For example, to detect lipolytic activity, 5-(4-hydroxy-3,5-dimethoxyphenylmethylene)-2-thioxothia-zolidin-4-one-3-ethanoic acid (SRA)-propionate, SRA-butyrate, SRA-octanoate, SRA-decanoate, SRA-laurate and SRA-myristate can be employed [37].

Given our high hit rates, i.e.1:440 in RWS and 1:1,047 in TWS, the multi-substrate screening approach was superior to approaches reported in the recent literature (Table 4). On the other hand, in both libraries the hit rates for the individual enzymes were <1:2,500, except for X-Gal in RWS (1:11,000). Comparing 15 different metagenomic libraries and 19 single substrates, we inferred an average hit rate of (hemi)cellulolytic activities of ~ 1:7,300. However, some approaches showed hit rates of < 1:2,000. Additionally, low hit found (in recent studies) using chromogenic substrates (e.g. X-Xyl) or azurine cross-linked polysaccharides. Interestingly, Zhao *et al.* (2010) [38], screening a BAC vector library produced from a cow rumen microbiome, reported a hit rate of 1:853 using xylan as the screening substrate. Nguyen *et al.* (2012) [39], screening their buffalo rumen metagenomic library, found hit rates of 1:108 and 1:2,500 using AZCL-HE-cellulose and AZCL-xylan, respectively. The authors suggested that the relatively rates were high hit rate on AZCL-HE-cellulose can be attributed to the use of ENZhance cell permeabilizing reagent. These results emphasize the advantages of combining large-insert libraries (maximizing the probability of identifying gene clusters whose components perform complementary functions), enriched-function systems (such as the cow rumen) and reagents that enhance the host cell

**Table 4. Literature comparison for functional metagenomic approaches to screen (hemi)cellulases.**

| Microbial source (enrichment bred from) | Vector / Host | Screened clones | Substrates for (hemi)cellulases screening | Hit rate * | Reference |
|---|---|---|---|---|---|
| Cow rumen | pCCl-BAC [a] / *E. coli* EPI300 | 15360 | Xylan | 1/853 | Zhao *et al.* (2010) [38] |
| Termite gut | pCC1FOS [b] / *E. coli* EPI100 | 40000 | 5-Bromo-3-indolyl-beta-D-xylopyranoside | 1/740 | Bastien *et al.* (2013) [32] |
| | | 40000 | 5-bromo-4-chloro-3-indolyl-alpha-L-arabinofuranoside | 1/1.212 | |
| | | 40000 | Oat AZCl-xylan | 1/4.444 | |
| Cow rumen | pCCFOS [b] / *E. coli* EPI300-T1 | 17000 | *p*NP-beta-D-glucopyranoside + *p*NP-beta-D-cellobioside | 1/5.666 | Del Pozo *et al.* (2012) [16] |
| Buffalo rumen | pCCFOS1 [b] / EPI300-T1 | 10,000 | AZCL-HE-cellulose | 1:108 | Nguyen *et al.* (2012) [39] |
| | | 10,000 | AZCL-xylan | 1:2.500 | |
| Avicel- enrichment culture (soil) | pUC19 [c] / *E. coli* DH5α | 57000 | Rimazol brilliant blue dyed-CMC | 1/7.125 | Mori *et al.* (2014) [29] |
| Pulp- enrichment culture (soil) | | 63000 | Rimazol brilliant blue dyed-xylan | 1/63.000 | |
| Ikaite tufa columns | pGNS-BAC [a] / *E.coli* DH10B | 2843 | 5-bromo-4-chloro-3-indolyl-beta-d-galactopyranoside | 1/1.421 | Vester *et al.* (2014) [49] |
| Grassland soil | pCC1FOS [b] / *E. coli* EPI300 | 4600 | Dye-labeled hydroxyethyl xylan | 1/4.600 | Nacke *et al.* (2012) [28] |
| | | 4600 | Dye-labeled hydroxyethyl cellulose | 1/2.300 | |
| Human gut | pCC1FOS [b] / *E. coli* EPI300 | 156000 | AZCL-xylan | 1/4.588 | Tasse *et al.* (2010) [50] |
| | | 156000 | AZCL-beta-glucan | 1/1.772 | |
| | | 136000 | AZCL-galactan | 1/1.271 | |
| Filter paper -enrichment culture (Earthworm) | pCCFOS [b] / *E. coli* EPI300 | 115000 | *p*NP-beta-D-glucopyranoside | 1/2.090 | Beloqui *et al.* (2010) [30] |
| Compost from pig manure | pCC2FOS [b] / *E. coli* EPI300 | 12380 | Xylan | 1/2.476 | Jeong *et al.* (2012) [51] |
| Forest soil | Lambda ZAP Express [c] / *E. coli* XL1 blue | 10000 | CMC | 1/10.000 | Wang *et al.* (2009) [52] |
| Elephant dung | | 10000 | CMC | 1/5.000 | Wang *et al.* (2009) [52] |
| Cow rumen | pCCFOS [b] / *E. coli* EPI300 | 12288 | *p*NP-alpha-galactopyranoside | 1/12.288 | Ferrer *et al.* (2012) [27] |
| | | 12288 | *p*NP-alpha-L-arabinofuranoside | 1/4.096 | |
| | | 12288 | CMC | 1/6.144 | |
| Alkaline-polluted soil | pGEM-3Zf[c] / *E. coli* DH5a | 30000 | Esculine | 1/15.000 | Jiang *et al.* (2011) [53] |
| Soil | pHBM803 [c] / *E. coli* XL10-Gold | 24000 | Remazol brilliant blue dyed-xylan | 1/24.000 | Hu *et al.* (2008) [54] |
| $^{13}$C-cellulose-enriched DNA (soil) [e] | pJC8 [d] / E. coli HB101 | 2876 | CMC | 1/360 | Verastegui *et al.* (2014) [40] |
| Wheat straw- enrichment culture (soil) | pCC2FOS [b] / *E. coli* EPI300 | 22000 | Mixture of 6 indolyl-monosaccharides | 1/440 | This study |
| Torrefied wheat straw- enrichment culture (soil) | | 22000 | Mixture of 6 indolyl-monosaccharides | 1/1,047 | |

[a] Bacterial artificial chromosome (BAC); [b] Fosmid; [c] Plasmid; [d] Cosmid; [e] High-molecular weight DNA from the $^{13}$C-cellulose-enriched SIP incubations for the three soils.
*Hit rate was determined based on the number positives clones in the initial screening / total of screened clones.
AZCL (azurine cross-linked)
CMC (carboxymethylcellulose)
*p*NP (para-nitrophenyl)

permeabilities, allowing releases of enzymes Here, "biased" communities were produced on wheat straw as the carbon source, which is thought to raise the relative abundance of target genes in the consortium and thus in the fosmid clones. However, Mori *et al.* (2014) [29], using a pUC19 library produced from pulp enrichments, obtained a hit rate of only 1:63,000 with rimazol brilliant blue dyed-Xylan. Beloqui *et al.* (2010) [30] reported a hit rate of 1:2,090 in a library prepared from filter paper -enrichments inoculated from earthworm gut extract, using as the screening substrate pNP-beta-D-glucopyranoside. Another interesting approach that potentially leads to highly efficient discovery of GH activities is the construction of metagenomic libraries prepared from DNA selected following stable isotope probing (DNA-SIP) using multiple labeled plant-derived carbon substrates. For example, Verastegui *et al.* (2014) [40] showed a hit rate of 1:360 using DNA from $^{13}$C-cellulose-enriched incubations on the basis of CMC as a substrate (Table 4). Clearly, the rates of obtaining positive clones are related to the cloning vector used, the metagenome source, the screening technique (substrates and desired activity) and the host cells. On top of that, in many cases stochastic (chance) factors play a role as well, which may relate to the relatively low sample sizes [36].

In functional screening of metagenomic libraries, proper selection of the substrate is highly recommended. Initial selection of substrate-active clones with "general" substrates or mixtures of substrates followed by more specific ones may represent a desirable "layered" approach. Recently, a new generation of multi-colored chromogenic polysaccharide substrates has been developed [41]. These substrates can be used to screen for GH activities (in this case, focusing on endo-enzymes). They show versatility and are convenient for high-throughput analyses for first-level screenings. Additionally, substrates representing -at least partially- the complexity of plant cell walls were produced, enabling activity screens on "real-world" plant polysaccharides.

In our study, four fosmid inserts carried genes that could be directly linked to the enzymatic activities based on sequence homologies to known enzymes and predicted and detected protein sizes. Thus, proteins of predicted sizes (1,027 and 1,028 amino acids, giving proteins of ~116 kDa) from fosmids NT2-2 and T17-2 (selected as positive on X-Gal) did transform pNPGal and MUFGal (zymogram). These were both annotated as beta-galactosidases of family GH2 (EC 3.2.1.23) (Table 3). Fosmids T5-5 and T4-1, positive on X-Xyl/pNPXyl and mixed chromogenic substrates, respectively, revealed the presence of genes predicted to produce proteins of 789 amino acids (~ 86 kDa), which were annotated as beta-xylosidases of family GH3 (EC 3.2.1.37). Interesting, fosmid T4-1 showed activity only on the substrate mixes, but not on single X-Xyl. This clone showed slight activity on pNPXyl (0.113 U/mg ± 0.002 at 40°C, pH 7.0) and revealed a protein of size between 75 to 100 kDa, which was likely encoded by gene 24 (Table 3). The protein was positive on the zymogram using MUFXyl as a substrate (Figure

5). The expression of this GH3 family gene may be regulated by the presence of the other substrates. Based on this rationale, on X-Xyl alone its activity might not be detected, whereas the presence of other substrates might spur activity, similar to what may happen in nature. Such a finding opens up a new paradigm in the screening of active enzymes from metagenomic libraries. Interestingly, in fosmid T4-1 a predicted xyloside transporter gene (*Xyn*T) was detected, which matched the CAZy GH17 family. Proteins from this family can have glucan endo-1,3-beta-glucosidase activity, suggesting a possible involvement in activity on mixed substrates.

Although fosmid NT18-17 was positive on X-Glu and pNPGlu, we did not detect any gene related with its predicted alpha-glucosidase activity. However, the detected genes for family GH58 (endo-N-acetylneuraminidase) and family GH27 (alpha-galactosidase) proteins might encode the activity (see Additional file 8 for a summary of activities associated with CAZy enzyme families described in this study). Similarly, fosmid NT18-21 showed alpha-glucosidase activity, with no GH family genes being detected in the insert. In this fosmid, genes for maltose/maltodextrin transporters were found. Maltose (an alpha 1-4 linked glucose dimer) resembles a cellulose dimer (albeit beta 1-4 linked). Maltose is released from starch by amylose/amylopectin-degrading activity (e.g. GH13-alpha-amylase, pullulanase or alpha-glucosidase). We surmised that starch that was initially present in the wheat straw used for consortium breeding incited the selection of such systems. The chromogenic substrates, e.g. X-Glu used in this study were surmised to report alpha-glucosidase activities, but genes for such enzymes were not detected. Possibly, gene 25 (annotated as an alpha-aspartyl dipeptidase; EC 3.4.13.21) or gene 24 (hypothetical protein) were responsible for the activity (Additional file 6).

Fosmid clone 10BT showed consistent activity only on mixed substrates. In addition, 10BT showed activity on pNPXyl, much below that shown by clone T5-5 (~14,8% of relative activity at 40°C, pH 7.0; Figure 3d). The finding of two genes producing proteins related to GH39 and GH53 families was revealing. Interestingly, family GH39 proteins have been linked to beta-xylosidase and alpha-L-iduronidase activities. Moreover, GH53 family proteins can have beta-1,4-galactanase activity (EC 3.2.1.89). The latter is possibly linked to the degradation of galactans and arabinogalactans, both integral parts of the pectin component of plant cell walls [41]. Interestingly, Jiménez *et al.* (2012) [43] recently found a novel cold-tolerant esterase, which had originally been annotated as a MarR family transcriptional regulator. Thus, we surmised that the gene originally predicted to encode an AraC transcriptional regulator may be responsible for the activity on pNPXyl. Similar to clone T4-1, this clone could require the presence of other types of substrates to enable detection of its full plethora of activities. However, given that we still don't know the mechanism involved, further studies are required, for example subcloning, transposon mutagenesis and detection of

activities on different substrates.

The high activity of fosmid NT2-2 compared to that of clones T5-5 and NT18-17 suggested a raised expression of the gene encoding beta-galactosidase (~116 kDa, as evident by the zymogram; Figure 5). Interestingly, the high activities of clones NT2-2 and T5-5 at 55°C and pH 10.0 pointed to their potential usefulness in pulp bleaching processes. The novelty attributed to these genes (3 and 12) was based on the low amino acid identities (less than 84%) and coverage values (less than 91%) versus the best hits in the NCBI database (Table 3). The functional analysis done by us directly from the metagenomic clones indicated substrate specificities and temperature/pH optima, and constitutes an easy way to select clones useful for biotechnology applications. In addition, subcloning, overexpression, induction and subsequent protein purification are labor-intensive and not always successful (e.g. due to the low solubility of the enzyme).

The leading industrial source of cellulase cocktails used for plant biomass biodegradation purposes is *Trichoderma reesei*. Several strains exist and their secretomes have been widely used to develop new commercial cocktails. However, *T. reesei* secretomes are dominated by endoglucanases and it usually produces low quantities of xylanases, arabinofuranosidases, galactosidases and beta-glucosidases. Hence, addition of exogenous enzymes to the secreted fraction could improve the hydrolytic efficiency [44]. Based on this premise, the enzymes detected in the clones NT2-2, T5-5 and NT18-17 might serve as components of new (hemi)cellulolytic cocktails. These may be combined with the commercial cellulases to improve plant biomass degradation for second-generation biofuel production. Additionally, thermo-alkaliphilic xylosidases are valuable with respect to their usefulness in pulp bleaching processes [13, 14].

In a previous study [45], Bacteroidetes-related genes for hemicellulases were found to be prominent amongst the dominant enzymes, whereas *Klebsiella*-related ones were less abundant. Both groups of organisms are key dominant types in our bacterial consortia bred on wheat straw. In the current study, evidence was found for the contention that *Klebsiella*-related organisms were at the basis of most cloned genes for biodegradative enzymes. The taxonomic closeness between this putative source organism (*Klebsiella*) and the heterologous host (*E. coli*) used, versus the remoteness in the case of Bacteroidetes, may have been a key factor explaining this finding. Unfortunately, the current study did not detect fosmids with activities on X-Fuc and X-Man. Such activities might be mostly associated with members of the Bacteroidetes (e.g. *Sphingobacterium*), as recently indicated by Jiménez *et al.* (2015) [45]. Finally, the differential association of fosmid NT2-2 and 10BT genes with *Pseudomonas putida* versus *Klebsiella* sp. was remarkable. IS-elements indicative of horizontal gene transfer were not detected, suggesting these fosmids might originate from fusions of two regions originating from different parental organisms. Alternatively, the insert may have come from a new Gammaproteobacteria species.

## Conclusions

Here, we propose a multi-substrate screening approach as a sound strategy that allows to detect multiple activities in a single initial assay. This methodology is less time-consuming than single-substrate approaches and can even be applied in high-throughput set-ups, as in agar plates. The strategy yielded high hit rates of genes for relevant enzymes compared with recent relevant literature data. Based on this methodology, we retrieved fosmids with beta-galactosidase, beta-xylosidase and alpha-glucosidase activities, whereas other fosmids showed activity only in the presence of mixed chromogenic substrates. Two fosmids, NT2-2 (GH2- beta-galactosidase) and T5-5 (GH3- beta-xylosidase), showed enzymatic activities at high temperatures and pH values, making these clones interesting sources for future biotechnological applications.

## Authors' contributions

MM performed most of the experiments, particularly the zymograms. DJJ helped to design the experiments, participated in the sequence analyses, performed the enzymatic assays and drafted the manuscript. PS contributed to the functional screening procedures. JDvE conceived the study, participated in its design and coordination and contributed to the drafting of the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank F. Dini-Andreote for his critical review of the manuscript and D. Devlitsarov for experimental support. Further thanks are due to H. Ruijssenaars and R. van Kranenburg for their scientific support. This work was supported by the Netherlands Ministry of Economic Affairs and the BE-Basic partner organizations (http://www.be-basic.org).

## Supplementary tables Chapter 2

Additional file 1: Complete gene annotation of fosmid insert NT2-2. (XLS 41 kb)
Additional file 2: Complete gene annotation of fosmid insert T5-5. (XLS 42 kb)
Additional file 3: Complete gene annotation of fosmid insert NT18-17. (XLS 43 kb)
Additional file 4: Complete gene annotation of fosmid insert T4-1. (XLS 43 kb)
Additional file 5: Complete gene annotation of fosmid insert 10BT. (XLS 43 kb)
Additional file 6: Complete gene annotation of fosmid insert NT18-21. (XLS 39 kb)
Additional file 7: Complete gene annotation of fosmid insert T17-2. (XLS 36 kb)
Additional file 8: Summary of activities associated with CAZy enzyme families described in this study.

These files can be downloaded directly in the article web page
http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2404-0#Sec25

## References


1. Limayem A, Ricke SC. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Prog Energy Combust Sci. 2012;38:449-467.
2. Himmel ME, Xu Q, Luo Y, Ding SY, Lamed R, Bayer EA. Microbial enzyme systems for biomass conversion: Emerging paradigms. Biofuels. 2010;1:323-341.
3. Dougherty MJ, D'haeseleer P, Hazen TC, Simmons BA, Adams PD, Hadi MZ. Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. BMC Biotechnol. 2012;12:38.
4. De Souza RW. Microbial degradation of lignocellulosic biomass. In: Chandel A, Da Silva, S. Sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization. Brazil: InTech; 2013. doi: 10.5772/54325.
5. Turner P, Mamo G, Karlsson E. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact. 2007;6:9.
6. Sims REH, Mabee W, Saddler JN, Taylor M. An overview of second generation biofuel technologies. Bioresour Technol. 2010;101:1570-1580.
7. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315:804-807.
8. Gowen CM, Fong SS. Exploring biodiversity for cellulosic biofuel production. Chem Biodivers. 2010;7:1086-1097.
9. Xing MN, Zhang XZ, Huang H. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnol Adv. 2012;30(4):920-929.
10. Jayani RS, Saxena S, Gupta R. Microbial pectinolytic enzymes: A review. Process Biochem. 2005;40:2931-2944.
11. Gilbert HJ, Stålbrand H, Brumer H. How the walls come crumbling down: Recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol. 2008;11:338-348.
12. Sweeney MD, Xu F. Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: Recent developments. Catalysts. 2012;2:244-263.
13. Clarke JH, Davidson K, Rixon JE, Halstead JR, Fransen MP, Gilbert HJ, *et al.* A comparison of enzyme-aided bleaching of softwood paper pulp using combinations of xylanase, mannanase and alpha-galactosidase. Appl Microbiol Biotechnol. 2000;53(6):661-667.
14. Sudha B, Veeramani H, Sumathi S. Bleaching of bagasse pulp with enzyme pre-treatment. Water Sci Technol. 2003;47(10):163-168.
15. Gao D, Uppugundla N, Chundawat SP, Yu X, Hermanson S, Gowda K, *et al.* Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. Biotechnol Biofuels. 2011;4:5.
16. Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, *et al.* Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnol Biofuels. 2012;5:73.
17. Wongwilaiwalin S, Rattanachomsri U, Laothanachareon T, Eurwilaichitr L, Igarashi Y, Champreda V. Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system. Enzyme Microb Technol. 2010;47:283-290.
18. Wang W, Yan L, Cui Z, Gao Y, Wang Y, Jing R. Characterization of a microbial consortium capable of degrading lignocellulose. Bioresour Technol. 2011;102:9321-9324.
19. Jiménez DJ, Korenblum E, Van Elsas JD. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl Microbiol Biotechnol. 2014;98:2789-2803.
20. Cheng J, Zhu M. A Novel co-culture strategy for lignocellulosic bioenergy production: A systematic review. Int J Mod Biol Med. 2012;1:166-193.

21. Deangelis KM, D'Haeseleer P, Chivian D, Simmons B, Arkin AP, Mavromatis K, *et al.* Metagenomes of tropical soil-derived anaerobic switchgrass-adapted consortia with and without iron. Stand Genomic Sci. 2013;7:382-398.
22. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, Tangphatsornruang S, Eurwilaichitr L, Igarashi Y, *et al.* Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. Appl Microbiol Biotechnol. 2013;97:8941-8954.
23. Zhou Y, Pope PB, Li S, Wen B, Tan F, Cheng S, *et al.* Omics-based interpretation of synergism in a soil-derived cellulose-degrading microbial community. Sci Rep. 2014;4:5288.
24. Simon C, Daniel R. Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol. 2009;85:265-276.
25. Li LL, Taghavi S, McCorkle SM, Zhang YB, Blewitt MG, Brunecky R, *et al.* Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases. Biotechnol Biofuels. 2011;4(1):23.
26. Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S, *et al.* Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. PLoS One. 2010;5:e8812.
27. Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, *et al.* Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. PLoS One. 2012;7:e38134.
28. Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett. 2012;34:663-675.
29. Mori T, Kamei I, Hirai H, Kondo R. Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. Springerplus. 2014;3:365.
30. Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni ME, Polaina J, *et al.* Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl Environ Microbiol. 2010;76:5934-5946.
31. Ekkers DM, Cretoiu MS, Kielak AM, van Elsas JD. The great screen anomaly-a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol. 2012;93:1005-1020.
32. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R, *et al.* Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. Biotechnol Biofuels. 2013;6:78.
33. Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, *et al.* Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol. 2007;127:575-592.
34. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina OV, Golyshin PN. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. Microb Biotechnol. 2015. doi: 10.1111/1751-7915.12309 [Epub ahead of print]
35. Jiménez DJ, Dini-Andreote F, Van Elsas JD. Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. Biotechnol Biofuels. 2014;7:92.
36. Uchiyama T, Miyazaki K. Functional metagenomics for enzyme discovery: Challenges to efficient screening. Curr Opin Biotechnol. 2009;20:616-622.
37. Gould SW, Chadwick M, Cuschieri P, Easmon S, Richardson AC, Price RG, *et al.* The evaluation of novel chromogenic substrates for the detection of lipolytic activity in clinical isolates of *Staphylococcus aureus* and MRSA from two European study groups. FEMS Microbiol Lett. 2009;297:10-16.
38. Zhao S, Wang J, Bu D, Liu K, Zhu Y, Dong Z, *et al.* Novel glycoside hydrolases identified by

screening a Chinese Holstein dairy cow rumen-derived metagenome library. Appl Environ Microbiol. 2010;76:6701-6705.
39. Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V, *et al.* Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. Biosci Biotechnol Biochem. 2012;76(6):1075-1084.
40. Verastegui Y, Cheng J, Engel K, Kolczynski D, Mortimer S, Lavigne J, *et al.* Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. MBio. 2014;5:e01157-14.
41. Kračun SK, Schückel J, Westereng B, Thygesen LG, Monrad RN, Eijsink VGH, *et al.* A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes. Biotechnol Biofuels. 2015;8:70.
42. Hinz SW, Pastink MI, van den Broek LA, Vincken JP, Voragen AG. *Bifidobacterium longum* endogalactanase liberates galactotriose from type I galactans. Appl Environ Microbiol. 2005;71:5501-5510.
43. Jiménez DJ, Montaña JS, Alvarez D, Baena S. A novel cold active esterase derived from Colombian high Andean forest soil metagenome. World J Microbiol Biotechnol. 2012;28:361-370.
44. Mohanram S, Amat D, Choudhary J, Arora A, Nain L. Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries. Sustain Chem Process. 2013;1:15.
45. Jiménez DJ, Chaves-Moreno D, Van Elsas JD. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. Sci Rep. 2015;5:13845.
46. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, *et al.* The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
47. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 2010;20:1574-1584.
48. Huson DH, Weber N. Microbial community analysis using MEGAN. Methods Enzymol. 531:465-485.
49. Vester JK, Glaring MA, Stougaard P. Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb Cell Fact. 2014;13:72.
50. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, *et al.* Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 2010;20:1605-1612.
51. Jeong YS, Na HB, Kim SK, Kim YH, Kwon EJ, Kim J, *et al.* Characterization of xyn10J, a novel family 10 xylanase from a compost metagenomic library. Appl Biochem Biotechnol. 2012;166:1328-1339.
52. Wang F, Li F, Chen G, Liu W. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res. 2009;164:650-657.
53. Jiang C, Li SX, Luo FF, Jin K, Wang Q, Hao ZY, *et al.* Biochemical characterization of two novel β-glucosidase genes by metagenome expression cloning. Bioresour Technol. 2011;102:3272-3278.
54. Hu Y, Zhang G, Li A, Chen J, Ma L. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol. 2008;80:823-830.

# Chapter 3

## Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library.

**Mukil Maruthamuthu and Jan Dirk van Elsas.**
*Biotechnol Biofuels.* 2017 Jun 2;10:142. doi: 10.1186/s13068-017-0808-y.

**Abstract**

Enzyme discovery is a promising approach to aid in the deconstruction of recalcitrant plant biomass in an industrial process. Novel enzymes can be readily discovered by applying metagenomics on whole microbiomes. Our goal was to select, examine and characterize eight novel glycoside hydrolases that were previously detected in metagenomic libraries, to serve biotechnological applications with high performance.

Here, eight glycosyl hydrolase (GHase) family candidate genes were selected from metagenomes of wheat-straw-degrading microbial consortia using molecular cloning and subsequent gene expression studies in *Escherichia coli*. Four of the eight enzymes had significant activities on either *p*NP-β-D-galactopyranoside, *p*NP-β-D-xylopyranoside, *p*NP-α-L-arabinopyranoside or *p*NP-α-D-glucopyranoside. These proteins, denoted as proteins 1, 2, 5 and 6, were his-tag purified and their nature and activities further characterized using molecular and activity screens with the *p*NP-labeled substrates. Proteins 1 and 2 showed high homologies with (1) a β-galactosidase (74%) and (2) a β-xylosidase (84%), whereas the remaining two (5 and 6) were homologous with proteins reported as a diguanylate cyclase and an aquaporin, respectively. The β-galactosidase- and β-xylosidase-like proteins 1 and 2 were confirmed as being responsible for previously found thermo-alkaliphilic glycosidase activities of extracts of *E. coli* carrying the respective source fosmids. Remarkably, the β-xylosidase-like protein 2 showed activities with both *p*NP-Xyl and *p*NP-Ara in the temperature range 40-50°C and pH range 8.0-10.0. Moreover, proteins 5 and 6 showed thermotolerant α-glucosidase activity at pH 10.0. *In silico* structure prediction of protein 5 revealed the presence of a potential "GGDEF" catalytic site, encoding α-glucosidase activity, whereas that of protein 6 showed a "GDSL" site, encoding a 'new family' α-glucosidase activity.

Using a rational screening approach, we identified and characterized four thermo-alkaliphilic glycosyl hydrolases that have the potential to serve as constituents of enzyme cocktails that produce sugars from lignocellulosic plant remains.

## Background

Plant biomass is considered to represent a sustainable source of sugars for biofuel production via fermentation. In this biomass, lignocellulosic material is the key source of 'renewable' energy [1]. Lignocellulose consists of the major compounds cellulose (40-50%) and hemicellulose (25-30%), next to lignin [2]. Hemicellulose is composed of different pentose (xylose, arabinose) and hexose (mannose, galactose, glucose) sugars that are linked by α- and/or β- glycosidic bonds [3]. All of these sugars are widely used in biotechnology to produce bio-based materials, such as biofuels, plastics and other chemicals [4,5]. In particular, hemicellulases are key to the degradation of plant biomass; the hemicellulose fraction of the plant biomass represents a rich source of D-xylose, which is considered to constitute a key sugar for further biotechnological approaches. Thus, in the light of the currently still imperfect substrate unlocking approaches [6], a great interest has arisen to enhance the industrial enzyme-mediated lignocellulose hydrolysis methods [7], with hemicellulose as a prime target.

Different sets of hydrolytic enzymes (glycosyl hydrolases) are likely required for the complete deconstruction of hemicellulose compounds in order to obtain a mixture of sugars. In nature, communities of microorganisms, which include fungi and bacteria, often produce mixtures of glycosyl hydrolases (GHases) that complete the lignocellulose breakdown processes [8,9]. Moreover, to increase the efficiency of industrial hydrolysis for lignocellulose breakdown, multi-species microbial consortia play vital roles [10–12]. However, despite large research efforts over the past decade, our limited understanding of how the glycosyl hydrolases and their associated enzymes and/or proteins function together to break down lignocellulosic materials remains a key limitation for many applications [13].

In a previous study, we constructed fosmid libraries from two wheat straw-degrading microbial consortia, which were subsequently screened for the presence of genes for (hemi)cellulose degrading enzymes using a multi-substrate approach [14]. In this endeavor, we screened for the presence of genes encoding 12 different GH family proteins (using CAZy database annotation), which were considered to possess the desired enzymatic activities, i.e. β-galactosidase, β-xylosidase and α-glucosidase. Heat- and alkali-tolerant enzymatic activities were found with extracts produced from four *E. coli* fosmid clones, denoted NT2-2, T4-1, T5-5 and NT18-17 [14]. The first three clones were identified as containing genes encoding proteins with β-galactosidase and β-xylosidase activities. On the other hand, clone NT18-17 presumably carried a gene for a protein with α-glucosidase activity, next to those for other glycoside hydrolase family enzymes, as predicted by CAZy database annotation. One more fosmid clone, 10BT, revealed enzymatic activity with mixtures of four substrates; in it, genes for proteins of families GH39 and GH53 were identified. However, the work with these five fosmid clones, into each of which up to 35 kb of metagenomic DNA (encompassing up to 30 genes), was cloned, precluded the precise determination of the exact function of each of the

predicted proteins.

In the current study, we selected eight genes from the aforementioned five fosmid clones, of which three were predicted to produce enzymes with novel thermo-alkaliphilic activity. The genes were subcloned in the pET28b(+) expression vector and (over)expressed in *E. coli*, after which the gene products were purified and biochemically characterized. The study explored, confirmed and refined the hypothesis that the selected proteins indeed have the thermo-alkaliphilic activities predicted from our previous study [14]. The characterization and *in vitro* studies identified two new glycoside hydrolase family enzymes with α-glucosidase activities.

## Materials and methods

### Cloning system

Cloning vector pET28b(+) (Novagen, Amsterdam, The Netherlands) was used for the expression of the selected genes. *Escherichia coli* JM109 competent cells (Promega, Leiden, The Netherlands), as well as BL21(DE3) and Origami2(DE3) pLysS cells (Novagen, Amsterdam, The Netherlands) were used as host strains for cloning and expression studies. Restriction enzymes (*Eco*R1, *Bam*H1, *Hind*III, and *Xho*1) and T4-DNA ligase were purchased from Fermentas (Amsterdam, The Netherlands) and used in accordance with the manufacturer's instructions.

### Extraction of DNA and molecular cloning into expression plasmids

Selected *E. coli* EPI 300 fosmid clones NT2-2, T4-1, T5-5, NT18-17 and 10BT (Figure 1) were cultured in 4ml of Luria Broth (LB) supplemented with 12.5 μl/ml chloramphenicol (Cm; Sigma-Aldrich Chemie B.V, Zwijndrecht, The Netherlands). Then, fosmid DNA was extracted as described [14]. PCR primers were designed in regions outside of each gene, adding specific restriction sites to their 5'-ends (Table 1). Thus, full-length genes were generated from the clones by each PCR [Initial denaturation at $98^0$C for 30 sec followed by 35 cycles of 10 sec at $98^0$C, 30 sec at $64^0$C and 1.5 min at $72^0$C, with a final extension step of $72^0$C (for 10 min)]. The PCR products were digested with selected restriction enzymes and then analyzed on 1% agarose gels. All patterns were in conformity with the predicted ones (Table 1). Then, full PCR products were run on gel and recovered from it using the Zymoclean™ Large Fragment DNA recovery kit (Zymo Research, Irvine, USA). Following recovery and purification, each DNA fragment was then ligated into expression vector pET28b(+), which was followed by transformation of *Escherichia coli* JM109 competent cells (Promega. Leiden, The Netherlands).

### Verification of inserts, and expression and purification of target proteins

All transformations were successful, and so eight recombinant constructs were produced. The success of cloning was further confirmed by colony PCR with the

respective primers (Table 1), yielding single amplicons of the expected sizes for all genes (Table 1). This was followed by single and double restriction of the constructs with the relevant enzymes (Table 1). Thus, *Eco*R1/*Bam*H1 was used for gene 1 (3,146 bp); *Hind*III/*Bam*H1 for gene 2 (2.4kb); *Eco*R1/*Hind*III for gene 3 (2.4kb); *Bam*H1/*Xho*1 for gene 4 and gene 8 (size of 624bp and 1134bp respectively) and *Hind*III/*Bam*H1 for gene 5(1.2kb), gene 6 (673bp) and gene 7 (2.0kb). Clones carrying the selected eight genes were then selected and purified, after which they were used to inoculate 4ml LB tubes supplemented with 50ug/ml of kanamycin (to select for the maintenance of the plasmid with insert). Tubes were incubated at 37$^0$C (shaking at 220rpm) for 16-19h. Following incubation, plasmid extraction was carried out using the QIAprep Spin Miniprep Kit (Qiagen, The Netherlands). All recombinant plasmids had the expected inserts; they were further checked by digestion with the aforementioned enzymes and enzyme combinations. Then, the full plasmids were introduced into *E. coli* strains BL21(DE3) and Origami2(DE3) pLysS (Novagen, Amsterdam, The Netherlands) competent cells, via transformation. These two strains facilitate the testing of the expression of the cloned genes. Selected transformants were purified and the presence of the correct inserts verified. They were then grown in kanamycin

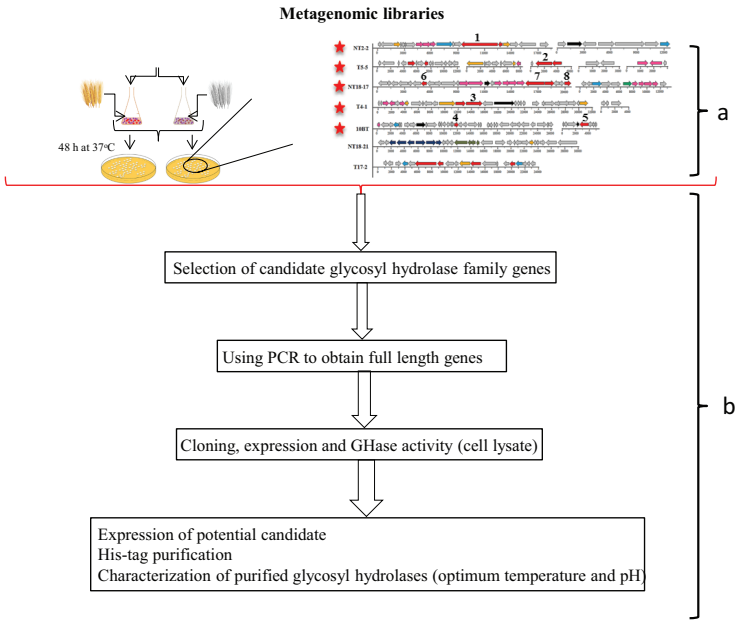



**Figure 1.** Candidate gene selection and cloning strategy used in this study. **a.** Selected candidate genes from functional screening of fosmid libraries [14]. **b.** Experimental setup

(50ug/ml)-supplemented 2X-PY medium (2 ml; 16g Bacto-tryptone, 10g yeast extract, 10g NaCl.$H_2O$/1L, pH 8.0) at 37$^0$C (220 rpm, overnight). A fresh (200ml) 2X-PY flask was then inoculated, establishing an initial $OD_{600}$ of 0.05, after which

the culture was grown at $37^0$C (shaking, 220rpm) to an $OD_{600}$ of 0.5-0.6. Afterwards, the culture was incubated for 1h at $18^0$C (shaking, 220rpm), after which gene expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.5M. Then, the culture was further incubated at $18^0$C for 16-19 h, after which cells were harvested at 4000xg ($4^0$C, 15min). The pellets were resuspended in 5mL of lysis buffer (50mM HEPES, pH 8.0, 300mM NaCl, 50µL 1M DTT (1,4-Dithiothreitol), 1 protease inhibitor mini tablet (Roche, Sigma-Aldrich Chemie B.V, Zwijndrecht, The Netherlands) and the mixtures kept on ice for 15min. Then, cells were disrupted using sonification with the following parameters: 40 cycles – 6sec on/15sec off – amplitude 6-10µm. After this treatment, the resulting cell lysates were centrifuged at 15000xg for 15 min at $4^0$C. The supernatants were removed and stored, and 10µl was checked with 12% SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), followed by staining with the Pierce™ 6xHis protein tag stain reagent set (ThermoFisher Scientific, Waltham, USA). The preparation was then heated to $60^0$C for 15-20 min and centrifuged at 15000xg to remove insoluble debris. Purification of his-tagged proteins from the crude extracts was then carried out by gravity flow chromatography through agarose. Thus, 600ul of Ni-NTA agarose (Qiagen, Hilden, Germany) was added to 10ml of lysis buffer. Incubation was for 5min (shaking, $4^0$C), before the mixture was centrifuged for 5 min at 800xg at $4^0$C. The supernatant was discarded and then 10ml of equilibration buffer (50mM HEPES and 300mM NaCl) was added, after which the mixture was incubated as mentioned above. A short spin followed. Then, the crude extract was added to the resin and incubated for 1h before it was transferred into a gravity flow column and incubated at $4^0$C until the resin bed settled down. The cell-free lysates were removed by gravity flow and unbound proteins were washed 3 times with 10 ml of wash buffer (50mM HEPES, 300mM NaCl, 20mM Imidazole). The bound enzyme was eluted with 3 ml of elution buffer (50mM HEPES, 300mM NaCl, 400mM Imidazole). The enzyme samples were concentrated using Amicon ultra-15 centrifugal filter units (Millipore, Amsterdam, The Netherlands) and quantified using the Bradford method (Bradford 1X dye, Biorad, Veenendaal, The Netherlands). The purity was then analyzed by running 12% SDS-PAGE followed by staining with Pierce™ 6xHis protein tag stain reagent set (Figure 2).

**Substrate specificity testing**

Fifty microliter volumes containing approximately 0.6ug of enzyme were used for testing enzyme activity on 3, 5 and 10mM of *p*NP substrate (*p*NP-β-D-galactopyranoside, *p*NP-β-D-xylopyranoside, *p*NP-α-L-arabinopyranoside and *p*NP-α-D-glucopyranoside) (Sigma-Aldrich Chemie B.V, Zwijndrecht, The Netherlands). Fifty mM of Tris-HCl buffer - pH 7.5 was used in the ratio of 1:1. The reaction mixture was incubated at $40^0$C for 0.5-1h, after which the reactions were deactivated on ice for 10min. To validate the enzyme activity, the

experiments were controlled with i) water and substrate, ii) enzyme with water, and iii) host protein lysate with substrate. The concentration of released *p*NP was determined by measuring the reaction mixture absorbance at 410nm using a calibration curve, as explained [14]. Effects of temperature and pH on hydrolysis activity for all recombinant enzymes were evaluated. For temperature, we used the range 10-70$^0$C at pH 7.5. For pH, we used the range 4.0-10.0 at 50$^0$C, using sodium citrate buffer (pH 4.0 and 5.0), Tris-HCl (pH 7.5, 8.0 and 9.0) and glycine-NaOH buffer (pH 10.0).

Kinetic parameters ($K_m$ and $V_{max}$) for all purified recombinant enzymes (protein 1, 2, 5 and 6) were evaluated by measuring the enzyme activity using 0–10 mM of respective *p*NP as substrate in 50 mM Tris-HCl buffer (pH 7.5) at 40°C for 15min. The data was plotted according to the Lineweaver-Burk method to calculate $K_m$ and $V_{max}$ values. For thermal stability assays, the purified enzymes were pre-incubated at 50$^0$C (according to preliminary data) in the absence of *p*NP substrates. After incubation for different time periods (0, 15, 30, 120 and 180 minutes), enzymatic activity was measured for each enzyme with specific *p*NP substrates, temperature and pH. The inhibitory effect of different concentrations (0 to 1.0%) of 5-hydroxymethylfurfural (5-HMF) and furfural was determined by incubating all enzymes (protein 1, 2, 5 and 6) with respective series of *p*NP dissolved in 50mM Tris-HCl buffer (pH 7.5). For the lignocellulose hydrolysis, the reaction mixtures contained 20 mg of raw wheat straw (RWS) substrate with the enzyme (protein 1, 2, 5, 6 and mixed) treatment, adjusted to 1.5 ml with sodium phosphate buffer (0.1 M, pH 6.0). The experiments were carried out at 50°C (12 - 24h, shaking at 250rpm). After incubation, the mixtures were centrifuged (12000xg for 15 min at 4°C), and the supernatants collected. The amount of total reducing sugars in the supernatants was measured by the dinitro salicylic acid (DNS) colorimetric method.

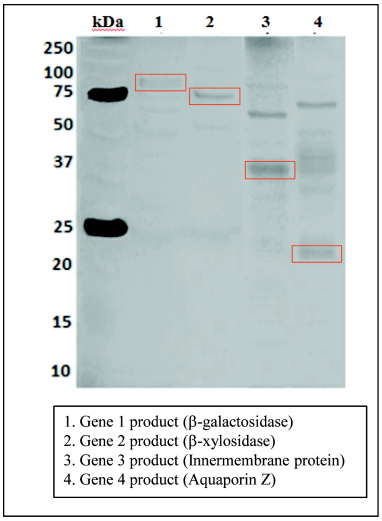



**Figure 2.** Analysis of His-tag purified proteins. Four selected proteins were checked using 12% SDS-PAGE.

**Activity of selected enzymes in the presence of different ions**
The effects of $MgCl_2$, $MnCl_2$ (5mM and 20mM) and NaCl (20mM and 2000mM) were assessed by measuring enzyme activities with specific substrates in the presence versus absence of these additives. Given the potential importance for application, we used a range of pH (4.0 to 10.0) and temperatures (40-50°C). Statistical comparisons among samples were performed using one-way ANOVA (Tukey's test) using the software Past3 (http://folk.uio.no/ohammer/past/).

**Amino acid sequence and structure analyses of the gene 5 and 6 products**
The products of genes 5 and 6 (protein 5 and 6) were investigated for homology with proteins in the non-redundant protein database (www.ncbi.nlm.nih.gov) using BLASTP[15]. In addition, protein domains were characterized by searching the protein family database Pfam [16]. The two proteins were initially annotated as (1) a diguanylate cyclase and (2) an aquaporin Z. For each gene, a multiple sequence alignment was constructed using COBALT, a tool that finds a collection of pairwise constraints derived from conserved domain database, protein motif database and sequence similarity, using RPS-BLAST, BLAST-P and PHI-BLAST [17]. In addition, domain analyses were done for function prediction by the conserved domain architecture retrieval tool (CDART)[18].

***Structure analyses*** - We used Phyre2 to reconstruct the tertiary structures of proteins 5 and 6 [19]. Next, we generated a list of proteins with structural similarities to the protein models using the Dali server[20]. A three-dimensional model of the best three matches for each protein was retrieved from the Protein Data Bank (PDB)[21]. For protein 5, these matches were: activated response regulators, namely a signaling protein [PDB: 1W25], a transferase [PDB: 2WB4] and a lyase [2V0N] from *Caulobacter vibrioides* (Alphaproteobacteria). For protein 6, the matches were membrane proteins (porins) from *Agrobacterium fabrum* [PDB: 3LLQ] and *Escherichia coli* [PDB: 2O9D] and a transport protein, also from *E. coli* [PDB: 3NKA]. Further manipulations, structural alignment and comparisons between the 3D models of the proteins were completed using PyMOL (http://www.pymol.org).

**Results**
**Selection of genes encoding glycosyl hydrolases from five fosmid clones**
In our previous study, we successfully screened two metagenomic libraries generated from wheat-straw-degrading microbial consortia produced by the dilution-to-stimulation approach for genes encoding (hemi)cellulose degrading enzymes [14]. In total, we identified 18 genes for proteins belonging to 12 different

**Table 1:** List of selected genes, PCR primers and predicted size

| Fosmid name | Predicted enzyme (Genes encoding) | Gene no. | Primers | Restriction enzymes | Size (bp) | Approx. Size (kDa) | Closets hit (Identity/coverage) |
|---|---|---|---|---|---|---|---|
| NT2-2 | β- Galactosidase | 1 | FP- TACgaattcACGGCATCAGCGCAATAGTC<br>RP- CAGggatccCGCTTCGCAACTGAGGTTAC | EcoRI<br>BamHI | 3146 | 116 | *Enterobacter hormaechei* (74%/81%) |
| T5-5 | β- Xylosidase | 2 | FP- GCGaagcttGTGTAGTCTTTGCCCATCTC<br>RP- CAGggatccCGCTTCGCAACTGAGGTTAC | HindIII<br>BamH I | 2410 | 85 | *Enterobacter mori*(83%/90%) |
| T4-1 | β- Xylosidase | 3 | FP- GCGgaattcCACCAGATTCAGCCCTACAG<br>RP- CAGaagcttTGTAGTCTTTGCCCATCTCC | EcoRI<br>HindIII | 2439 | 86 | *Enterobacter mori* (83%/90%) |
| 10BT | Trancription Regulator | 4 | FP- GCGggatccGCACAACAAGATGGCTTTAC<br>RP- CAGctcgagTAGTCGCATTTTGACGGAAC | BamHI<br>XhoI | 627 | 22 | *Klebsiella oxytoca* (76%/87%) |
| | Innermembrane protein | 5 | FP- GCGaagcttGTGCAAAGTATCGCTTTAAC<br>RP- CAGggatccAGCAGCATGAATAAGGACAC | HindIII<br>BamHI | 1262 | 45 | *Enterobacter cloacae* (79%/88%) |
| NT18-17 | Aquaporin | 6 | FP- GCGaagcttGGGATCGGCTAGATCTTGAG<br>RP- CAGggatccCAAGCGTCATGAATAAACTG | HindIII<br>BamHI | 673 | 22 | *Hyphomonas neptunium* (65%/80%) |
| | Betahexamidase | 7 | FP- GCGggatccTTTGCGGAGCCTACTCCTTC<br>RP- CAGaagcttTTACAATGGCATCTGATTGG | BamHI<br>HindIII | 1944 | 70 | *Rhizobium leguminosarum* 3841 (50%/64%) |
| | Hypothetical protein | 8 | FP- GCGggatccCCGGATACTCCACACAAAGG<br>RP- CAGctcgagTTATCGCAAGTCGATCAAGG | BamHI<br>XhoI | 1134 | 37 | *Hyphomicrobium denitrificans* (56 %/71 %) |

glycosyl hydrolase (GH) families and another 21 genes for carbohydrate-active enzymes (CBMs, AAs, GTs and CEs; as evidenced by CAZy database annotation), in seven fosmid clones. Five of these fosmid clones, notably NT2-2, T4-1, T5-5, 10BT and NT18-17, were selected here in order to study their enzyme activities in greater detail and match structure with function. Thus, eight genes, denoted genes 1 through 8, encoding putative GHases, were initially selected in these clones (Table 1). This was done on the basis of homologies with genes for known GH active enzymes, as well as protein sizes. Interestingly, three of the five selected fosmid clones, denoted NT2-2, T5-5 and NT18-17, had been predicted to encode putative thermo-alkaliphilic enzymes[14].

Here, we provide the specifics of the eight selected genes. Gene 1 (fosmid NT2-2) was annotated as a gene for a family GH2 β-galactosidase (EC. 3.2.1.23), with a molecular size of ~116 kDa. Its predicted amino acid sequence revealed 74% identity with a β-galactosidase of *Enterobacter hormaechei* by Blast-P. Genes 2 (fosmid T5-5) and 3 (fosmid T4-1) were both predicted to encode a family GH3 β-xylosidase of ~86 kDa (EC. 3.2.1.37; tracked to *Enterobacter mori* with identities of 84% and 83%, respectively). These two genes revealed 86% identity between them. The remaining five genes, i.e. genes 4 and 5 (source fosmid 10BT) and 6, 7 and 8 (source fosmid NT18-17), were predicted to encode enzymes belonging to CAZy families GH39, GH53, GH27, GH20 and GH58, respectively. Blast-P comparisons showed homologies of the predicted gene products with a suite of different proteins, i.e. a transcriptional regulator of the AraC family (protein 4; GH39; 76% identity; organism *Klebsiella oxytoca*), an inner membrane protein (protein 5; GH53; 79% identity; *Enterobacter cloacae*), an aquaporin Z (protein 6; GH27; 65% identity; *Hyphomonas neptunium*), a beta-hexosaminidase (protein 7; GH20; 50% identity; *Rhizobium leguminosarum*) and a hypothetical protein (protein 8; GH58; 56% identity; *Hyphomicrobium denitrificans*), respectively (Table 1).

**Characterization of cloned genes and gene expression**

The eight selected genes (Table 1) were all cloned into the pET28b(+) vector [22], with poly-histidine tag sequences at both the N and C termini. All genes were, thus, successfully introduced into *E. coli* strains BL21(DE3) and Origami2 (DE3) pLysS. To confirm the fidelity of the cloning, the pET28b(+) plasmids with inserts were extracted from selected clones, per gene, of each of the two strains and the presence of the cloned fragments detected by restriction with specific restriction enzymes followed by gel electrophoresis. The observed sizes of the restriction fragments (data not shown) were consistent with the fragment sizes as predicted from the fosmid-derived gene sequences (Table 1).

Expression of the eight cloned genes was then investigated using extracts of grown cultures of both *E. coli* transformant strains (BL21(DE3) and Origami2(DE3) pLysS) in 2x-PY medium, following induction by IPTG. Both the cell and the soluble fractions (culture supernatants) were used in the tests. We

investigated the effects of IPTG concentration (0.25, 0.5 and 1 mM), temperature (18, 30 and 37°C; at 200 rpm) and glucose concentration (0.25, 0.5, 0.75 and 1 mM) on the gene expression levels. In strain BL21(DE3), the target proteins were, unfortunately, mostly found in the cell fractions, except the product of gene 2, where ~50% of protein occurred in the soluble fraction (data not shown). In contrast, the use of *E. coli* Origami2 (DE3) pLysS resulted in seven of the eight proteins being present in the soluble fractions, albeit at different levels. Unfortunately, the single remaining protein (product of gene 4) was only detectable in the cell fraction. The collective data revealed that maximal protein production took place, in the selected strain *E. coli* Origami2 (DE3) pLysS, with 0.5 mM IPTG at 18°C, in the absence of glucose. This was true for all genes (data not shown).

We thus used the total protein-containing lysates of the cultures of *E. coli* Origami2 (DE3) pLysS for analysis of the selected seven gene products, i.e. proteins 1, 2, 3, 5, 6, 7 and 8, on a series of *p*NP-labeled substrates. Indeed, protein 1 revealed high activity on *p*NP-*β*-*D*-galactopyranoside, indicating beta-galactosidase-like activity. This was consistent with the activity that had previously been detected in its source fosmid NT2-2. Interestingly, the cultures of *E. coli* with cloned genes 2 and 3 both yielded soluble fractions that revealed dual activities, i.e. transformation of *p*NP-*β*-*D*-xylopyranoside (β-xylosidase) and *p*NP-*α*-*L*-arabinopyranoside (α-arabinosidase). These genes had been selected from fosmid clones T5-5 and T4-1, respectively, which had previously shown high β-xylosidase activities, but had not shown α-arabinosidase activity. Expectedly, the cultures with cloned gene 4, with predicted β-xylosidase activity, did not yield lysates with any activity on the substrates used. Clearly, the expressed protein of gene 4 was in inclusion bodies, and unfortunately we were unsuccessful in several attempts to recover native forms of it by refolding. In fact, its source fosmid clone 10BT had previously been found to have the gene 4 encoded GH39 family protein linked to *β*-xylosidase activity [14]. Finally, proteins 5 and 6 [14]) both revealed activities towards *p*NP-α-D-glucopyranoside, and there was no detectable enzymatic activity with any of the other *p*NP substrates ( such as *p*NP-*β*-*D*-xylopyranoside and *p*NP-*α*-*L*-arabinopyranoside). Gene 5 had also been selected from fosmid clone 10BT, which had shown consistent enzyme activity on *p*NP-*β*-*D*-xylopyranoside, but not on α-D-glucopyranoside [14]. Gene 6, originating from fosmid clone NT18-17, yielded a product with α-glucosidase activity, which was in line with the predicted activity in its source fosmid. Finally, and against our expectations, the total protein lysates of genes 7 (GH20 family - ®-hexosaminidase; EC 3.2.1.52) and 8 (GH58 family - endo-N-acetylneuraminidase) did not reveal activities with any of the substrates.

For all further work, we selected the four genes 1, 2, 5 and 6 that had yielded soluble proteins with key promising activities, i.e. *β*-galactosidase, *β*-

xylosidase and $\alpha$-glucosidase. Specifically, gene 2 was selected instead of the similar gene 3, because our previous study showed that its mother fosmid clone, T5-5, had yielded extracts in which proteins with thermo-alkaliphilic $\beta$-xylosidase activity were present, whereas fosmid clone T4-1 (source for gene 3) showed only slight $\beta$-xylosidase activity [14].

**Enzyme sizing and activity**

We scaled up the production of the gene 1, 2, 5 and 6 products in order to obtain sufficient soluble protein for further testing. The products were thus purified from large overnight *E. coli* Origami2 (DE3) pLysS cultures containing copies of intact genes 1, 2, 5 and 6 in the pET28b(+) expression vector (See Figure 2). The products of genes 1 (~116 kDa), 2 (~85 kDa), 5 (~45 kDa) and 6 (~22 kDa) had the predicted molecular sizes, as estimated from the poly acrylamide gels. The purified proteins were then further examined for GHase activity under various conditions of temperature (10 - 70°C) and pH (4.0 - 10.0), as discussed below and shown in Figure 3.

***Protein 1*** - The activity of the gene 1 encoded protein was examined with 3mM *p*NP-Gal as the substrate. It was active over a temperature range from 30 to 50°C (Figure 3a), with maximal activity at 50°C, of 68.6 U/mg. The effect of pH was then determined at 50°C (Figure 3b). Activity was observed in the pH range 4.0-10.0, with maximal activity at pH 8.0 (58.7 U/mg). This activity did not decrease much at pH 9.0 (98%) and 10.0 (78%).

***Protein 2*** - Initially, the gene 2 encoded protein was screened for enzymatic activity on *p*NP-Xyl and *p*NP-Ara (both 3mM), revealing both activities. The purified enzyme was found to be active on *p*NP-Xyl in the temperature range 40 - 60°C (Figure 3c), showing highest activity at $50^0$C, with the release of approximately 117.8 U/mg of p-nitrophenol. As in Figure 3d, activity was further observed between pH 6.0 and 10.0, with maximal activity at pH 9.0, at $50^0$C. The activity, overall, remained >60% between pH 4.0 and 10.0. On *p*NP-Ara, the enzyme showed activity (~16.8 U/mg) between 20 and $70^0$C, with maximal activity at $40^0$C (Figure 3e). We then checked the pH sensitivity of the enzyme using *p*NP-Ara at $40^0$C, and found optimal activity at pH 6.0, with approximately 16.62 U/mg of p-nitrophenol being released (Figure 3f).

***Protein 5*** - To define the optimal temperature for the activity testing of protein 5, tests were performed with *p*NP-$\alpha$-D-glucopyranoside (10 mM) at pH 7.5 and at temperatures ranging from $10^0$C to $70^0$C. An optimal temperature of $50^0$C for protein 5 activity was found (Figure 3g). Moreover, the protein also relased 30.6 U/mg of p-nitrophenol. The latter activity (at $50^0$C) was maximal at pH 10.0 (102.54 U/mg; Figure 3h).

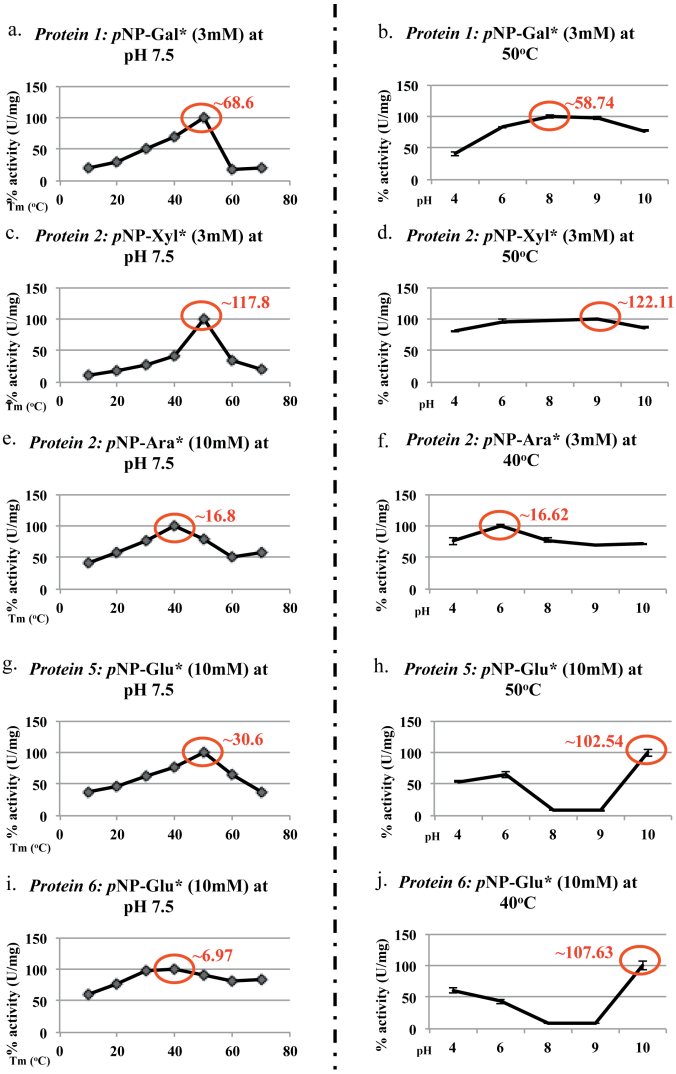



**Figure 3** Characterization of four selected candidate GHases: Enzyme activities measured against *p*NP-substrates under different temperatures with constant pH and different pH with optimum temperature. Protein 1 with *p*NP-β-D-galactopyranoside **a.** Tm 50°C; pH 7.5; 68.6 U/mg and **b.** Tm 50°C; pH 8.0; 58.74 U/mg; Protein 2 with *p*NP-β-D-xylopyranoside **c.** Tm 50°C; pH 7.5; 117.8 U/mg and **d.** Tm 50°C; pH 9.0; 122.11 U/mg; Protein 2 with *p*NP- α-L-arabinopyranoside **e.** Tm 40°C; pH 7.5; 16.8 U/mg and **f.** Tm 40°C; pH 6.0; 16.62 U/mg; Protein 5 with *p*NP-α-D glucopyranoside **g.** Tm 50°C; pH 7.5; 30.6 U/mg and **h.** Tm 50°C; pH 10.0; 102.54 U/mg; Protein 6 with *p*NP-α Dglucopyranoside **i.** Tm 40°C; pH 7.5; 6.97 U/mg and **j.** Tm 40°C; pH 10.0; 107.63 U/mg.

***Protein 6*** - The gene 6 product was tested with 10mM of *p*NP-α-D-glucopyranoside at pH 7.5 and in a temperature range of $10^0$C to $70^0$C. Protein 6 worked optimally at $40^0$C, with 6.97 U/mg of p-nitrophenol being released (Figure 3i). Furthermore, the enzyme showed maximal activity, i.e. 107.63 U/mg of p-

nitrophenol being released, at pH10.0 (Figure 3j), and this was slightly higher than the product of gene 5.

*Comparison of gene 5 and gene 6 products* - The products of genes 5 and 6 showed similar enzymatic activities on *p*NP-α-D-glucopyranoside at pH 4.0–6.0 under the same conditions. Unexpectedly, the activities of both proteins 5 and 6 dropped to near zero at pH 8.0 to 9.0 (Figure 3h and 3j). Moreover, protein 5 showed significantly higher α-glucosidase activity (30.6±0.17 U/mg) than protein 6 (6.9±0.36 U/mg). The collective results suggest that the four selected genes encode proteins with different thermo-alkaliphilic activities, as they all were optimally functional at temperatures of $40^0$C to $50^0$C and pH values of up to 10.0 (Figure 3).

The kinetic parameters of all enzymes were calculated from Lineweaver-Burke plots of specific activities at various substrate concentrations (0 -10mM). The $K_m$ and $V_{max}$ values for the enzymes Protein 1 (with *p*NP-β-D-galactopyranoside) were 0.1mM and 58.8U/mg, for Protein 2 (*p*NP-β-D-xylopyranoside and *pNP*-α-L-arabinopyranoside) showed 1.0mM, 0.3mM and 666.7U/mg, 102U/mg, respectively. The $K_m$ values of Protein 5 and Protein 6 with *p*NP-α-D-glucopyranoside were 7.4 and 21.4mM, and the $V_{max}$ were 196 and 588.2U/mg, respectively.

**Stability of the selected enzymes at elevated temperatures**

In a second set of experiments, we determined the stabilities of the enzymes encoded by genes 1, 2, 5 and 6, at elevated temperatures. The gene 1 (*p*NP-Gal), 5 (*p*NP-Glu) and 6 (*p*NP-Glu) products retained 95% of activity after over 120min of incubation at $50^0$C, after which the activities decreased - at 180min - to about 50% of the initial levels. At $50^0$C, the gene 2 product showed 100% activity with *p*NP-Xyl for 15 min, and then enzyme activity even increased, to over 180% of the control, maintaining the raised activity level until 180min. On *p*NP-Ara, the gene 2 encoded protein retained 100% activity for 30min, showing a 20% increased activity in the period between 120 to 180min (Figure 4).

**Effects of inhibitors (5-HMF and furfural) and degradation of complex polysaccharide**

Two of the major inhibitors in lignocellulose hydrolysis are 5-hydroxymethylfurfural (5-HMF) and furfural, were tested for their effects on the activity of protein 1, 2, 5 and 6 using selected concentrations of *p*NP substrates (3mM of *p*NP-Gal, *p*NP-Xyl and *p*NP-Ara; 10 mM of *p*NP-Glu) (Figure 5A). The activities of all enzymes were in most cases strongly blocked in the presence of 1.0% (w/v) 5-HMF and furfural (approximately 90 to 95%). Interestingly, the presence of 0.5% (w/v) 5-HMF inhibited all enzymes by 50-60%, whereas protein 2 with *p*NP-Ara showed nearly 100% inhibition. At the lower dosages (0.05-0.1%

w/v), 5-HMF inhibited all enzymes by 10 to 40% (Figure 5a). Next to that, the presence of 0.5% (w/v) furfural resulted in an inhibitory effect of 50 % to Protein 1 and 5, but in one close to 100% for Protein 2 (with *p*NP-Xyl and *p*NP-Ara) (Figure 5a). Notably, low levels (0.05-0.1% w/v) of furfural showed about 10-70% inhibition.

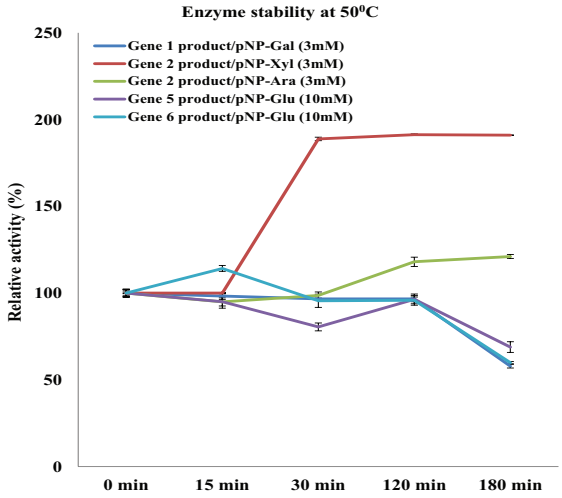
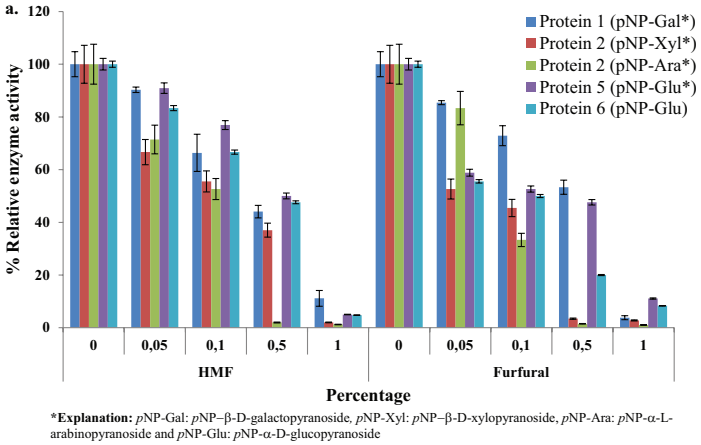



**Figure 4.** The effect of temperature on enzyme stabilities. Explanation: activities of proteins 1, 5 and 6 were stable up to 120 min (at 50 °C); protein 2 showed increased activity up to 180 min at 50 °C.

The efficiency of degradation of raw wheat straw by the four enzymes was determined by measuring the total sugars released. The data suggested that the 24 h treatment showed higher hydrolysis effect for all enzymes (Protein 1, 2, 5 and 6) when compared with the 12 h treatment (Figure 5b). The amount of total sugars was higher in the presence of Protein 6 (0.44 mg/ml) when compared with the other three proteins. Moreover, the addition of Protein 1 yielded 0.35mg/ml of total sugar. Interestingly, addition of all enzymes together (mixed) increased the 24h yield up to 0.55mg/ml (Figure 5b).

## Enzyme activities in the presence of ions

To examine the effects of ions (including NaCl) on the four selected enzymes (Table 2), activity tests under different ionic regimes were carried out with the appropriate substrates. In the presence of all $MgCl_2$, $MnCl_2$ and NaCl levels, the enzymes encoded by genes 1, 2 and 6 showed increased activities of between 10 and 25% of the control (without ions). Protein 2, on *p*NP-Ara, showed significantly increased ($P<0.05$) activity (50%) in the presence of 5mM of $Mg^{2+}$, as compared to the controls without ions, and almost 80% elevated activity with




***Explanation:** *p*NP-Gal: *p*NP–β-D-galactopyranoside, *p*NP-Xyl: *p*NP–β-D-xylopyranoside, *p*NP-Ara: *p*NP-α-L-arabinopyranoside and *p*NP-Glu: *p*NP-α-D-glucopyranoside

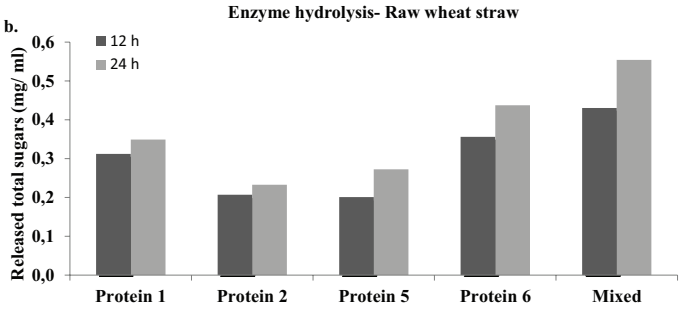



**Figure 5.** Relative activities of proteins 1, 2, 5 and 6. **a.** different concentrations (zero to 1.0% (w/v)) of furfural and 5-hydroxymethylfurfural (5-HMF); **b.** effect of lignocellulosic polysaccharide hydrolysis (raw wheat straw)

200mM of NaCl. Finally, protein 6 showed increased α-glucosidase activity (50%) in the presence of 200mM of NaCl when compared to the controls without NaCl. In fact, all additives, except 5mM of $Mg^{2+}$, significantly increased this enzymatic activity. Strikingly, the protein 5 α-glucosidase activity decreased in the presence of all additives; the activity was significantly lowered upon addition of both 5mM of $Mg^{2+}$ and 2000mM NaCl ($p<0.05$).

## Analysis of the sequences of proteins 5 and 6

***Evidence for a new α-glucosidase encoded by gene 5*** - Analyses of protein 5, using BLAST-P, multiple sequence alignment and subsequent tree building with the first 19 matches, revealed between 81 and 92% identity of the protein with proteins defined as 'diguanylate cyclase' from, respectively, *Kluyvera cryocrescens* and *Enterobacter cloaceae* (Additional Figure 1). Further characterization of the domains of protein 5 using Pfam revealed a quite complex domain architecture. Two putative conserved domains were identified, denoted as 'HAMP' and 'GGDEF'

(Figure 6). The HAMP domain has been described as a signaling mediator, being mostly found in histidine kinases, adenyl cyclases, methyl-accepting proteins and phosphatases [23]. HAMP domains consist of 16-residue amphiphilic helices that are often part of a two-component signal transduction pathway [24]. They can be found in association with other domains, such as the GGDEF and EAL domains. Remarkably, gene 5 is here described as a gene encoding a glycoside hydrolase family GH53 protein (analysis by the CAZymes analysis toolkit (CAT) server), with predicted endo-β-1, 4-galactanase activity based on the (catalytic) GGDEF domain. Therefore, we here posit that protein 5 is a GGDEF family protein, which is part of a signal-responsive system (given its HAMP domain), with $\alpha$-glucosidase activity as shown with *p*NP-$\alpha$-*D*-glucopyranoside.

**Table 2:** Effects of metal ions and NaCl on the enzymatic activities of recombinant enzymes (each treatment had three biological replicates) [a]Substrate: refer to Fig. 5a

| Enzymes (product of:) | Substrate[a] | No additive | Relative activities (%) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $MgCl_2$ | | $MnCl_2$ | | NaCl | |
| | | | 5mM | 20mM | 5mM | 20mM | 200mM | 2000mM |
| Gene 1 | *p*NP-Gal | 100± 1.8 | 113.1± 9.1 | 114.1± 9.6 | 97.7± 6.5 | 95.5± 9.2 | 91.7± 3.5 | 90.9± 3.6 |
| Gene 2 | *p*NP-Xyl | 100± 4.9 | 124.5± 0.7 | 129.9± 4.9 | 108.6± 4.4 | 105.0± 3.2 | 112.4± 12.4 | 118.6± 9.3 |
| Gene 2 | *p*NP-Ara | 100± 17.6 | 150.5± 19.4 | 106.4± 26.0 | 135.5± 7.6 | 81.6± 1.2 | 183.9± 53.1 | 75.1± 1.9 |
| Gene 5 | *p*NP-Glu | 100± 2.7 | 85.4± 3.4 | 94.0± 6.3 | 95.6± 2.3 | 88.7± 1.0 | 102.7± 3.8 | 59.9± 2.4 |
| Gene 6 | *p*NP-Glu | 100± 4.1 | 105.5± 1.4 | 112.8± 5.5 | 92.8± 2.3 | 111.1± 1.0 | 159.5± 3.6 | 119.3± 5.1 |

Furthermore, we analyzed the protein 5 amino acid sequence based on domain architecture and revealed potential "GDSL" motifs (by CDART); the presence of such a GDSL motif may indicate the potential for multiple functional properties, such as broad substrate specificity, given its active site flexibility. Due to this, multiple activities, such as lipase and esterase activities, could be present, which are of use in hydrolysis processes of biological interest [25]. Further 3D modeling of protein 5 (Figure 8a) indicated the presence of hypothetical ligands of both the GGDEF and GDSL domains that may allow binding of a particular substrate, supporting a catalytic mechanism.

### Evidence for a new α-glucosidase encoded by gene 6

In the case of the gene 6 product, a BLAST-P search initially showed high similarity of it to typical porin-like proteins, with up to 97% identity. Specifically, the predicted protein clustered with porins from the genera *Marterella endophytica* (80%) and *Devosia sp. root635* (97%), which are both Alphaproteobacteria from the order Rhizobiales (Additional Figure 2). Connected to this, the characterization of the different domains in the predicted protein yielded a match with

transmembrane-channel-forming proteins of the major intrinsic protein (MIP) family from *Hyphomonas johnsonii* and *Brevundimonas diminuta* . The predicted protein 6 was found to contain the characteristic Asn-Pro-Ala (denoted as NPA) signature motif located at the segment interface between helices M3 and M7 in residues 103-105 (Figure 7, in yellow). In our previous study, the source fosmid clone NT18-17 produced proteins with α-glucosidase activity. Here, the product of the cloned gene 6 showed 100% increased α-glucosidase activity at pH 10.0 (Figure 3j). Finally, protein 6 was analyzed with respect to domain architecture. A potential "GDSL" motif was also found by the CDART software, the potential function and characterization of this motif being as above. Using 3D modeling (Figure 8b), hypothetical substrate-binding ligands were identified, supporting the tenet that a catalytic site is present in this presumed trans-membrane protein.

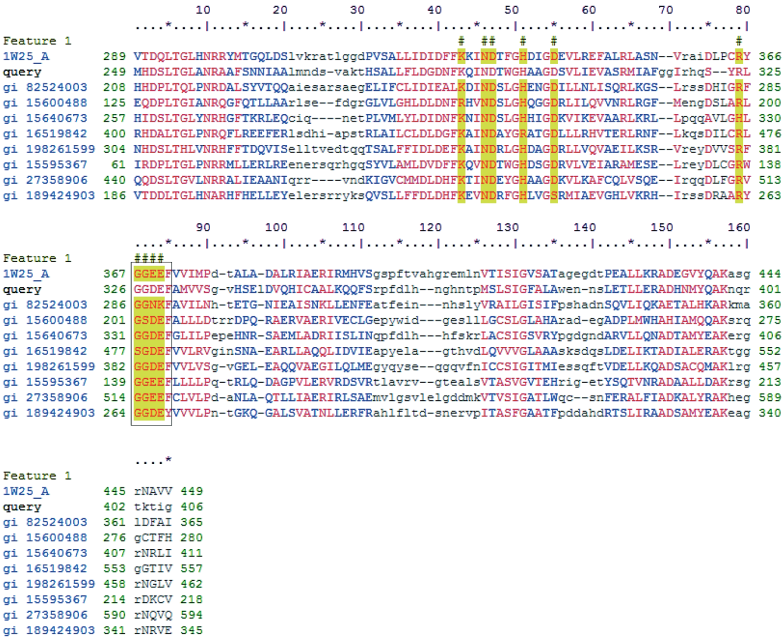


**Figure 6.** Alignment of amino acid sequences revealing that protein 5 contains the (typical for class III nucleotidyl cyclases [37]) conserved residues K 290: lysine, N 293: asparagine, D 294: aspartic acid, H 298: histidine, D 302: aspartic acid, and a "GGDEF" domain (R 324: arginine, G 326: glycine, G 327: glycine, D 328: aspartic acid, E 329: glutamic acid and F 330: phenylalanine). 1W25_A diguanylate cyclase [*Caulobacter vibrioides*]; gi 82524003-Hypothetical protein [uncultured gamma *proteobacterium*]; gi 15600488-hypothetical protein PA5295 [*Pseudomonas aeruginosa* PAO1]; gi 15640673- c-di-GMP phosphodiesterase A-like protein [*Vibrio cholerae*]; gi 16519842-diguanylate cyclase/phosphodiesterase [*Sinorhizobium fredii* NGR234]; gi 198261599-Sensory box/ggdef domain/eal domain protein [gamma *proteobacterium* HTCC5015]; gi 15595367-Hypothetical protein PA0169 [*Pseudomonas aeruginosa* PAO1]; gi 27358906-FOG: GGDEF domain protein [*Vibrio vulnificus* CMCP6]; gi 189424903-GAF sensor-containing diguanylate cyclase [*Geobacter lovleyi* SZ]

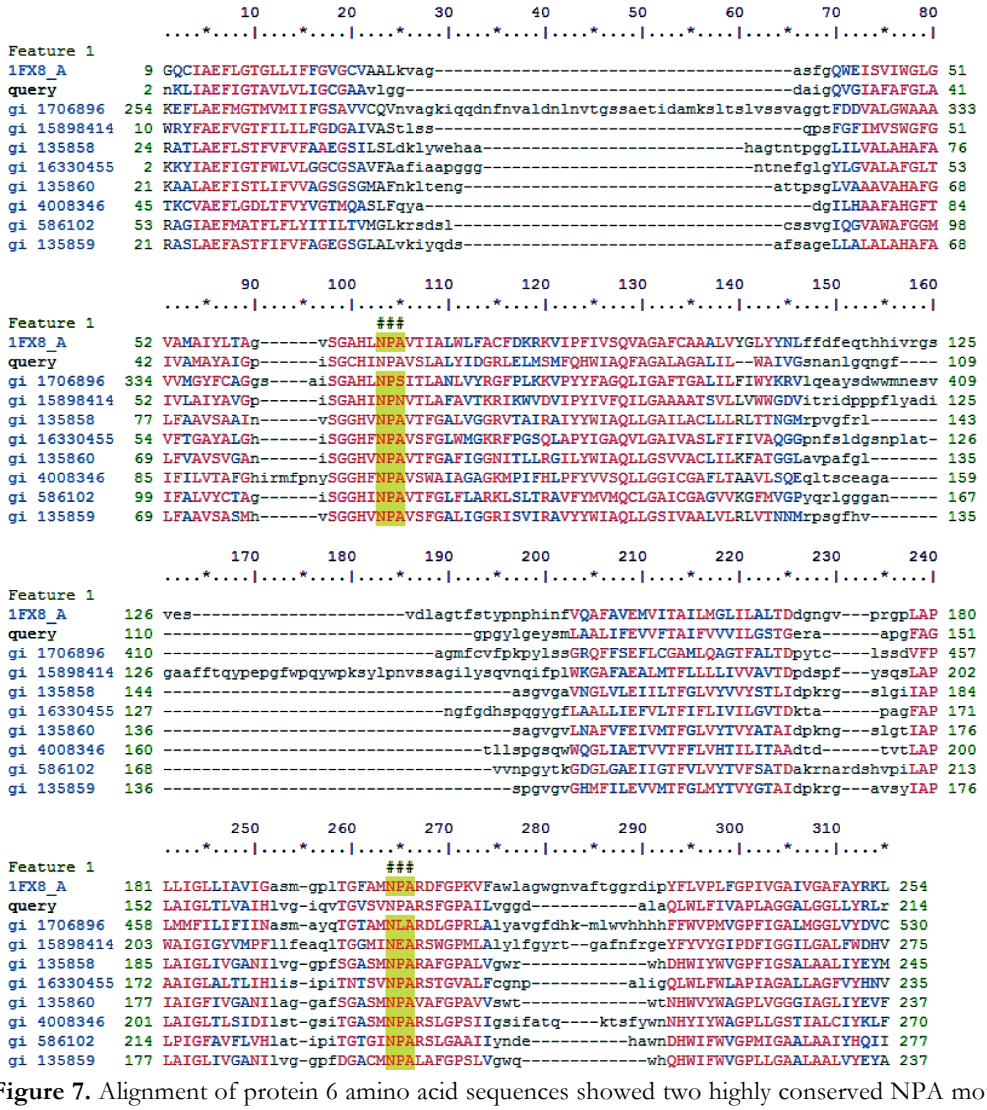



**Figure 7.** Alignment of protein 6 amino acid sequences showed two highly conserved NPA motifs (asparagine–proline–alanine; highlighted yellow). 1FX8_A-Membrane protein [*Escherichia coli*]; gi 1706896-Glycerol uptake/efflux facilitator protein [*Saccharomyces cerevisiae* S288C]; gi 15898414-transposon ISC1229 Orf1 [*Sulfolobus solfataricus* P2]; gi 135858-Aquaporin TIP3-1 [*Arabidopsis thaliana*]; gi 16330455-Aquaporin Z [*Synechocystis* sp. PCC 6803]; gi 135860 Aquaporin TIP1-1 [*Arabidopsis thaliana*]; gi 4008346-Major intrinsic protein [*Caenorhabditis elegans*]; gi 586102-Membrane protein [*Solanum lycopersicum*]; gi 135859-Aquaporin TIP-type alpha [*Phaseolus vulgaris*]

## Discussion

Metagenomics constitutes a true 'breakthrough tool' that allows to examine natural and manipulated microbiomes for the presence of genes for novel enzymes that may fuel biomedical and industrial applications [26]. In a previous study involving two wheat-straw-degrading microbial consortia, we identified two genes encoding proteins with putative thermo-alkaliphilic activities[14]. With respect to activity (β-

galactosidase and β-xylosidase), these were stable at alkaline pH and elevated temperature. Hence, they may have interesting potential applications in the process of enzyme-assisted pulp bleaching [14]. In the current study, we further mined our fosmids that were predicted to encode such GHases. Eight thus selected genes were cloned into two *E. coli* host strains, BL21(DE3) and Origami2 (DE3) pLysS, and gene expression was examined. Strain BL21(DE3) is useful for gene expression studies from a T7-type promoter on the vector. In our study, BL21(DE3) cell growth was arrested, possibly as a result of the emergence of inhibitory compounds, i. e. the intended proteins, in the culture. In the case of *E. coli* Origami2 (DE3) pLysS, the T7 lysozyme produced suppresses the basal expression of T7 RNA polymerase prior to induction. Thus, we found *E. coli* Origami2 (DE3) pLysS to produce more proteins in the soluble fractions than BL21(DE3), giving rise to five (genes 1, 2, 3, 5 and 6) measurable enzyme activities. In our previous work[14], fosmid clone 10BT extracts showed activity with multiple substrates and also light (14.8%) β-xylosidase activity. There was no detectable α-glucosidase activity. Interestingly, protein 5 (originating from clone 10BT) showed α-glucosidase activity from the strong inducible promoter of the vector. Thus, presumably the native promoter of this gene did not allow for expression under such conditions. Concerning genes 7 and 8, no activity was found with any of the test substrates, of which we ignore the cause. However, problems of instability and/or aggregation (under the prevailing conditions of osmolarity, pH, redox potential, cofactors) may have played key roles.

Based on the initial screening, we thus considered the selected five genes for further study of their products. Among these, proteins 2 and 3 were similar in activity and size. With reference to our previous study, here we selected gene 2 instead of gene 3 (from fosmid clone T5-5) based on the high thermo-alkaliphilic activity with *p*NP-Xyl, whereas the gene 3 mother fosmid clone T4-1 had shown only slight activity against *p*NP-Xyl [14].

### *Protein 1*

With respect to the identification of protein 1 as an EC. 3.2.1.23 (GH2), several new β-galactosidase family-GH2 enzymes have recently been discovered and characterized [27–29]. Specifically, a purified (β-galactosidase) family-GH2 enzyme, denoted BglA, from *Arthrobacter psychrolactophilus F2* showed maximal enzymatic activity (alkaliphilic) at pH 8.0/ 10 °C (33.3 U/mg). Such β-galactosidases are widely used in the food industry given their capabilities to hydrolyse lactose at extreme (low or high) temperatures, and to produce glucose and galactose. Remarkably, our gene 1 encoded protein showed increased enzymatic activity at raised temperature and similar pH when compared to the BglA enzyme, with increased p-nitrophenol release (58.7 U/mg; Table 3). Moreover, protein 1 showed a lower $K_m$ and slightly increased $V_{max}$ as compared to those of aforementioned *A. psychrolactophilus* enzyme, as well as a *Lactobacillus sakei*

Lb790 enzyme [27,29]. Therefore, protein 1 shows promising hydrolytic activity, which may be of direct relevance to industrial processes.

**Table 3:** Comparison with enzymes (GH2 and GH3/43) from other studies (reactions with the substrate ONPG and/or PNPG)

| Enzyme family | Functions | Strain | Protein (kDa) | Optimum Tm (°C) | Optimum pH | Activity (U/ mg) | Reference |
|---|---|---|---|---|---|---|---|
| GH 2 | β-galactosidase | *Arthrobacter psychrolactophilusF2* | 130 | 10 | 8 | 33.3 | Nakagawa et al. (2006) [27] |
| GH 2 | β-galactosidase | *Arthrobacter sp. 20B* | 113.7 | 25 | 6–8 | 0.84 | Bialkowska et al. (2009) [56] |
| GH 2 | β-galactosidase | *Arthrobacter SB* | 114 | 18 | 7 | 26.9 | Coker and Brenchley (2006)[57] |
| GH 2 | β-galactosidase | *Flavobacterium sp. 4214* | 114.3 | 42 | 7.5 | - | Sorensen et al. (2006) [58] |
| GH 2 | β-galactosidase | *Halomonas sp. S62* | 63 | 45 | 7 | - | Wang et al. (2013) [59] |
| GH 2 | β-galactosidase | *Enterobacter hormaechei* | 120 | 50 | 8 | 58.7 | This study |
| GH 3 | Dual function (β-xylosidase/α-arabinofuranosidase) | *Rumen/Parabacteroides distasonis ATCC 8503* | 80 | 50 | 6 | 29.5 | Zhou et al. (2012) [30] |
| GH 43 | Dual function (β-xylosidase/α-arabinofuranosidase) | *Rumen/Prevotella bryantii* | 45 | 40 | 7 | 36.3/ 14.2 | Zhou et al. (2012) [31] |
| GH3- BglX | Dual function (β-xylosidase/α-arabinofuranosidase) | *C. crescentus* | 90 | 60/50 | 6 | 4.3/ 3.0 | Justo et al. (2015) [32] |
| GH 3 | Dual function (β-xylosidase/α-arabinofuranosidase) | *Enterobacter mori* | 80 | 50/40 | 9 and 6 | 122.11/ 16.6 | This study |

*Protein 2*

Gene 2 was annotated as a gene encoding a protein with β-xylosidase/α-arabinosidase activity belonging to CAZy family GH3 (EC. 3.2.1.37). As shown in Table 3, the enzyme had raised activities on *p*NP-Xyl and *p*NP-Ara, with optimum pH values of 9.0 (and 6.0) at temperatures of 50 °C and 40 °C. This was superior as compared to the previously-reported enzymes Rubgx1, GH43 (several) and XynB5. Specifically, the Rubgx1 (GH3 family) enzyme (β-glucosidase/β-xylosidase activities; optima at pH 6.0 and 50°C) [30], the GH43 family one (β-xylosidase/α-arabinosidase; optima at pH 7.0 and 40°C) [31] and, finally, XynB5 (β-glucosidase/β-xylosidase/α-arabinosidase; optimum pH 6.0 and temperature 50°C) had lower activities than protein 2 [32]. In addition, protein 2 revealed high kinetic parameters, such as $V_{max}$ (666.6 and 102U/mg), than a previously-characterized bifunctional (β-xylosidase/α-arabinosidase) enzyme [31]. In another study, we recently characterized GH43 family enzyme XylM1989 with β-xylosidase/α-arabinosidase activity and showed $V_{max}$ values of 285.71/78.12 U/mg, respectively (unpublished data). Also, protein 2 progressively showed higher enzyme activity (with the corresponding substrates) with increasing pH. Such changes in pH will alter the attractions between groups in the side chains of the protein, potentially modifying the protein domain shape. Moreover, the binding of substrate to the active site may be modulated and/or it cannot undergo catalysis.

As indicated above, protein 2 was similar to the GH3 family proteins Rubgx1 (81.4kDa) and XynB5 (95kDa, from *Caulobacter crescentus*) [30,32]. Its *p*NP-Xyl and *p*NP-Ara dual activity may have been due to two distinct catalytic properties in the same polypeptide chain, usually catalyzing two consecutive reactions [33]. Dual- or even multi-functional properties are common in proteins of the GH3 (and GH43) family. For instance, the GH3 family protein XynB5 showed rather similar β-glucosidase/β-xylosidase/α-arabinosidase activities [32]. Moreover, GH3 family enzymes include β-xylosidases (EC 3.2.1.37) and α-arabinosidases (EC 3.2.1.55). Whereas the aforementioned enzymes Rubgx1 and XynB5 showed (β-xylosidase) activities with optimum pH and temperature of 6.0 and 50 °C, respectively, protein 2 showed maximal activity at pH 9.0 and 50°C (122.11U/mg) (Figure 3d). This preference for alkaline instead of slightly acid conditions is likely connected to amino acid substitutions that we have-as yet-not addressed. The mechanisms by which an enzyme's catalytic properties may be affected were beyond the scope of the current study [34].

*Protein 5*

Interestingly, protein 5 showed α-glucosidase activity, whereas no such activity was detected from its mother fosmid clone 10BT[14]. Fosmid clone 10BT had shown activity in the presence of multiple substrates, revealing some activity with *p*NP-

Xyl. Thus, expression of the α-glucosidase encoding gene may have been repressed at the genomic level, becoming expressed from the inducible promoter of the expression system of the used vector. Protein 5 clearly belongs to the "GGDEF" domain protein family [35,36]. This domain was first identified in a response regulator involved in cell differentiation in *C. crescentus* [37]. It was also observed in *Salmonella enterica* enzymes involved in cellulose biosynthesis and biofilm formation [38]. For more than 20 years, all GGDEF domain enzymes have been classed as diguanylate cyclases and/or phosphodiesterases [39]. The former enzymes produce cyclic di-GMP (cdiG), a messenger that regulates the key bacterial lifestyle transition from a motile to a sessile, biofilm-forming, state [40]. However, most bacteria are known to possess large numbers of genes that encode a range of GGDEF domain proteins [38], allowing functional diversity across them. The function of most GGDEF domain containing proteins has not yet been experimentally proven [41]. Surprisingly, the GGDEF domain family protein 5 revealed α-glucosidase activity (in the context of (hemi)cellulose degradation), with a 100% raised activity at pH 10.0 (102.54 U/mg) as compared to pH 7.0. At pH 10, the large $OH^-$ oversupply may have caused a change in the shape and/or charge of the enzyme's active site, spurring activity. Further detailed structural studies could reveal the dynamics of such glycoside hydrolase activities at varying pH values in the lignocellulose degradation processes. Also, protein 5 was found to have a different motif, of the "GDSL" class by CDART prediction.

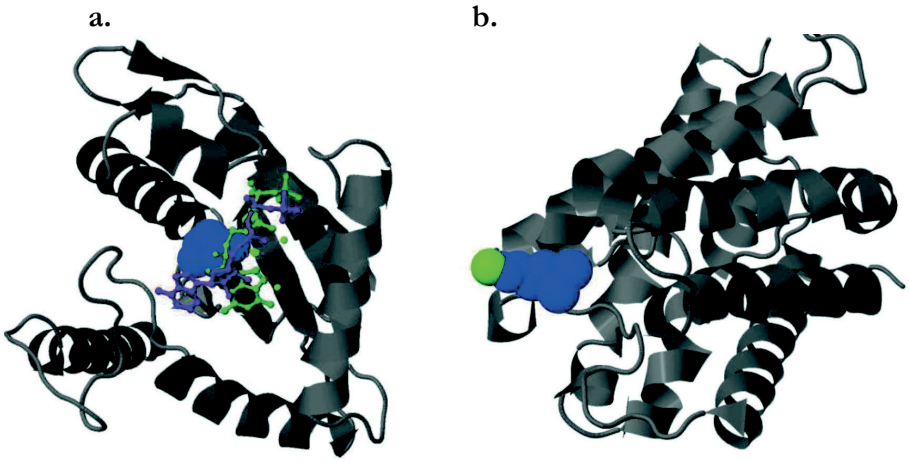



**Figure 8.** Three-dimensional (3D) models of **a.** protein 5 and **b.** protein 6. Background helices shown in gray, Active sites shown in blue, Hypothetical ligands shown in magenta and green

### *Protein 6*

Remarkably, protein 6 was found to possess two different conserved motifs, NPA and GDSL. The NPA motif is a key structural feature of proteins that play crucial roles as water channels across membranes, supporting membrane localization of the protein. Surprisingly, protein 6 also belongs to the GDSL hydrolase family

(encompassing esterases/lipases), not sharing sequence homology with any of the CAZy database glycosyl hydrolases. GDSL motif enzymes constitute a rather new group of proteins, with characteristics that have not yet been precisely described [42]. Furthermore, a subgroup of this GDSL family was categorized as a so-called SGNH-hydrolase, and our protein 6 belongs to this subgroup. The SGNH family of hydrolytic enzymes has a wide range of catalytic functions, such as lipase, protease, carbohydrate esterase, (thio)esterase, arylesterase and acyltransferase activities [43]. All GDSL hydrolase family enzymes were found to have flexible substrate binding and/or active sites [43]. Indeed, the Koshland induced-fit theory indicates that the active sites of enzymes may become modified in the presence of substrates, involving structural and catalytic site modifications[44]. The active and/or binding sites of protein 6 are presumably flexible if they are to follow the induced-fit theory. For instance, the newly-identified carbohydrate esterase family 3 (CE3) gene *axe2* product (acetyl xylan esterase), which has a GDSL motif, removes acetyl groups from the hemicellulose polymer xylan [45]. Such activity, i.e. carbohydrate esterase, likely gives protein 6 the capability to hydrolyze substituents on the xylan backbone, supporting its ability to competently degrade hemicellulose [46]. Indeed, protein 6 (and 5), based on an SGNH-type hydrolase, may exhibit the specific α-glucosidase activity in (hemi)cellulose degradation.

In recent years, several enzyme cocktails have been proposed that can enhance plant biomass degradation rates [47]. Such cocktails were produced from several sources, including the producer organisms *Trichoderma reesei, Thermobifida fusca* and *Clostridium thermocellum*. The degradation of lignocellulose requires the intensive action of multiple enzymes. This is due to structural aspects that are typical for the substrate, i.e. its high molecular weight and the association of hemicellulose with cellulose and lignin. The required hemicellulases need to hydrolyze glycosidic bonds. In many habitats, the degradation process is rather slow because of the poor substrate accessibility, inhibitors (5-HMF and furfural), and untimely presence of efficient enzymes. However, in industrial processes, such enzyme availabilities can be steered [48,49]. The inhibitory by-products 5-HMF and furfural are released during lignocellulose degradation, constitute a major inhibitors of subsequent fermentation [50,51]. However, the amount of 5-HMF and furfural in plant biomass hydrolysates is strongly linked to the pretreatment methods and the source of the plant biomass. In this line, the levels of 5-HMF and furfural in corn, poplar and pine were 0.017% and 0.022% (w/v), respectively [52]. Interestingly, the four enzymes (Protein 1, 2, 5 and 6) showed 50-80% tolerance to 0.1% 5-HMF, and about 40-50% tolerance at 0.5% (w/v). In addition, all enzymes also revealed tolerance (30-72%) to furfural. Remarkably, protein 5 (with α-glucosidase activity) revealed 50% activity at 0.5% of 5-HMF and furfural. Therefore, we posit that such inhibitor-tolerant galactosidases and glucosidases have great interest in food and biorefinery industries. Because of this strong tolerance, our enzymes may work better than other ones in the presence of certain

levels of inhibitors. As of now, no previous studies reported 5-HMF- and furfural-tolerant β-galactosidases and α-glucosidases.

The four (thermostable) enzymes described here may pave the way towards improved (hemi) cellulolytic action of existing enzyme cocktails. Using the his tags, the enzymes, provided they are sufficiently stable, can be easily recovered from the treatment process and re-used in continuous applications, allowing alkaline and enhanced temperature conditions [53]. The here-detected activity of our enzymes (specifically proteins 1 and 2) under thermo-alkaliphilic conditions thus indicates their usefulness for plant biomass degradation, spurring the production of glucose and xylose/arabinose from (hemi)cellulosic material and also for pulp bleaching processes. Moreover, thermo-stable enzymes like the ones reported here are often specific for activity on particular bonds and therefore offer promise for specific industrial purposes [54,55]. Several amylases and cellulases are already in use in industries, including enzymes with wide pH range [4.0-10.0] and high thermo-tolerance. They are clearly useful under conditions that restrict microbial growth. Furthermore, enzymes with optimized properties, such as enhanced thermal and alkali tolerance, and an ability to function without additives, recently improved several industrial applications [51,52]. In a subsequent study, we will characterize the role of the novel enzymes as modulators of the degradation of hemicellulose compounds.

## Authors' contributions

MM designed, constructed the experiments and drafted the manuscript. JDvE conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.


## Acknowledgements

We thank Maryam Chaib de Mares (for bioinformatics), Diego Javier Jiménez and Patricia Stevens for the support and suggestions for this study. Further thanks are due to Harald Ruijssenaars and Richard van Kranenburg for scientific support. This work was supported by the Netherlands Ministry of Economic Affairs and the BE-Basic partner organizations (http://www.be-basic.nl).


## References


1. Sørensen A, Lübeck M, Lübeck PS, Ahring BK. Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials. Biomolecules. 2013;3:612–631.
2. Gray KA, Zhao L, Emptage M. Bioethanol. Curr Opin Chem Biol. 2006;10:141–146.
3. Dougherty MJ, D'haeseleer P, Hazen TC, Simmons BA, Adams PD, Hadi MZ. Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. BMC Biotechnol. 2012;12:38.
4. Turner P, Mamo G, Karlsson E. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact. 2007;6:9.
5. Sims REH, Mabee W, Saddler JN, Taylor M. An overview of second generation biofuel technologies. Bioresour Technol. 2010;101:1570–1580.

6. Mohanram S, Amat D, Choudhary J, Arora A, Nain L, Bringezu S, et al. Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries. Sustain Chem Process. 2013;1:15.
7. Gowen CM, Fong SS. Exploring biodiversity for cellulosic biofuel production. Chem Biodivers. 2010;7:1086–1097.
8. 8. Bayer EA, Chanzy H, Lamed R, Shoham Y. Cellulose, cellulases and cellulosomes. Curr Opin Struct Biol. 1998;8:548–557.
9. Ljungdahl LG. The cellulase/hemicellulase system of the anaerobic fungus *Orpinomyces* PC-2 and aspects of its applied use. Ann N Y Acad Sci. 2008;1125:308–321.
10. Wongwilaiwalin S, Rattanachomsri U, Laothanachareon T, Eurwilaichitr L, Igarashi Y, Champreda V. Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system. Enzyme Microb Technol. 2010;47:283–290.
11. Wang W, Yan L, Cui Z, Gao Y, Wang Y, Jing R. Characterization of a microbial consortium capable of degrading lignocellulose. Bioresour Technol. 2011;102:9321–9324.
12. Jiménez DJ, Korenblum E, Van Elsas JD. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl Microbiol Biotechnol. 2014;98:2789–2803.
13. Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnol Bioeng. 2012;109:1083–1087.
14. Maruthamuthu M, Jiménez DJ, Stevens P, Van Elsas JD. A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. BMC Genomics. 2016;17:86.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–410.
16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2015;44:D279-285.
17. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics. 2007;23:1073–1079.
18. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res. 2002;12:1619–1623.
19. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the phyre server. Nature Protoc. 2009;4:363–371.
20. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010;38:W545-549.
21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000;28:235–242.
22. Wulff NA, Carrer H, Pascholati SF. Expression and purification of cellulase Xf818 from *Xylella fastidiosa* in *Escherichia coli*. Curr Microbiol. 2006;53:198–203.
23. Aravind L, Ponting CP. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett. 1999;176:111–116.
24. Parkinson JS. Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. Annu Rev Microbiol. 2010;64:101–122.
25. Messaoudi A, Belguith H, Gram I, Hamida J Ben. Classification of EC 3.1.1.3 bacterial true lipases using phylogenetic analysis. African J Biotechnol. 2010;9:8243–8247.
26. Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol. 1991;173:4371–4378.
27. Nakagawa T, Fujimoto Y, Ikehata R, Miyaji T, Tomizuka N. Purification and molecular characterization of cold-active beta-galactosidase from *Arthrobacter psychrolactophilus* strain F2. Appl Microbiol Biotechnol. 2006;72:720–725.

28. Schwab C, Sørensen KI, Gänzle MG. Heterologous expression of glycoside hydrolase family 2 and 42 β-galactosidases of lactic acid bacteria in *Lactococcus lactis*. Syst Appl Microbiol. 2010;33:300–307.
29. Iqbal S, Nguyen T-H, Nguyen HA, Nguyen TT, Maischberger T, Kittl R, et al. Characterization of a heterodimeric GH2 β-galactosidase from *Lactobacillus sakei* Lb790 and formation of prebiotic galacto-oligosaccharides. J Agric Food Chem. 2011;59:3803–3811.
30. 3Zhou J, Bao L, Chang L, Liu Z, You C, Lu H. Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. Lett Appl Microbiol. 2012;54:79–87.
31. Zhou J, Bao L, Chang L, Zhou Y, Lu H. Biochemical and kinetic characterization of GH43 β-d-xylosidase/α-l-arabinofuranosidase and GH30 α-l-arabinofuranosidase/β-d-xylosidase from rumen metagenome. J Ind Microbiol Biotechnol. 2012;39:143–152.
32. Justo PI, Corrêa JM, Maller A, Kadowaki MK, da Conceição-Silva JL, Gandra RF, et al. Analysis of the xynB5 gene encoding a multifunctional GH3-BglX β-glucosidase-β-xylosidase-α-arabinosidase member in *Caulobacter crescentus*. Antonie Van Leeuwenhoek. 2015;108:993–1007.
33. Vrzheshch P V. Steady-state kinetics of bifunctional enzymes. Taking into account kinetic hierarchy of fast and slow catalytic cycles in a generalized model. Biochem Biokhimiia. 2007;72:936–943.
34. Tijskens LM, Greiner R, Biekman ES, Konietzny U. Modeling the effect of temperature and pH on activity of enzymes: the case of phytases. Biotechnol Bioeng. 2001;72:323–330.
35. Cotter PA, Stibitz S. c-di-GMP-mediated regulation of virulence and biofilm formation. Curr Opin Microbiol. 2007;10:17–23.
36. Sinha SC, Sprang SR. Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases. Rev Physiol Biochem Pharmacol. 2006;157:105–140.
37. Hecht GB, Newton A. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. J Bacteriol. 1995;177:6223–6229.
38. García B, Latasa C, Solano C, García-del Portillo F, Gamazo C, Lasa I. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol Microbiol. 2004;54:264–277.
39. Römling U, Galperin MY, Gomelsky M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev. 2013;77:1–52.
40. Hallberg ZF, Wang XC, Wright TA, Nan B, Ad O, Yeo J, et al. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3′, 3′-cGAMP). Proc Natl Acad Sci. 2016;113:1790–1795.
41. Galperin MY, Nikolskaya AN, Koonin E V. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 2001;203:11–21.
42. Upton C, Buckley JT. A new family of lipolytic enzymes? Trends Biochem Sci. 1995;20:178–179.
43. Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F. GDSL family of serine esterases/lipases. Prog Lipid Res. 2004;43:534–552.
44. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. Proc. National Academy of Sciences. 1958;44:98–104.
45. Alalouf O, Balazs Y, Volkinshtein M, Grimpel Y, Shoham G, Shoham Y. A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from *Geobacillus stearothermophilus*. J Biol Chem. 2011;286:41993–42001.
46. Kabel MA, Yeoman CJ, Han Y, Dodd D, Abbas CA, De Bont JAM, et al. Biochemical characterization and relative expression levels of multiple carbohydrate esterases of the xylanolytic rumen bacterium *Prevotella ruminicola* 23 grown on an ester-enriched substrate. Appl Environ Microbiol. 2011;77:5671–5681.
47. Cherry JR, Fidantsef AL. Directed evolution of industrial enzymes: an update. Curr Opin

Biotechnol. 2003;14:438–443.
48. Subramaniyan S, Prema P. Biotechnology of microbial xylanases: enzymology, molecular biology, and application. Crit Rev Biotechnol. 2002;22:33–64.
49. Beg QK, Kapoor M, Mahajan L, Hoondal GS. Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol. 2001;56:326–338.
50. Garrote G, Domínguez H, Parajó JC. Hydrothermal processing of lignocellulosic materials. Holz als Roh- und Werkstoff. 1999;57(3):191–202
51. Van der Pol EC, Bakker RR, Baets P, Eggink G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl Microbiol Biotechnol. 2014;9579–9593.
52. Du B, Sharma LN, Becker C, Chen S-F, Mowery RA, Van Walsum GP, et al. Effect of varying feedstock-pretreatment chemistry combinations on the formation and accumulation of potentially inhibitory degradation products in biomass hydrolysates. Biotechnol Bioeng. 2010;107:430–440.
53. Fischer F, Mutschler J, Zufferey D. Enzyme catalysis with small ionic liquid quantities. J Ind Microbiol Biotechnol. 2011;38:477–487.
54. Vester JK, Glaring MA, Stougaard P. Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb Cell Fact. 2014;13:72.
55. Haki G. Developments in industrially important thermostable enzymes: a review. Bioresour Technol. 2003;89:17–34.
56. Białkowska AM, Cieśliński H, Nowakowska KM, Kur J, Turkiewicz M. A new β-galactosidase with a low temperature optimum isolated from the *Antarctic Arthrobacter* sp. 20B: gene cloning, purification and characterization. Arch Microbiol. 2009;191:825–835.
57. Coker JA, Brenchley JE. Protein engineering of a cold-active β-galactosidase from *Arthrobacter* sp. SB to increase lactose hydrolysis reveals new sites affecting low temperature activity. Extremophiles. 2006;10:515–524.
58. Sørensen HP, Porsgaard TK, Kahn RA, Stougaard P, Mortensen KK, Johnsen MG. Secreted β-galactosidase from a *Flavobacterium* sp. Isolated from a low-temperature environment. Appl Microbiol Biotechnol. 2006;70:548–557.
59. Wang G, Gao Y, Hu B, Lu X, Liu X, Jiao B. A novel cold-adapted β-galactosidase isolated from *Halomonas* sp. S62: gene cloning, purification and enzymatic characterization. World J Microbiol Biotechnol. 2013;29:1473–1480.


Figure S1. Protein 5 Blast-P multiple sequence alignment

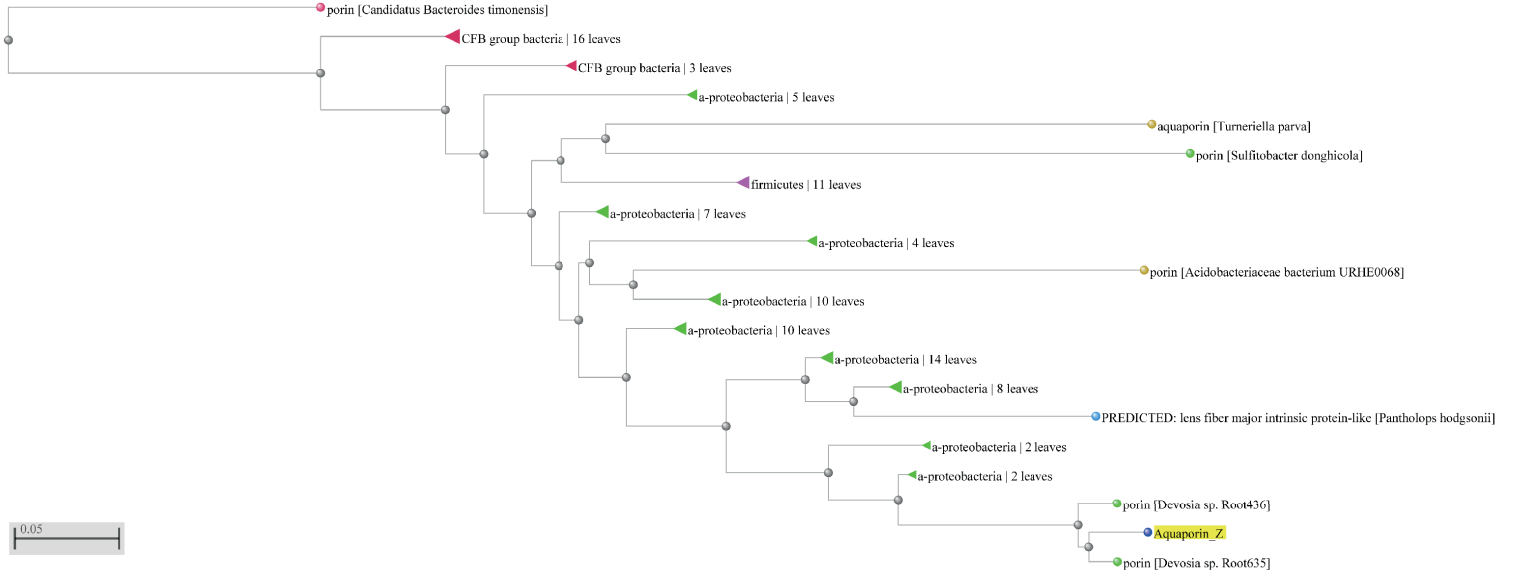



Figure S2. Protein 6 Blast-P multiple sequence alignment.

# Chapter 4

## Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach.

**Mukil Maruthamuthu[*], Diego Javier Jiménez[*], and Jan Dirk van Elsas.**
*[*]Equal contribution*

*J Appl Microbiol.* 2017 May 10. doi: 10.1111/jam.13484.

**Abstract**

**Aim:** The aim of the study was to characterize ten hemicellulolytic enzymes obtained from a wheat straw-degrading microbial consortium.

**Methods and results:** Based on previous metagenomics analyses, ten glycosyl hydrolases were selected, codon-optimized, synthetized, cloned and expressed in *Escherichia coli*. Nine of the overexpressed recombinant proteins accumulated in cellular inclusion bodies, whereas one, a 37.5 kDa protein encoded by gene *xyl*M1989, was found in the soluble fractions. The resulting protein, denoted XylM1989, showed β-xylosidase and α-arabinosidase activities. It fell in the GH43 family and resembled a *Sphingobacterium* sp. protein. The XylM1989 showed optimum activity at 20°C and pH 8.0. Interestingly, it kept approximately 80% of its β-xylosidase activity in the presence of 0.5% (w/v) furfural and 0.1% (w/v) 5-hydroxymethylfurfural. Additionally, the presence of $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$ ions increased the enzymatic activity and conferred complete tolerance to 500 mmol $l^{-1}$ of xylose. Protein XylM1989 is also able to release sugars from complex polysaccharides.

**Conclusion:** We report the characterization of a novel bifunctional hemicellulolytic enzyme obtained through a targeted synthetic metagenomics approach.

**Significance and impact of the study:** The properties of XylM1989 turn this protein into a promising enzyme that could be useful for the efficient saccharification of plant biomass.

## Introduction

Plant-derived lignocellulose represents an abundantly available and renewable energy source. Lignocellulose comprises cellulose, hemicellulose and lignin moieties. Hemicellulose consists of hetero-polymers that are composed of pentoses and hexoses. In this fraction, xylan is the major component, constituting nearly one third of all renewable carbon in nature. Xylan (or arabinoxylan) is composed of β-1,4-linked D-xylose units, which may be substituted by different side groups, such as D-galactose, L-arabinose, glucuronic acid, acetyl, feruloyl and p-coumaroyl residues [1,2]. In the enzyme-mediated catalysis that is required for hemicellulose degradation, microbial glycoside hydrolases (GHs) are key enzymes. These cleave the glycosidic linkages between carbohydrate residues, allowing to produce sugars that are released. Xylan can be degraded through the action of a set of different GHs. For instance, β-1,4-endoxylanase (EC 3.2.1.8), which cleaves the backbone into small oligosaccharides, and β-1,4-xylosidase (EC 3.2.1.37), which cleaves these oligosaccharides into xylose. Next to breaking the side chains of xylan, enzymes like α-L-arabinosidases, α-D-glucuronidases and acetyl esterases can play vital roles [3–5]. Such enzymes are thought to be valuable for diverse industrial (e.g. food, pharma, plastics and biofuels) applications [6].

In spite of the promise of using microorganisms from natural settings as sources of novel GHs, the discovery of novel enzymes or activities has been hampered by problems of unculturability. Thus, recent research has set out to analyze lignocellulose-enriched microbial consortia by DNA-based approaches (as also known as targeted metagenomics) [7]. For example, recently, Jiménez *et al.* (2016) [8] reported the analysis of three soil-derived microbial consortia cultivated on biologically pretreated plant biomass. They analyzed the microbial structure, GH profile and extracellular enzymatic activities. Moreover, in a process known as "synthetic metagenomics", GH-encoding genes can be custom-synthesized and codon-optimized, after which their efficient expression can be achieved in a suitable host. In this respect, Dougherty *et al.* (2012) [9] identified, synthetized and expressed a total of 19 GHs originating from the metagenome of a switchgrass-adapted compost community. In the same way, Gladden *et al.* (2014) [10] discovered 18 active GHs that were tolerant to 10% of 1-ethyl-3-methylimidazolium acetate (ionic liquid used in the pretreatment of plant biomass).

In previous work, we developed two wheat straw-degrading microbial consortia derived from forest soil, in which substantially enriched (hemi)cellulolytic genes and activities were found [11]. In order to explore these consortia further, here we performed a targeted synthetic metagenomics approach. Thirteen large contigs - produced previously on the basis of shotgun sequencing of metagenomics DNA extracted from the aforementioned consortia - were selected and screened for GH-encoding genes [12]. In the current study, we report the selection of ten such genes on the basis of a combination of criteria. The genes were codon-optimized, synthesized and expressed, after which they were further

tested. Here, we describe the full analysis, placing a focus on a gene for a key furan aldehyde-tolerant β-xylosidase/α-arabinosidase (CAZy family GH43) enzyme that is proposed for biorefining processes, especially the saccharification of pretreated plant biomass.

## Materials and methods

### Identification and selection of GHs from a wheat straw-degrading consortial metagenome

Previous analyses of contigs constructed following shotgun metagenomics and sequencing of DNA from two wheat straw-degrading microbial consortia identified thirteen novel Bacteroidetes-derived hemicellulose utilization loci containing 39 GHs [12]. From the contigs, we selected ten predicted GH-encoding genes on the basis of the following criteria: *i*) genes encode highly enriched GH families compared with the original soil inoculum; *ii*) GHs are predicted to allow deconstruction of xylan, xyloglucan and galacto(gluco)mannan; *iii*) GHs are flanked by genes for membrane transporters and genes involved in sugar metabolism (i.e. coherent genomic context); *iv*) predicted GHs have low amino acid identity (e.g. <80%) compared to proteins in databases; *v*) GHs contain identifiable start and stop codons and a complete intact reading frame. In FigureS1 (supporting information), the selected genes and their genomic context are shown.

### Cloning and expression of ten GH-encoding genes recovered from the metagenome assemblages.

The ten genes were selected based on directed choices with respect to CAZy [13] family allocation. Specifically, we selected genes for families GH92, GH43, GH2, GH95 and GH29, as follows: M4684 and M3030 (GH92); M7068, M7073, M1989 and M8244 (GH43); M1927 and M20752 (GH2); M1916 (GH95) and M8239 (GH29) (Table 1). All genes were codon-optimized for expression in *Escherichia coli*, synthesized and cloned into the pET21b+ vector, with the help of a commercial partner (GenScript, Piscatawy, NJ, USA). For codon optimization, we used the OptimumGene™ algorithm and the Codon Adaptation Index (CAI). Additionally, other major codon usage biases, such as premature Poly-A sites, GC contents, internal chi sites and ribosome-binding sites, repeat sequences (direct, reverse and dyad repeats), as well as restriction sites that might interfere with cloning, were changed. The expression clones were introduced into *E. coli* BL21(DE3) competent cells (Invitrogen, Carlsbad, CA, USA) using the manufacturer's instructions. Following clone selection and purification, plasmid extractions were done for each of the ten cloned genes. Thereafter, the nature of the cloned fragments was checked by restriction fragment analyses. Specifically, *Xba*1/*Xho*1 were used for genes M3030, M1916, M1927, M1989 and M20752; *Mlu*1/*Xho*1 for genes M8239, M8244 and M7068; *Sal*1/*Xho*1 for M4684; and *Sac*1/*Xho*1 for M7073 (Figure S2). A single colony of each verified clone was then

introduced into an Erlenmeyer containing 10ml of fresh LB medium containing Overnight Express™ Autoinduction System 1 reagents (Novagen, Darmstadt, Germany) and ampicillin (100 μg $ml^{-1}$). The bacterial cultures were incubated overnight at 37°C with constant shaking at 200rpm. The cell pellets were harvested by centrifugation at 10000 *g* for 10 min and resuspended in 2 ml of lysis buffer (20 mmol $l^{-1}$ Tris-HCL pH 7.5, 100 mmol $l^{-1}$ NaCl, 1 mmol $l^{-1}$ EDTA, 0.1% Triton, 5 mmol $l^{-1}$ CHAPS and a mini tablet of protease inhibitor-Roche, Mannheim, Germany - to 50 ml). Subsequently, the resuspended cells were sonicated on ice (6sec on, 15sec off, 30 cycles with amplitude of 10-15 microns) and the lysates centrifuged at 14000*g* for 10min at 4°C in order to separate the soluble and insoluble protein fractions. Insoluble proteins were washed twice with 750μl of 20 mmol $l^{-1}$ Tris-HCL (pH 8.0) and solubilized with 2 mol $l^{-1}$ urea following the freeze-thawing method [13]. Protein concentrations were determined by the Bradford method using bovine serum albumin as the standard. Protein fractions were analyzed by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [15].

**Zymographic analysis and detection of enzymatic activity using para-nitrophenol-glycosides**

Zymograms were used to detect xylanase activity on SDS polyacrylamide gels (4% stacking, 10% resolving gels) containing 0.2% of xylan from beechwood (Sigma-Aldrich, Zwijndrecht, The Netherlands). Each well was loaded with 20μg (in 20μl) of total proteins per sample. After running the gels at 4°C, they were soaked for 1h in 2.5% of Triton and washed thoroughly in water prior to incubation (1h at 30°C) in 50 mmol $l^{-1}$ of sodium citrate buffer pH 6.0. The gel was stained with 0.1% Congo red for 30min and then de-stained for 2h in 1 mol $l^{-1}$ NaCl to reveal zones of clearing. Additionally, protein fractions (soluble and insoluble) were recovered and tested for activity using *p*-nitrophenyl β-D-xylopyranoside (*p*NP-Xyl), *p*-nitrophenyl α-L-arabinofuranoside (*p*NP-Ara), *p*-nitrophenyl α-L-fucopyranoside (*p*NP-Fuc) and *p*-nitrophenyl α-D-mannopyranoside (*p*NP-Man). The reaction mixtures consisted of 180μl of 2 mmol $l^{-1}$ of each *p*-nitrophenol-glycoside (diluted in 20 mmol $l^{-1}$ of Tris-HCL pH 7.0) and 20μl of each protein fraction. The mixtures were incubated at 37°C for 30min, after which the reactions were stopped on ice. Three negative controls were used for all assays: *i*) reaction mixture without substrate; *ii*) reaction mixture using the protein fractions from *E.coli* BL21(DE3) transformed with pET21b+ vector; and *iii*) reaction mixture without proteins. Activity was detected by the presence of yellow color in the reaction plate.

**Table 1.** Features of the overexpressed selected GH-encoding genes

| Gene ID | CAZy family | Gene length (bp) | Amino acids | PSI-BLASTp best hit [Taxa] (Accession number) | QC* | Identity | pI** | kDa |
|---|---|---|---|---|---|---|---|---|
| M4684 | GH92 | 3,111 | 1,032 | Hypothetical protein [*Parabacteroides* sp.] (WP_010801039.1) | 100% | 70% | 6.04 | 117 |
| M3030 | GH92 | 1,983 | 657 | Alpha-1,2-mannosidase [*Sphingobacterium spiritivorum*] (WP_002997981.1) | 98% | 83% | 9.24 | 74.2 |
| M7068 | GH43 | 996 | 328 | Glycosyl hydrolase family 32 [*Paraprevotella clara*] (WP_008623026.1) | 88% | 69% | 6.51 | 37.9 |
| M7073 | GH43 | 1,935 | 641 | Glycosyl hydrolase [*Flavobacterium johnsoniae*] (YP_001195445.1) | 98% | 66% | 7.11 | 73.3 |
| M1916 | GH29 | 1,641 | 543 | Alpha-1,3/4-fucosidase [*Capnocytophaga canimorsus*] (YP_004740108.1) | 98% | 63% | 6.53 | 61.5 |
| M1927 | GH2 | 3,216 | 1,068 | Beta-galactosidase [*Sphingobacterium spiritivorum*] (WP_002995396.1) | 85% | 50% | 7.08 | 120 |
| M1989 | GH43 | 981 | 323 | Hypothetical protein [*Sphingobacterium* sp.] (WP_021189556.1) | 100% | 98% | 4.70 | 37.5 |
| M20752 | GH2 | 1,293 | 427 | Beta-galactosidase/beta-glucuronidase [*Flavobacterium* sp.] (WP_007809792.1) | 90% | 56% | 8.63 | 48.5 |
| M8239 | GH95 | 1,302 | 430 | Alpha-L-fucosidase [*Pedobacter saltans*] (YP_004274942.1) | 98% | 72% | 6.74 | 48 |
| M8244 | GH43 | 1,605 | 531 | Hypothetical protein [*Sphingobacterium* sp.] (WP_021189555.1) | 55% | 99% | 5.17 | 61.1 |

* query coverage; ** isoelectric point.

**Bioinformatics analysis, phylogenetic tree and structural modeling of protein XylM1989**

For one protein that was successfully produced into the soluble fraction (gene M1989), the translated gene (*xyl*M1989) was analyzed based on BLASTp searches against the NCBI nonredundant protein database. In addition, catalytic domains were identified using the pFam database. The protein XylM1989 was aligned by ClustalW against thirty-five proteins from different origin that belong to CAZy families GH43, GH3 [16] and AA10 (outgroup sequences). For the multiple protein alignment and phylogenetic analyses, the software's MEGA v6.0 and PRALINE [17,18] were used. In order to detect the catalytic and substrate-binding sites, a XylM1989 protein data bank file was generated using Phyre2 [19]. With this prediction, the closest homolog of XylM1989 was a β-xylosidase protein 4MLG (Protein data bank ID) from an uncultivable bacterium (EC 3.2.1.37, GH43 family). This protein was used as a template for structural predictions using the PyMOL platform (http://www.pymol.org).

**Biochemical properties of protein XylM1989**

The optimum temperature for activity was determined in the range 10–70°C using *p*NP-Xyl and *p*NP-Ara (at pH 7.0). The pH optimum was determined in a pH range from 3.0 to 10.0 (at 30°C) using the following buffers: 50 mmol $l^{-1}$ sodium citrate (pH 3.0 to 6.0), 50 mmol $l^{-1}$ Tris–HCl (pH 7.0–9.0) and 50 mmol $l^{-1}$ glycine-NaOH (pH 10.0). The reaction mixture consisted of 280µl of 0.5 mmol $l^{-1}$ of *p*NP-Xyl or *p*NP-Ara and 20µl of soluble protein XylM1989 (approximately 1.7 mg $ml^{-1}$). The kinetic parameters ($K_m$ and $V_{max}$) of XylM1989 were determined with *p*NP-Xyl and *p*NP-Ara concentrations ranging from 0 to 50 mmol $l^{-1}$ in 20 mmol $l^{-1}$ Tris–HCl (pH 8.0) at 30°C for 15min. The data were plotted according to the Lineweaver-Burk method (double reciprocal plot). The effects of lignocellulosic hydrolysate inhibitors (furfural, 5-hydroxymethylfurfural and acetic acid) and chemical additives (ions, sugars, NaCl, EDTA, detergents and organic solvents), at different concentrations, on the activity of the XylM1989 protein were evaluated with *p*NP-Xyl (pH 8.0) at 30°C for 30min. Additionally, the effect of xylose (ranging from 0 to 1000 mmol $l^{-1}$) in the presence of 5 mmol $l^{-1}$ of $Mg^{2+}$, $Ca^{2+}$ and $Mn^{2+}$ was evaluated with the above parameters. Enzymatic activities were determined from the measured absorbance units using a standard calibration curve. The amount of para-nitrophenol (*p*NP) liberated was measured by absorbance at 410 nm. One unit (U) of enzyme activity was defined as the activity required for the formation of 1 µmol of *p*NP per min under the above conditions.

**Activity of the XylM1989 protein on complex polysaccharides**

The enzymatic activity of XylM1989 was evaluated on three complex polysaccharides (xylan from beechwood, oat spelt xylan and soluble arabinoxylan). The reaction mixtures (500µl) contained 1% of each polysaccharide (diluted in 20

mmol $l^{-1}$ of Tris-HCl pH 8.0) and 150µl of soluble protein XylM1989 (approximately 1.7 mg $ml^{-1}$). The mixtures were incubated at 30°C for 72h, subsequently the reactions were stopped on ice and centrifuged for 10min at 12000*g*. The enzymatic activity was determined by measuring the amount of reducing sugars in the supernatant by the 3,5-dinitrosalicilic acid (DNS) method [20]. A standard calibration curve was used, as previously constructed with different concentrations of xylose. In addition, the types of sugars and their concentrations, released by the enzymatic reaction, were analyzed by high-performance liquid chromatography (HPLC). Two negative controls were set up: *i*) reaction without substrate and *ii*) reaction without protein.

## Results

### Synthesis, cloning, expression and enzymatic analysis of ten genes predicted to encode GHs

From two wheat-straw-degrading microbial consortia, we here selected ten genes encoding predicted GHs for synthesis. All ten genes were codon-optimized for *E. coli*, synthetized and cloned into the pET21b+ expression vector with specific restriction enzyme sites on both sides (*Nde*1 and *Xho*1). For all genes, the fragment sizes - as measured on agarose gels - were consistent with the predicted sizes of the sequences (Table 1; Figure S2). Then, using overnight cultures (in LB medium) of each selected purified clone, gene expression was induced. The data revealed that, among the cultures from the ten cloned genes, only one protein (gene M1989) occurred in the soluble fraction (~80-90% pure), whereas the remainder was mainly present in inclusion bodies (Figure 1a). To enhance solubility, we applied solubilization methods to the latter, so as to recover and refold each aggregated protein into its native state. The nine inclusion bodies were thus isolated, purified and then treated with 2 mol $l^{-1}$ urea following the freeze-thawing method. Fractions were diluted (1-, 10-, and 100- fold) into PBS buffer in order to decrease the urea concentration and improve the refolding of the protein [13]. Unfortunately, the final suspensions did not show any enzymatic activities, suggesting persistent misfolding or aggregation of the nine proteins (data not shown). However, zymogram analysis with beechwood xylan revealed that the "insoluble" protein fractions of genes M7068, M1916, M20752, M8239 and M8244 represented a clear zone with enzymatic activity. Moreover, the (soluble) M1989 cell lysate also had activity. Thus, the products of six in ten genes produced in *E. coli* had xylanase activity, of which only one, M1989, appeared in the soluble fraction (Figure 1b).

***Testing for different enzymatic activities*** - Based on the prediction of the activities of the gene products by CAZy database annotation (Table 1), we selected four *p*NP-labelled substrates, i.e. *p*NP-Xyl, *p*NP-Ara, *p*NP-Fuc and *p*NP-Man, to evaluate the putative enzymatic activities (using both insoluble and soluble

fractions). However, with the exception of the lysate of clone M1989 (soluble fraction), none of the lysates showed hydrolytic activity with any of the selected substrates (Figure 1c). Indeed, the product of clone M1989 showed dual activity, i.e. with *p*NP-Xyl and *p*NP-Ara, but not with *p*NP-Man and *p*NP-Fuc. On the basis of its activity, protein M1989 will be denoted XylM1989. It is the basis of the further results described below.

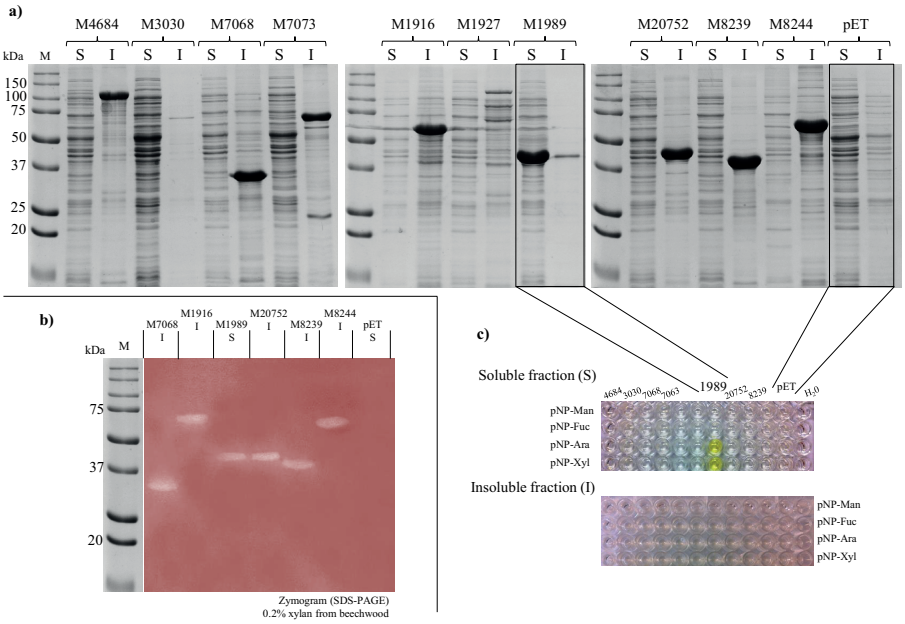



**Figure 1**. (a) SDS-PAGE of the overexpressed GHs in Escherichia coli Bl21(DE3) cells. (b) Zymogram analysis of six GHs that showed hydrolytic activity and one negative control (soluble protein fraction from E. coli BL21(DE3) transformed with pET21b+ vector; lane pET S). (c) Detection of enzymatic activities of the overexpressed GHs by using *p*NP-labelled substrates. Supernant fractions for the cell extracts are denoted as S, and for the insoluble pellets of the lysates are denoted as I. M, protein marker.

## Analysis of the XylM1989 protein – CAZy family, phylogeny and structural prediction

The protein XylM1989 has a calculated isoelectric point (pI) of 6.16 and a molecular weight of 37.5 kDa. In addition, XylM1989 was predicted to belong to the GH43 family, which contains mostly β-xylosidases (EC 3.2.1.37), α-arabinosidases (EC 3.2.1.55), galactan 1,3-β-galactosidases (EC 3.2.145) and endo-α-arabinases (EC 3.2.1.99) [16,21]. Based on the BLASTp analysis, the amino acid sequence of XylM1989 showed 95% identity (100% coverage) with an uncharacterized GH43 family protein (ACX30651) encoded by a chromosomal segment of *Sphingobacterium* sp. TN19 [22]. In addition, protein XylM1989 showed 63% identity with a characterized bifunctional GH43 family xylosidase/arabinosidase (*xyn*B; CAA89208) from *Prevotella bryantii* [23]. The more

detailed phylogenetic analyses (including different types of family GH3 and GH43 enzymes) further showed that protein XylM1989 clustered with uncharacterized family GH43 xylosidases and arabinosidases from different Bacteroidetes, specifically belonging to species of *Sphingobacterium*, *Draconibacterium*, *Proteiniphilum*, *Dysgonomonas* and *Chryseobacterium*. XylM18989 revealed a relatively low degree of similarity with characterized family GH43 bacterial xylosidases (EC 3.2.1.37) next to fungal endo-arabinases (Figure 2). Moreover, it showed 71% identity protein with protein 4MLG (structure of RS223-β-xylosidase) [24]. Based on the predicted 3D structure using protein 4MGL as the template and multiple alignments with phylogenetically closer proteins, we identified a catalytic triad (Asp15 – Asp135 – Glu222) and a substrate-binding site (Trp83 – Ile134 –Thr271) (Figure 3).

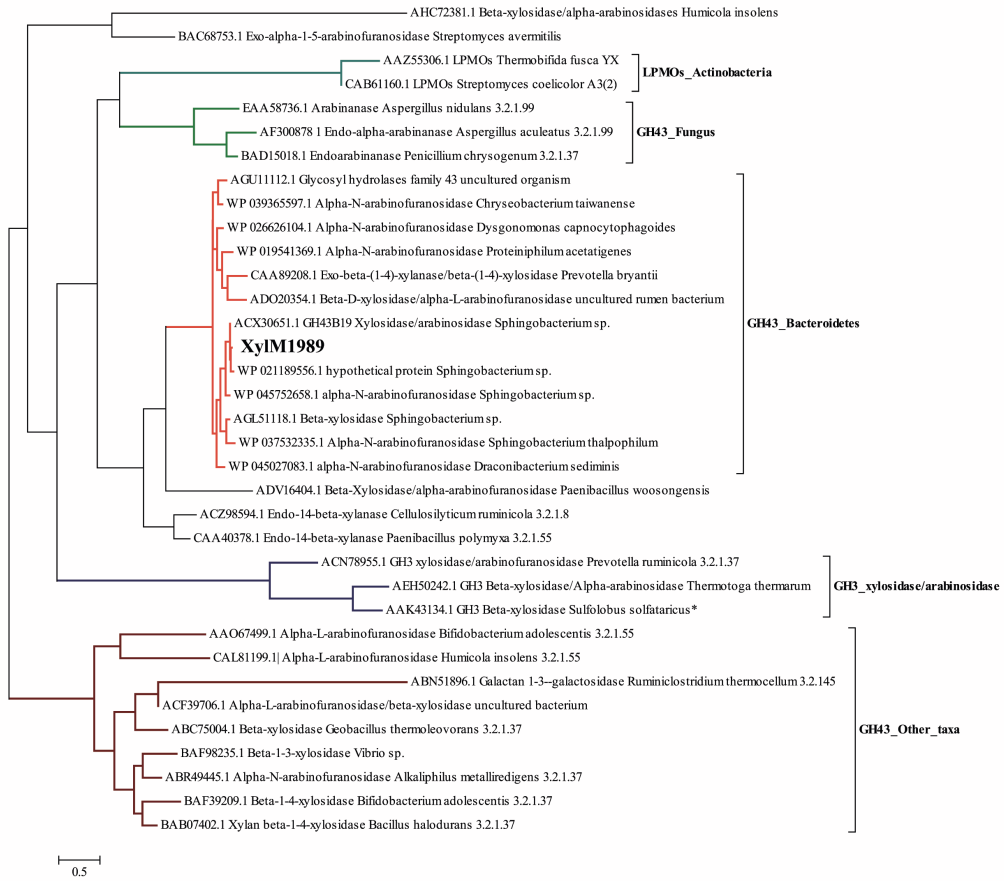



**Figure 2.** Maximum likelihood phylogenetic analysis of XylM1989 with closest related proteins. The amino acid sequences of the other GH43 and GH3 enzymes were obtained from published data (Lagaert et al. 2014). Two Actinobacteria-derived lytic polysaccharide monooxygenases (LPMOs) were used as outgroup (CAZy family AA10). The tree is drawn to scale, with branch lengths measured conform the number of substitutions (amino acids) per site. All positions containing gaps and missing data were eliminated and the evolutionary analyses were conducted in MEGA ver. 6.0.

**Biochemical characterization of protein XylM1989**

***Effects of temperature and pH on protein XylM1989 activity*** - Based on the finding that protein XylM1989 had β-xylosidase and α-arabinosidase activities, these activities were characterized with respect to temperature and pH ranges. Protein XylM1989 exhibited maximal activities at 20°C in the presence of 0.5 mmol $l^{-1}$ of (buffer pH 7.0) *p*NP-Xyl (3.36 U $mg^{-1}$ of protein) and *p*NP-Ara (0.65 U $mg^{-1}$ of protein) (Figure 4a). The activity of XylM1989 decreased, to approximately 20% of the maximal activity, when the temperature was raised from 40°C to 70°C for both *p*NP-Xyl and *p*NP-Ara. As shown in Figure 4b, this activity was then assessed at pH values between 3.0 and 10.0. In this analysis, the maximal β-xylosidase activity was reached at pH 8.0 (30°C, 0.5 mmol $l^{-1}$ of *p*NP-Xyl). The activity, overall, remained at ~70% of the maximum at pH 9.0, whereas it was completely lost at pH 10.0. Moreover, activity on *p*NP-Ara was also maximal at pH 8.0. Indeed, the latter activity was still at 60% at pH 9.0, suggesting that the protein is considerably active under slightly to strongly alkaline conditions.

***Kinetic analysis of the XylM1989 protein*** - The XylM1989 hydrolytic activity was measured with respect to catalytic properties and by assessment of kinetic parameters, i.e. the $K_m$ (Michaelis-Menten constant) and $V_{max}$ values (maximal reaction velocities), using *p*NP-Xyl and *p*NP-Ara as the substrates, under optimal conditions. The $K_m$ values of XylM1989 for *p*NP-Xyl and *p*NP-Ara were 1.2 mmol $l^{-1}$ and 0.781 mmol $l^{-1}$, and the $V_{max}$ values 285.71 U $mg^{-1}$ and 78.12 U $mg^{-1}$, respectively. Given the fact that the $K_m$ and $V_{max}$ values of protein XylM1989 for *p*NP-Xyl were higher than that for *p*NP-Ara, *p*NP-Xyl was used for further analysis.

***Effects of additives on protein XylM1989 activity*** - The effects of different additives on protein XylM1989 β-xylosidase activity were assessed (Table 2). Remarkable increases in the β-xylosidase activity were observed in the presence of 5 mmol $l^{-1}$ of $CaCl_2$ (10-fold), $MgCl_2$ (12-fold) and $MnCl_2$ (7-fold). In addition, a 2.5-fold enhanced activity was observed with 50 mmol $l^{-1}$ L-arabinose, but this activity decreased by 47% in the presence of xylose. Furthermore, the activity decreased by 60% with 50 mmol $l^{-1}$ of EDTA. It increased slightly (118.6 ± 24) upon addition of 20% glycerol. Concurrently, the activity dropped by 75% in the presence of all organic solvents, i.e. ethanol, methanol and isopropanol. Moreover, the activity decreased by 50% in the presence of 10% DMSO.

***Effects of plant biomass hydrolysate-derived compounds on protein XylM1989 activity*** - Three lignocellulosic hydrolysate-derived compounds with potential inhibitory activity, i.e. furfural, 5-hydroxymethylfurfural (5-HMF) and acetic acid, were tested for their effects on the activity of protein XylM1989 with *p*NP-Xyl (0.5 mmol $l^{-1}$; pH 8.0 at 30°C) (Figure 5a). Interestingly, in the presence of 0.3% (w/v)

acetic acid, inhibition was high and the relative β-xylosidase activity was nearly zero. Moreover, it was also strongly blocked by 0.7% (w/v) 5-HMF (~80% inhibition), whereas low levels (0.05-0.1% w/v) of 5-HMF showed only 15-20% of inhibition. The presence of 0.7% (w/v) furfural also resulted in around 60% of inhibition. However, at lower concentrations (0.05–0.5% w/v of furfural) the activity inhibition remained at 20-40%.

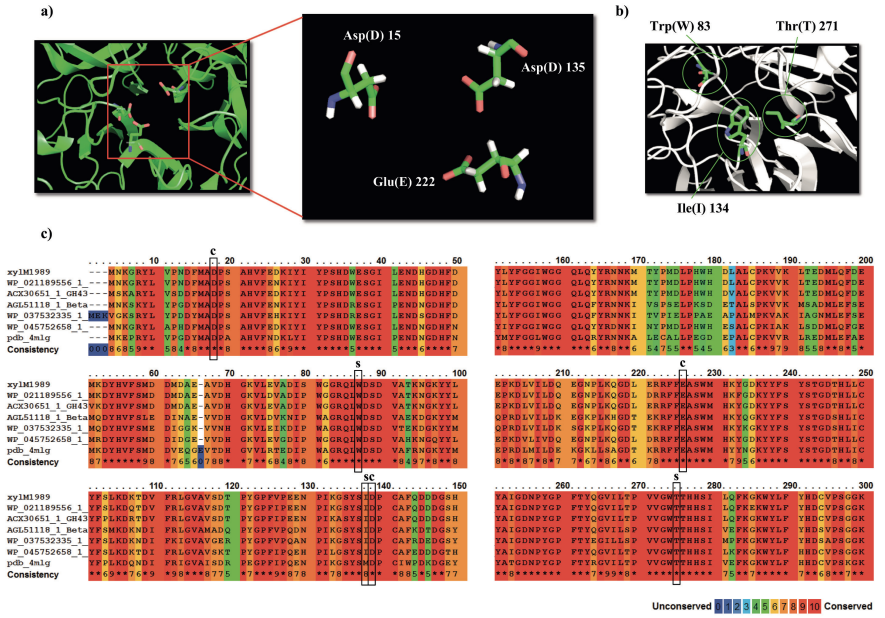



**Figure 3.** a) *In silico* 3D structure prediction of the enzyme XylM1989, generated by molecular modelling, showing the catalytic and b) substrate binding sites. c) Conserved blocks in the deduced amino acid sequences of XylM1989. Highly conserved residues are red shaded. Residues of the catalytic triad and substrate binding pocket are denoted by C and S, respectively. The aligned GH43 sequences came from *Sphingobacterium*-related organism (WP_021189556.1; ACX30651.1; WP_045752658.1; WP_037532335.1; AGL51118.1) and uncultured organism (Protein data bank ID: 4MLG).

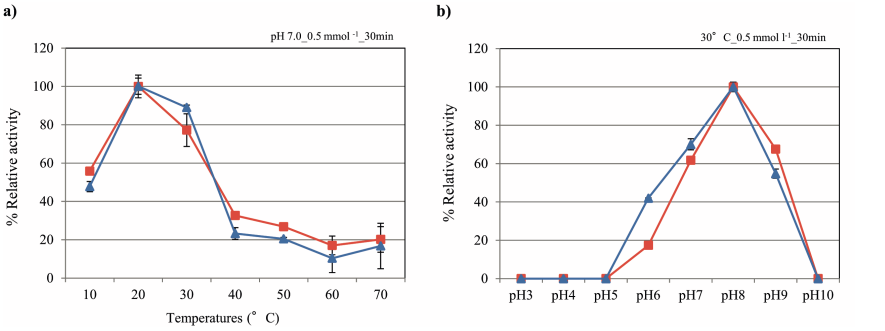



**Figure 4.** Relative enzymatic activity of the XylM1989 protein measured against pNP-β-D-xylopyranoside (pNP-Xyl) (■) and pNP-α-L-arabinopyranoside (pNP-Ara) (▲). a) Different temperature with constant pH 7.0. b) Different pH values at 30°C.

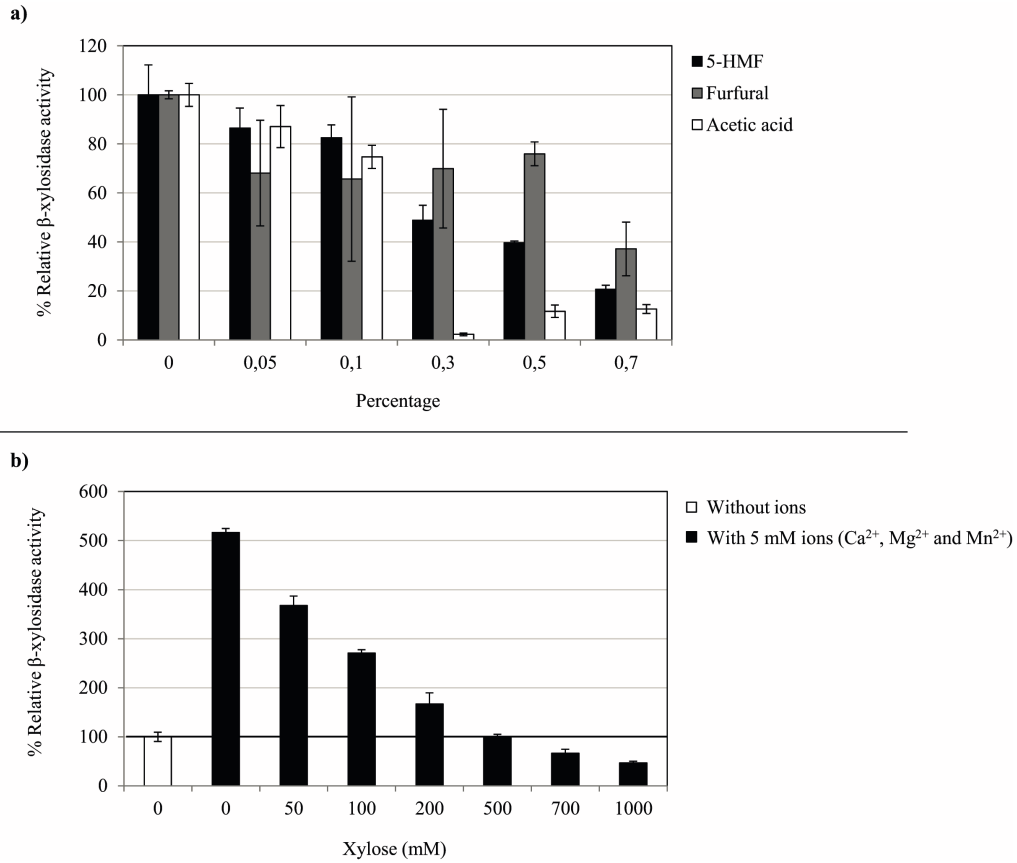



**Figure 5.** Relative β-xylosidase activity of the protein XylM1989 at **a)** different concentrations (zero to 0.7%)% w/v) of furfural, 5-hydroxymethylfurfural (5-HMF) and acetic acid; and **b)** different concentrations (zero to 1000 mmol $l^{-1}$) of xylose in presence of 5 mmol $l^{-1}$ ions ($Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$).

*Effect of xylose on protein XylM1989 activity* - The β-xylosidase activity of XylM1989 was inhibited by addition of xylose (Table 2). However, the activity increased approximately 3.5- fold in the presence of 50 mmol $l^{-1}$ xylose with ions (5 mmol $l^{-1}$ of $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$) over the control (without xylose and ions). In the presence of each of the three bivalent cations, without xylose, the xylM1989 β-xylosidase activity was 5-fold increased over that of the control (Figure 5b). At 200 mmol $l^{-1}$ xylose (with ions), protein XylM1989 still showed an activity of 80% over that of the control without xylose and ions. Finally, the enzyme activity dropped to around 50% of that of the control at high concentrations of xylose (700 and 1000 mmol $l^{-1}$).

**Activity of protein XylM1989 on semi-natural substrates**

The hydrolytic activity of protein XylM1989 on 1% of xylan from beechwood (XB), oat spelt xylan (OX), and arabinoxylan (ARB) was tested, using 5 mmol $l^{-1}$ of

**Table 2.** Effects of different additives on the XylM1989 β-xylosidase activity

| Type of additive | Reagent | % Relative activity | |
|---|---|---|---|
| **Ions** | | **5 mmol $l^{-1}$** | **20 mmol $l^{-1}$** |
| | $CaCl_2$ | 1065.23 ± 74 | 546.58 ± 39 |
| | $MgCl_2$ | 1214.18 ± 16 | 691.11 ± 120 |
| | $CoCl_2$ | 6.27 ± 1.1 | 1.33 ± 0.3 |
| | $NiCl_2$ | 8.67 ± 1.5 | 2.01 ± 1.2 |
| | $CuCl_2$ | 0.63 ± 0.1 | 1.94 ± 0.4 |
| | $MnCl_2$ | 738.93 ± 8.4 | 172.88 ± 20 |
| | $NH_4$ | 69.32 ± 1.0 | 63.61 ± 4.3 |
| **Sugars** | | **20 mmol $l^{-1}$** | **50 mmol $l^{-1}$** |
| | Glucose | 69.30 ± 1.3 | 99.28 ± 4.9 |
| | Xylose | 55.82 ± 2.5 | 53.71 ± 1.4 |
| | Arabinose | 164.89 ± 5.4 | 249.26 ± 3.8 |
| | Cellobiose | 82.32 ± 2.0 | 90.32 ± 14 |
| | Galactose | 80.38 ± 1.8 | 90.99 ± 12 |
| **Salt** | | **200 mmol $l^{-1}$** | **2000 mmol $l^{-1}$** |
| | NaCl | 96.06 ± 6.1 | 36.34 ± 0.6 |
| **Chelating agent** | | **20 mmol $l^{-1}$** | **50 mmol $l^{-1}$** |
| | EDTA | 61.66 ± 0.8 | 30.20 ± 0.6 |
| **Detergents** | | **2%** | **5%** |
| | SDS | 8.68 ± 0.7 | 6.12 ± 4.0 |
| | Triton | 107 ± 5.6 | 32.70 ± 0.8 |
| **Organic solvents** | | **10%** | **20%** |
| | Glycerol | 101.9 ± 4.7 | 118.6 ± 24 |
| | Ethanol | 26.63 ± 0.1 | 11.40 ± 0.2 |
| | Methanol | 26.47 ± 0.2 | 16.63 ± 1.9 |
| | Isopropanol | 26.10 ± 0.6 | 12.53 ± 1.9 |
| | DMSO | 50.56 ± 1.0 | 24.55 ± 0.1 |

$Mg^{2+}$. Interestingly, 68.73 ± 3.52 mg of sugars $g^{-1}$ of polysaccharide were released from XB and 62.28 ± 6.96 mg $g^{-1}$ of polysaccharide from OX, whereas only 14.33 ± 1.64 mg $g^{-1}$ of polysaccharide was produced from ARB. Based on HPLC data (not shown), the most abundant sugars released in the reaction with XB were (listed in order of estimated quantity): xylose>glucose>galactose>xylobiose. Regarding the OX and ARB reactions, we observed that the xylose and glucose levels exceeded that of galactose, arabinose and xylobiose. On the basis of these results, we posit that protein XylM1989 works avidly on xylan from beechwood, oat spelt xylan and arabinoxylan, releasing sugars in accordance with the specifics of these substrates.

## Discussion

In this study, ten GH-encoding genes, retrieved from two wheat straw-degrading microbial consortia [12], were selected for codon optimization, synthesis, cloning, expression and characterization. These genes all originated from different Bacteroidetes, a dominant phylum in wood-degrading communities [25] and were found in different microbial consortia cultivated on agricultural residues [26,27]. The microbial origin, the genomic context and the annotation of the genes for these selected GHs all suggest that they play an important role in plant biomass degradation [12]. Inspired by recent data [28], we intended to increase the expression levels of such GH-encoding genes (in *E. coli*), using codon swapping by

a commercial routine. However, such codon optimization comes with potential drawbacks that are related to the fast depletion of the precise cognate tRNAs in the host organism, which can incite translational errors. An exact "smart mix" of major and minor codons appears to be necessary in each case, which may be gene and host dependent. Indeed, the feasibility of any protocol for heterologous protein expression is often not theoretically foreseeable [29]. In our study, nine of the ten selected GHs were found in inclusion bodies inside the *E. coli* host, which was likely due to overexpression. Such inclusion body location is actually an advantage in the industrial production of high quantities of proteins. However, due to misfolding and aggregation, the included proteins may become inactive. To tackle this problem, recently different types of expression host (e.g. *E. coli* origami; Novagen), protein refolding kits and new protocols have been developed [30,31]. In our study, the finding that the products of five genes (in the insoluble fractions) apparently had xylanase activity, whereas no activity could be detected with any of the *p*NP substrates, indicated that protein aggregates may have been differentially dissociated and thus (ephemerally) active. For instance, in the zymography assay, the removal of SDS by Triton may refold the protein correctly in the gel with subsequent evidence of activity [32,33]. Notably, by using zymography we showed, for the first time, that proteins (M1916 and M8239) annotated as fucosidases have xylanolytic activity (Figure 1b; Table 1). Remarkably, in the presence of the *p*NP substrates, protein XylM1989 showed β-D-xylosidase activity and α-L-arabinosidase activity. Given the problems with the insolubility of the other proteins, we henceforth only studied the soluble protein XylM1989 in detail. Interestingly, the genomic context of the XylM1989-encoding gene resembled that of the gene for an uncharacterized enzyme (GH43B19) located in a 37.5 kb chromosome fragment of *Sphingobacterium* sp. TN19 (Figure S3). This strain had been isolated from the gut of *Batocera horsfieldi* larvae, a beetle which develops in woody tissues. In the chromosome fragment, genes for three xylanolytic enzymes have been characterized (XynB19, GH43A19, XynA19), suggesting their importance in hemicellulose depolymerization [22,34]. We here found a shared synteny (with our contig_248) of genes encoding transketolases, xylulokinases, xylose isomerases and ABC transporters. These findings suggest that the flanking genes to the gene encoding XylM1989 are involved in xylan degradation, sugar transport and xylose metabolism (Figure S3).

In earlier work, several β-xylosidases and α-arabinosidases of CAZy family GH43 were recovered and characterized from various organisms (e.g. *Bacillus*, *Fibrobacter*, *Thermobifida*, *Clostridium*, *Thermotoga*, *Enterobacter* and *Paenibacillus*) [35–42]. However, this was often accompanied by an incomplete characterization which did not allow the understanding of their full potential. For instance, the tolerance to high levels of (inhibitor) compounds derived from lignocellulosic hydrolysates has not been adequately addressed [43]. It is worth noting that these GHs constitute key components of enzyme cocktails used for the improved sacch-

**Table 3.** Comparison of the protein XylM1989 with GH43 family enzymes from other studies

| Enzyme | Microbial Source | kDa | Substrate | Optimal pH | Optimal °C | Activity (U $mg^{-1}$) | Km (mmol $l^{-1}$) | $V_{max}$ (µmol $min^{-1}$ $mg^{-1}$) | References |
|---|---|---|---|---|---|---|---|---|---|
| XylM1989 | *Sphingobacterium* sp. | 37.5 | *p*NP-X/*p*NP-A | 8.0 | 20 | 9.65/1.22 | 1.2/0.781 | 285.71/78.12 | This work |
| S2 | *Penicillium herquei* | 37.4 | *p*NP-X | 6.5 | 30 | 225 | *ND | *ND | Ito *et al*. 2003 [57] |
| Abn2 | *Bacillus subtilis* | 46.0 | *p*NP-A | 7.0 | 50 | *ND | *ND | *ND | Inácio and de Sa-Nogueira 2008 [38] |
| FSUAXH1 | *Fibrobacter succinogenes* | 84.0 | *p*NP-X | 7.5 | 45 | *ND | *ND | *ND | Yoshida *et al*. 2010 [42] |
| XylB | *Aspergillus oryzae* | 37.4 | *p*NP-X | 7.0 | 30 | 6.1 | 0.48 | 42.6 | Suzuki *et al*. 2010 [58] |
| PtXyl43 | *Paecilomyces thermophila* | 52.3 | *p*NP-X | 7.0 | 55 | 45.4 | 4.5 | 90.2 | Teng *et al*. 2011 [59] |
| TlXyl43 | *Thermomyces lanuginosus* | 45.0 | *p*NP-X | 6.5 | 55 | 45.4 | 3.9 | 107.6 | Chen *et al*. 2012 [60] |
| RuXyn1 | *Prevotella bryantii* | 45.0 | *p*NP-X/*p*NP-A | 7.0 | 40 | 36.3/14.2 | 3.43/2.23 | *ND | Zhou *et al*. 2012 [46] |
| Xyl43A | *Humicola insolens* | 37.0 | *p*NP-X | 6.5 | 50 | 20.5 | 12.2 | 203.8 | Yang *et al*. 2014 [61] |
| Xyl43B | *Humicola insolens* | 62.0 | *p*NPX | 7.0 | 50 | 1.7 | 1.29 | 2.18 | |

*ND: not determined.

-arification of plant biomass and production of biofuels [44]. The aforementioned studies showed that some, but not all, GH43 family enzymes have bifunctional activities, in particular β-xylosidase and α-arabinosidase. For instance, protein XylC, isolated from *Paenibacillus woosongensis*, had dual activity [39], and some proteins encoded by genes isolated from rumen and compost-derived metagenomes also had dual activity [9,45,46]. In contrast, the enzyme His-Xyl43 showed β-xylosidase activity [36], whereas proteins Abn2 and AbnZ2 had only endo-arabinanase activity [38,41]. We speculated, on the basis of the foregoing, that the β-D-xylosidase/α-L-arabinosidase activity of our XylM1989 enzyme may constitute a key asset in wheat straw biomass saccharification, with specific involvement in xylan and arabinan degradation. Indeed, compared to previously described family GH43 enzymes, it showed a raised $V_{max}$ (also compared with fungal enzymes) (Table 3). Thus XylM1989 has a higher reaction speed, enabling a faster substrate processing when the protein is saturated with the substrate.

Regarding the biochemical characterization, the three bivalent cations $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$ clearly enhanced the activity of XylM1989, thus indicating that such cations are important as enzyme cofactors. A role in the enzymatic reaction, e.g. by binding and stabilizing the substrate at the active site, may be invoked [35,47,48]. In contrast, $Cu^{2+}$, $Ni^{2+}$, $Co^{2+}$ and $NH^{4+}$ strongly inhibited the β-xylosidase activity of XylM1989. Similar to XylM1989, a xylanase produced by a gene isolated from a bovine rumen metagenome showed enhanced β-xylosidase activity with $Mn^{2+}$, whereas $Cu^{2+}$, $Fe^{2+}$, $Ag^{2+}$ and $Zn^{2+}$ ions inhibited the activity [49]. Helper molecules like sugars and ions can control enzyme activities by "setting" proteins "on" and "off" in response to environmental changes. Thus, feedback inhibition may occur due to allosteric regulation, in which molecules bind to the catalytic site of enzymes, altering their structural shape and changing the protein to an active or inactive form. The fact that XylM1989 activity was slightly stimulated by L-arabinose may relate to the binding of this sugar to the substrate-binding site, thus enabling the XylM1989 catalytic residues to react effectively.

Furfural and 5-HMF are major byproducts from the pretreatment of lignocellulosic materials [50,51]. These aromatic compounds released in hydrolysates are considered to be major inhibitors of fermentation processes [43]. During plant biomass pretreatment, several released products (including furanic compounds and monosaccharides) can inhibit the activity of the enzymes (or cocktails) used for the subsequently saccharification process. In addition, several studies has been showed a strong inhibition of ethanol production due to the presence of lignocellulosic byproducts [43]. For example, concentration of ~0.5% (w/v) and ~0.7% (w/v) of furfural and 5-HMF, respectively, can inhibit the growth rate and production of ethanol by *Issatchenkia orientalis* [52]. The concentration of furfural and 5-HMF in plant biomass hydrolysates depends on the pretreatment conditions and the feedstock. However, analyses of pretreated corn stover, poplar and pine materials showed furfural concentrations up to 0.22 g

$l^{-1}$ (0.022% w/v) and 5-HMF concentrations up to 0.17 g $l^{-1}$ (0.017% w/v) [53]. Interestingly, protein XylM1989 was tolerant to furfural, as its β-xylosidase activity was still at approximately 80% at 0.5% (w/v) furfural. It also showed some (restricted) tolerance to 0.3% (w/v) 5-HMF (50%). We posit here that such furan-tolerant xylanases have great potential for use in the biorefining industries, as they would presumably work well in the presence of expected levels of furfurals and related compounds. As far as we know, this is the first report of a β-xylosidase/α-arabinosidase that is tolerant to furfural and 5-HMF. Moreover, as was explained before, β-xylosidases are key in the conversion of xylo-oligosaccharides to xylose as the end-product. However, xylose is one of the major inhibitors of β-xylosidase activity [54]. Thus, xylose-tolerant β-xylosidases are important in hemicellulose conversion. The sensitivity of most β-xylosidases to xylose (as tested with fungal-produced xylosidases, such as those from *Arxula adeninivoran, Aureobasidium pullulans* and *Trichoderma reesei*) is striking, with *Ki* (concentration of inhibitor) values for xylose ranging from 2-10 mmol $l^{-1}$ [55]. In contrast, *Thermotoga thermarum* Tth xynB3 β-xylosidases showed high xylose-tolerant activity at 500 mmol $l^{-1}$ [56]. Interestingly, our novel xylM1989 protein showed 100% activity at xylose concentrations of 500 mmol $l^{-1}$, in the presence of ions (either $Ca^{2+}$, $Mg^{2+}$ or $Mn^{2+}$). Strikingly, the XylM1989 activity in the presence of xylose (with ions) was even enhanced by relatively low levels of xylose. In conclusion, our enzyme is inhibited by xylose, but this inhibition is alleviated by the presence of $Ca^{2+}$, $Mg^{2+}$ or $Mn^{2+}$.

Finally, five key features makes that the protein XylM1989 is a good candidate for use in industrial process related with plant biomass saccharification: *i*) active at alkaline pH, *ii*) higher reaction speed ($V_{max}$), *iii*) tolerance to lignocellulosic hydrolysates-derived inhibitors, *iv*) tolerance to high concentration of xylose in presence of $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$ and *v*) activity and release of sugars from complex polysaccharides such as a XB, OX and ARB. The properties of XylM1989 turn this enzyme into a promising puzzle part for the design of enzyme cocktails useful for the saccharification of (pretreated) plant biomass.

**Acknowledgements**

This work was supported by the BE-Basic foundation (http://www.be-basic.org). We thank H. Ruijssenaars and R. van Kranenburg for scientific support.

**References**

1. Saha BC. Hemicellulose bioconversion. J Ind Microbiol Biotechnol. 2003;30:279–291.
2. De Souza WR. Microbial degradation of lignocellulosic biomass in sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization. InTech; 2013.
3. Barker IJ, Petersen L, Reilly PJ. Mechanism of Xylobiose Hydrolysis by GH43 β-Xylosidase. J Phys Chem B. 2010;114:15389–15393.
4. Bokhari SAI, Latif F, Akhtar MW, Rajoka MI. Characterization of a β-xylosidase produced by a mutant derivative of Humicola lanuginosa in solid state fermentation. Annals of Microbiolog.

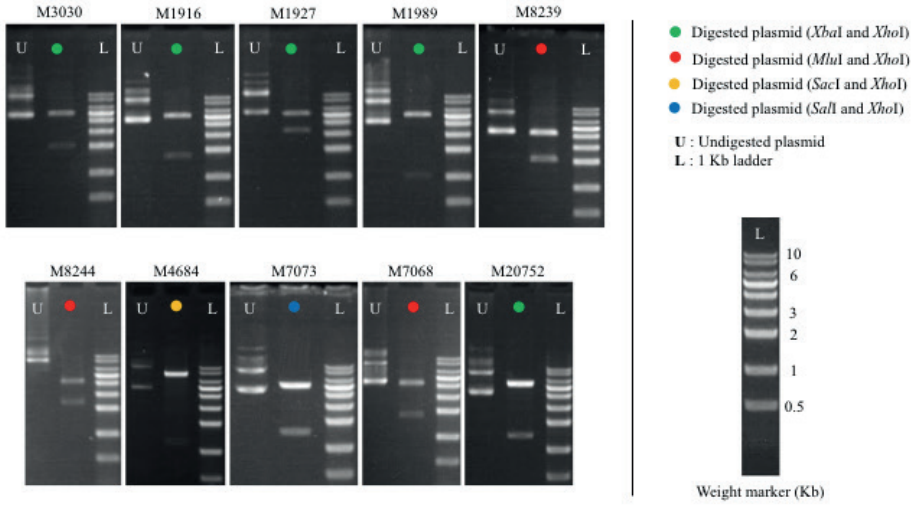
2010;60(1):21–29
5. Van den Brink J, de Vries RP. Fungal enzyme sets for plant polysaccharide degradation. Appl Microbiol Biotechnol. 2011;91:1477–1492.
6. Gao D, Uppugundla N, Chundawat SP, Yu X, Hermanson S, Gowda K, et al. Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. Biotechnol Biofuels. 2011;4:5.
7. D'haeseleer P, Gladden JM, Allgaier M, Chain PSG, Tringe SG, Malfatti SA, et al. Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. PLoS One. 2013;8:e68465.
8. Jiménez DJ, de Lima Brossi MJ, Schückel J, Kračun SK, Willats WGT, Van Elsas JD. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. Appl Microbiol Biotechnol. 2016;100(24):10463-10477
9. Dougherty MJ, D'haeseleer P, Hazen TC, Simmons BA, Adams PD, Hadi MZ. Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. BMC Biotechnol. 2012;12:38.
10. Gladden JM, Park JI, Bergmann J, Reyes-Ortiz V, D¿haeseleer P, Quirino BF, et al. Discovery and characterization of ionic liquid-tolerant thermophilic cellulases from a swithchgrass-adapted microbial community. Biotechnol Biofuels. 2014;7:15.
11. Jiménez DJ, Dini-Andreote F, Van Elsas J. Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. Biotechnol Biofuels. 2014;7:92.
12. Jiménez DJ, Chaves-Moreno D, Van Elsas JD. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. Sci Rep. 2015;5:13845.
13. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490-495.
14. Qi X, Sun Y, Xiong S. A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. Microb Cell Fact. 2015;14:24.
15. Sambrook J, Russell DW. Molecular cloning : a laboratory manual. 2001;3rd editio.
16. Lagaert S, Pollet A, Courtin CM, Volckaert G. β-Xylosidases and α-l-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. Biotechnol Adv. 2014;32:316–332.
17. Simossis VA, Heringa J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Res. 2005;33:w289-w294
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–2729.
19. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the phyre server. Nature Protoc. 2009;4:363–371.
20. Miller GL. Use of Dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem. 1959;31:426–428.
21. Jordan DB, Wagschal K, Grigorescu AA, Braker JD. Highly active β-xylosidases of glycoside hydrolase family 43 operating on natural and artificial substrates. Appl Microbiol Biotechnol. 2013;97:4415–4428.
22. Zhou J, Meng K, Yang P, Shi P, Wang Y, Luo H, et al. Characterization of a chromosomal segment showing xylanolytic activity from the symbiotic *Sphingobacterium* sp. TN19. World J Microbiol Biotechnol. 2010;26:761–765.
23. Gasparic A, Martin J, Daniel AS, Flint HJ. A xylan hydrolase gene cluster in *Prevotella ruminicola* B(1)4: sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and beta-(1,4)-xylosidase activities. Appl Environ Microbiol. 1995;61:2958–2964.
24. Jordan DB, Braker JD, Wagschal K, Lee CC, Chan VJ, Dubrovska I, et al. X-ray crystal structure of divalent metal-activated β-xylosidase, RS223BX. Appl Biochem Biotechnol.

2015;177:637–648.
25. Van der Lelie D, Taghavi S, McCorkle SM, Li LL, Malfatti SA, Monteleone D, et al. The metagenome of an anaerobic microbial community decomposing poplar wood chips. PLoS One. 2012;7(5):e36740.
26. De Lima Brossi MJ, Jiménez DJ, Cortes-Tolalpa L, Van Elsas JD. Soil-derived microbial consortia enriched with different plant biomass reveal distinct players acting in lignocellulose degradation. Microb Ecol. 2016;71:616–627.
27. Cortes-Tolalpa L, Jiménez DJ, De Lima Brossi MJ, Salles JF, Van Elsas JD. Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. Appl Microbiol Biotechnol. 2016;100(17):7713-7725.
28. Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol. 2014;5:21.
29. Kurland C, Gallant J. Errors of heterologous protein expression. Curr Opin Biotechnol. 1996;7:489–493.
30. Vallejo L, Rinas U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact. 2004;3:11.
31. Yamaguchi H, Miyazaki M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules. 2014;4:235–251.
32. Peterson R, Grinyer J, Nevalainen H. Extracellular hydrolase profiles of fungi isolated from koala faeces invite biotechnological interest. Mycol Prog. 2011;10:207–218.
33. Vandooren J, Geurts N, Martens E, Van den Steen PE, Opdenakker G. Zymography methods for visualizing hydrolytic enzymes. Nature Methods. 2013;10:211–220.
34. Zhou J, Huang H, Meng K, Shi P, Wang Y, Luo H, et al. Molecular and biochemical characterization of a novel xylanase from the symbiotic *Sphingobacterium* sp. TN19. Appl Microbiol Biotechnol. 2009;85:323–333.
35. Ahmed S, Luis AS, Bras JLA, Ghosh A, Gautam S, Gupta MN, et al. A Novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (Ct43Araf) from *Clostridium thermocellum*. PLoS One. 2013;8:e73575.
36. Campos E, Negro Alvarez MJ, Sabarís di Lorenzo G, Gonzalez S, Rorig M, Talia P, et al. Purification and characterization of a GH43 β-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. Microbiol Res. 2014;169:213–220.
37. Fekete CA, Kiss L. Purification and characterization of a recombinant β-D-xylosidase from *Thermobifida fusca* TM51. Protein J. 2012;31:641–650.
38. Inacio JM, de Sa-Nogueira I. Characterization of abn2 (yxiA), encoding a *Bacillus subtilis* GH43 Arabinanase, Abn2, and its role in arabino-polysaccharide degradation. J Bacteriol. 2008;190:4272–4280.
39. Kim YA, Yoon K-H. Characterization of a *Paenibacillus woosongensis* beta-xylosidase/alpha-arabinofuranosidase produced by recombinant *Escherichia coli*. J Microbiol Biotechnol. 2010;20:1711–1716.
40. Shi H, Ding H, Huang Y, Wang L, Zhang Y, Li X, et al. Expression and characterization of a GH43 endo-arabinanase from *Thermotoga thermarum*. BMC Biotechnol. 2014;14:35.
41. Wang S, Yang Y, Zhang J, Sun J, Matsukawa S, Xie J, et al. Characterization of abnZ2 (yxiA1) and abnZ3 (yxiA3) in *Paenibacillus polymyxa*, encoding two novel endo-1,5-α-l-arabinanases. Bioresour Bioprocess. 2014;1:14.
42. Yoshida S, Hespen CW, Beverly RL, Mackie RI, Cann IKO. Domain analysis of a modular alpha-L-arabinofuranosidase with a unique carbohydrate binding strategy from the fiber-degrading bacterium *Fibrobacter succinogenes* S85. J. Bacteriol. 2010;192:5424–5436.
43. Van der Pol EC, Bakker RR, Baets P, Eggink G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl Microbiol Biotechnol. 2014;98:9579–9593.
44. Machado CB, Citadini AP, Goldbeck R, de Lima EA, Figueiredo FL, Da Silva TM, et al.

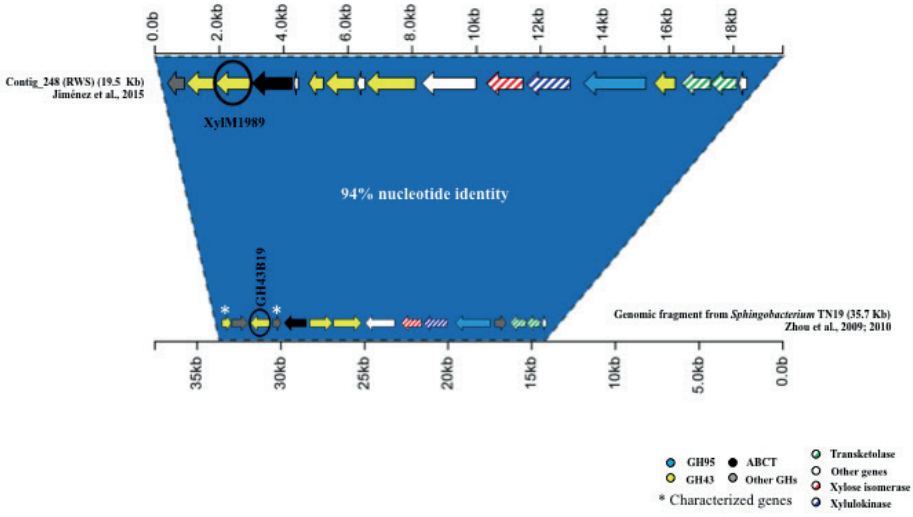
Increased biomass saccharification by supplementation of a commercial enzyme cocktail with endo-arabinanase from *Bacillus licheniformis*. Biotechnol Lett. 2015;37:1455–1462.
45. Matsuzawa T, Kaneko S, Yaoi K. Screening, identification, and characterization of a GH43 family β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome. Appl Microbiol Biotechnol. 2015;99:8943–8954.
46. Zhou J, Bao L, Chang L, Zhou Y, Lu H. Biochemical and kinetic characterization of GH43 β-d-xylosidase/α-l-arabinofuranosidase and GH30 α-l-arabinofuranosidase/β-d-xylosidase from rumen metagenome. J Ind Microbiol Biotechnol. 2012;39:143–152.
47. Moraïs S, Salama-Alber O, Barak Y, Hadar Y, Wilson DB, Lamed R, et al. Functional association of catalytic and ancillary modules dictates enzymatic activity in glycoside hydrolase family 43 β-xylosidase. J Biol Chem. 2012;287:9213–9221.
48. Santos CR, Polo CC, Costa MCMF, Nascimento AFZ, Meza AN, Cota J, et al. Mechanistic strategies for catalysis adopted by evolutionary distinct family 43 arabinanases. J Biol Chem. 2014;289:7362–7373.
49. Gong X, Gruniniger RJ, Forster RJ, Teather RM, McAllister TA. Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. Appl Microbiol Biotechnol. 2013;97:2423–2431.
50. Garrote G, Cruz JM, Moure A, Domínguez H, Parajó JC. Antioxidant activity of byproducts from the hydrolytic processing of selected lignocellulosic materials. Trends Food Sci Technol. 2004;15:191–200.
51. Klinke HB, Thomsen AB, Ahring BK. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl Microbiol Biotechnol. 2004;66:10–26.
52. Kwon Y-J, Ma A-Z, Li Q, Wang F, Zhuang G-Q, Liu C-Z. Effect of lignocellulosic inhibitory compounds on growth and ethanol fermentation of newly-isolated thermotolerant *Issatchenkia orientalis*. Bioresour Technol. 2011;102:8099–8104.
53. Du B, Sharma LN, Becker C, Chen S-F, Mowery RA, van Walsum GP, et al. Effect of varying feedstock-pretreatment chemistry combinations on the formation and accumulation of potentially inhibitory degradation products in biomass hydrolysates. Biotechnol Bioeng. 2010;107:430–440.
54. Yan QJ, Wang L, Jiang ZQ, Yang SQ, Zhu HF, Li LT. A xylose-tolerant beta-xylosidase from *Paecilomyces thermophila*: characterization and its co-action with the endogenous xylanase. Bioresour Technol. 2008;99:5402–5410.
55. Zanoelo FF, Polizeli Md Mde L, Terenzi HF, Jorge JA. Purification and biochemical properties of a thermostable xylose-tolerant beta- D-xylosidase from *Scytalidium thermophilum*. J Ind Microbiol Biotechnol. 2004;31:170–176.
56. Shi H, Li X, Gu H, Zhang Y, Huang Y, Wang L, et al. Biochemical properties of a novel thermostable and highly xylose-tolerant β-xylosidase/α-arabinosidase from *Thermotoga thermarum*. Biotechnol Biofuels. 2013;6:27.
57. Ito T, Yokoyama E, Sato H, Ujita M, Funaguma T, Furukawa K, et al. Xylosidases associated with the cell surface of *Penicillium herquei* IFO 4674. J Biosci Bioeng. 2003;96:354–359.
58. Suzuki S, Fukuoka M, Ookuchi H, Sano M, Ozeki K, Nagayoshi E, et al. Characterization of *Aspergillus oryzae* glycoside hydrolase family 43 β-xylosidase expressed in *Escherichia coli*. J Biosci Bioeng. 2010;109(2):115-117.
59. Teng C, Jia H, Yan Q, Zhou P, Jiang Z. High-level expression of extracellular secretion of a β-xylosidase gene from *Paecilomyces thermophila* in *Escherichia coli*. Bioresour Technol. 2011;102:1822–1830.
60. Chen Z, Jia H, Yang Y, Yan Q, Jiang Z, Teng C. Secretory expression of a β-xylosidase gene from *Thermomyces lanuginosus* in *Escherichia coli* and characterization of its recombinant enzyme. Lett Appl Microbiol. 2012;55:330–337.
61. Yang X, Shi P, Huang H, Luo H, Wang Y, Zhang W, et al. Two xylose-tolerant GH43

bifunctional β-xylosidase/α-arabinosidases and one GH11 xylanase from *Humicola insolens* and their synergy in the degradation of xylan. Food Chem. 2014;148:381–387.

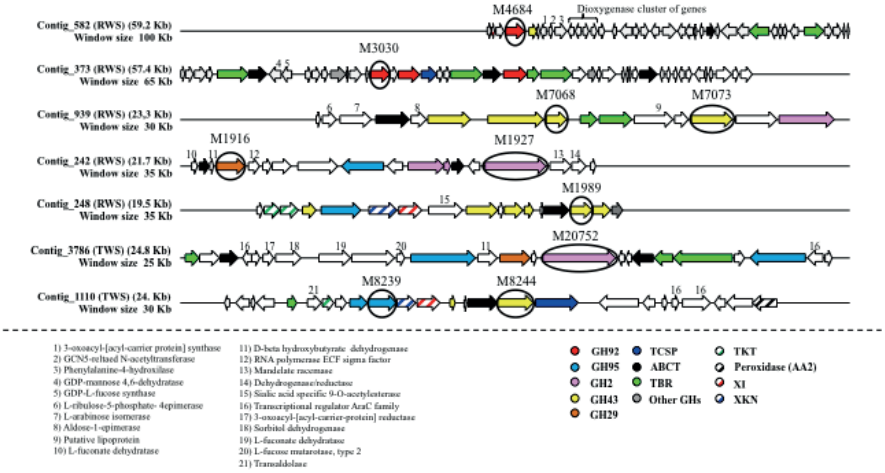
## Supplementry figures




**Supplementary Figure S2.** Restriction profile of the GHs-encoding genes cloned in the pET21b(+) expression vector .




**Supplementary Figure S3.** Genomic context of the XylM1989 in comparison with a genomic fragment from *Sphingobacterium* TN19. ABCT means ABC transporters




**Supplementary Figure S1.** Genomic context of the selected GHs-encoding genes that were subsequently used for codon optimization, synthesis, cloning and expression. Abbreviations: Two-component system proteins (TCSP), ABC transporters (ABCT), TonB-dependent receptors (TBR), transketolases (TKT), xylose isomerase (XI) and xylulose kinase (XKN) genes.

# Chapter 5

# Improved (hemi) cellulolytic enzyme cocktails for the conversion of lignocellulosic biomass into sugars.

**Mukil Maruthamuthu, Maria Julia de Lima Brossi and Jan Dirk van Elsas**

*(Submitted)*

**Abstract**
Enzymatic hydrolysis is a powerful approach for biomass conversion in the bio-refinery industries, as it allows to produce fermentable sugars from complex polysaccharides. The goal of this study was to examine if the wheat straw conversion rate of a commercial enzyme preparation could be enhanced. Different combinations of (hemi)cellulolytic enzymes (denoted P1, P2, P5 and P6), next to secretomes from wheat straw degrader consortia, were tested with the commercial enzyme Celluclast® 1.5 L. Thus the efficiency of the resulting mixtures in releasing fermentable sugars from steam-exploded (SE) versus untreated wheat straw was assessed by using dinitro salicylic acid (DNS) colorimetry and high-performance anion exchange chromatography (HPAEC). The admixture of either of two secretomes did not yield any tangible enhancement of the wheat straw degradation activity. In contrast, addition of the purified enzymes P1, P5 and/or P6 significantly enhanced the release of sugars, in particular from SE wheat straw. In these systems, xylose, glucose and arabinose were released to greater extents (50, 5 and 35%, respectively) from this substrate. Moreover, the release of sugars from SE wheat straw was significantly higher than that from untreated wheat straw. Overall, combinations of the P1 with the P5 and/or P6 enzymes significantly enhanced the Celluclast-driven enzymatic hydrolysis of SE wheat straw.

## Introduction

Lignocellulosic biomass (LCB) from agricultural waste is often proposed as a key source of energy and carbon that may serve multiple purposes, e.g. the production of biofuels or of value-added compounds [1]. For the conversion of LCB into the respective monomers, i.e. sugars, efficient enzymes are important, as they catalyze and so enhance the rate of substrate depolymerization [2,3]. Many microorganisms (bacteria and fungi) are capable of deconstructing the building blocks of LCB, i.e. cellulose, hemicellulose and pectin [4,5]. In contrast, lignin is known to be less easily decomposed. However, the use of whole microbial cells has limitations due to the often restricted availability of efficient enzymes and the scavenging, by the organisms, of the released sugars [6,7]. Nevertheless, microbial consortia that are adapted to growth on LCB constitute a highly-valued source of potentially novel enzymes [8–11]. If these offer enhancement of catalytic rates or product yields, they may subsequently be used in current cell-free LCB deconstruction approaches.

The production of efficient enzyme cocktails for LCB deconstruction will require an intricate understanding of the specific enzymes that can – jointly - deconstruct the polymeric matrix [7,12]. Most currently available commercial enzyme cocktails that are often produced on the basis of fungal (e.g. *Trichoderma reesii*)-excreted compounds are quite complex. Often, different proteins (more than 100) are present, with some of these being inefficient in the LCB degradation [13]. Moreover, the specific roles of the different components of such products in the decomposition are poorly understood [12]. Accordingly, little knowledge has been gained so far on the working of current enzyme cocktails, which are presumably often still sub-optimal [7,14]. To overcome this hurdle, such products might be enhanced using novel enzymes with proven catalytic ability or even whole secretomes of different (bacterial) origin [15].

Previous work has shown that the properties of each of the key enzymes and their relative abundances, next to potential enzyme-enzyme interactions, are key determinants of the activity of enzyme cocktails on LCB [16–21]. For instance, six fungal enzymes (with different loading concentrations) were used to promote corn stover hydrolysis following an ammonia fiber expansion pretreatment [22]. Within this cocktail, an endoglucanase, denoted EG1, played a vital role in maximizing hydrolysis yields. Barr et al (2012) developed an enzyme mixture (containing β-xylosidases (βX), β-glucosidases (βG), cellobiohydrolases (CBH), endoglucanases (EG), endoxylanases (EX) and acetylxylan esterases (AEX)), which were subsequently found to effectively degrade pretreated poplar and switchgrass biomass [23]. On the other hand, testing of the *Trichoderma reesei* derived enzymes Cel7a (cellobiohydrolase CBH1), Cel6a (CBH2), Cel7b (endoglucanase EG1), Cel5a (endoglucanase EG1), Cel12a (endoglucanase EG3) and Xyn11a (xylanase) for their activity on hydrothermally-pretreated wheat straw revealed that hydrolysis was still quite incomplete and further enhancement was required [24]. When these

enzymes were applied on SE wheat straw, Cel7b, Cel5a and Xyn11a turned out to be the key hydrolytic enzymes that promoted the release of sugars [25]. Here, the SE pretreatment was a key asset, as it made the substrate more accessible to specific enzymes.

In the current study, we selected two secretomes from wheat-straw degrading (hemi)cellulolytic microbial consortia [26], next to four biochemically-characterized enzymes with (hemi) cellulase activity obtained from a recent study [27]. The latter consisted of one β-galactosidase (P1; GH2 family), one bifunctional β-xylosidase/α-arabinosidase (P2; GH3) and two α-glucosidases (P5 and P6). We examined the potential roles of all of these enzyme sources in the deconstruction of either untreated or SE wheat straw. The underlying hypothesis of this study was that the addition of efficient enzymes (either purified proteins or secretomes) to the commercially-available enzyme preparation Celluclast enhances the release of reducing sugars from the substrate, thus paving the way towards improved enzymatic hydrolysis strategies.

## Materials and methods

### Substrates

Wheat straw was collected from a local farm in Groningen, The Netherlands. The plant biomass raw material was transported to the laboratory (<24 h following sampling) at room temperature (T = 20 °C) for further processing. The raw material was air-dried at 50 °C for 24 h before grinding using a hammer mill, yielding pieces <1 mm. Part of the wheat straw was pretreated using steam explosion, whereas the remaining part was kept as non-treated ('raw') wheat straw. The pretreatment was performed in a laboratory-scale steam explosion machine, at the University of Groningen, The Netherlands. Briefly, 30 g of the pre-dried and milled straw was soaked in 70 mL of sterile distilled water. The treatment was carried out at 180°C, with an incubation time of 15 min before explosion. The pretreated material was dried at 50°C for 24h. A total amount of 300 g of material was produced in various steam explosion rounds. Thus, two different substrates were used in the analysis: non-treated wheat straw (NTWS) and steam-exploded wheat straw (SEWS).

### Secretomes

In this study we tested two different secretomes. Extractions of extracellular protein fractions (secretomes) from the microbial consortia (10 enrichments of soil microbial community in wheat straw as unique carbon source) were performed as described in Jiménez et al. (2014a). This source consortium was then bred further, for up to 6 days, on either NTWS and SEWS (1%, as carbon and energy sources) in MSM medium [7 g/l $Na_2HPO_4$; 2 g/l $K_2HPO_4$; 1 g/l $(NH_4)SO_4$; 0.1 g/l $Ca(NO_3)_2$; 0.2 g/l $MgCl_2$, pH 7.5] supplemented with either NTWS or SEWS (1%, as carbon and energy sources. Both consortia presented progressively increasing

cell densities, from the inoculum level (around 5 log cells/ ml) to around 9 log bacterial cells/ml after 5-6 days. The resulting cultures were centrifuged (12,000xg, 10 min; Eppendorf centrifuge, Hamburg, 351 Germany) to remove cells, and the supernatants passed through 0.22-μm syringe filters (Whatman FP30/0.22, Little Chalfont, UK). This yielded two secretomes, denoted SNT (secretome of non-treated) and SSE (secretome of steam-explosion treated), respectively.

**Selection of novel enzymes**

Four novel enzymes were tested in this study, i.e. P1 (GH2 family, β-galactosidase), P2 (GH3 family, β-xylosidase/ α-arabinosidase), and P5 and P6, both with α-glucosidase activity. All had been retrieved from a metagenomic library, as described previously [11]. The enzyme P1 had a size of 116 kDa and revealed considerable sequence identity with a similar protein from *Enterobacter hormaechei* (74 %), the P2 enzyme was 85 kDa and had 83% of sequence identity with a similar protein from *Enterobacter mori*. Moreover, the enzymes P5 (45 kDa) and P6 (26 kDa) revealed sequence identity with proteins from *Enterobacter cloacae* (79 %) and *Hyphomonas neptunium* (65 %). In our study, we thus tested combinations of the secretomes (SNT and SSE) and the four enzymes with the commercially available enzyme preparation Celluclast 1.5L (Cellulases from *Trichoderma reesei* ATCC 26921; Sigma – Aldrich).

**Production and purification of enzymes**

*Escherichia coli* cells containing plasmid Origami2(DE3) pLysS (Novagen, Amsterdam, The Netherlands) loaded with genes for each of the four enzymes were grown in kanamycin (50ug/ml)- supplemented 2X-PY medium (2 ml; 16g Bacto-tryptone, 10g yeast extract, 10g NaCl/1L, pH 8.0) at $37^0$C (220 rpm, shaking incubator, overnight). Then, aliquots of the grown cultures were transferred to fresh 200mL 2xPY, setting an initial OD600 of 0.05, after which the cultures were incubated at $37^0$C (220 rpm) until they reached an $OD_{600}$ of 0.5-0.6. Subesequently, the cultures were incubated for 1 h at $18^0$C (shaking at 220rpm), to induce expression by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.5M. Following this, the cultures were further grown at $18^0$C for 16-19 h, after which cells were harvested (3,000-4,000xg; $4^0$C, 15 min). The pellets were resuspended in 5 mL of lysis buffer (50mM HEPES, pH 8.0, 300mM NaCL, 50μL 1M DTT, 1 protease inhibitor mini tablet (Roche, Sigma-Aldrich Chemie B.V, Zwijndrecht, The Netherlands) and kept on ice for 15 min. Then, cells were disrupted using sonification with the following parameters (40 cycles – 6s ON/ 15s OFF – amplitude 6-10 μm. After disruption, the resulting cell lysates were centrifuged at 15,000g for 15 min at $4^0$C. The supernatants were removed and stored (and the pellets were discarded), and 10 μL supernatants checked for protein content with 12% SDS-PAGE. The preparation was then heated to $60^0$C for 15-20 min and centrifuged again at 15000xg to remove insoluble debris. Purification of the his-

tagged proteins from the crude extracts was then carried out by gravity-flow chromatography through agarose. Thus, 600ul of Ni-NTA Agarose (Qiagen, Hilden, Germany) was added to 10 ml of lysis buffer. Incubation was for 5 min (shaking, $4^0$C), before the mixture was centrifuged for 5 min at 800g at $4^0$C. The supernatant was discarded and then 10 ml of equilibration buffer (50mM HEPES and 300mM NaCl) was added and incubated as above, before centrifugation. The crude extracts were added to the resin and incubated for 1 h as above, before being transferred to a gravity flow column and incubated at $4^0$C until the resin bed settled down. The cell-free lysates were removed by gravity flow and unbound proteins removed 3 times with 10 ml of wash buffer (50mM HEPES, 300mM NaCl, 20mM Imidazole). The bound enzyme was then eluted with 3 ml of elution buffer (50mM HEPES, 300mM NaCl, 400mM Imidazole). The resulting preparations were then concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore, Amsterdam, The Netherlands). Finally, the protein concentrations were determined using the Quick Start Bradford protein 1x dye reagent (Biorad, Veenendaal, The Netherlands) with bovine serum albumin (BSA) as a standard [28] using the Infinite M200®spectrophotometer (Tecan, Männedorf, Switzerland) at 595 nm. The purity of the preparations was analyzed using 12% SDS-PAGE [29].

The secretomes collected from the degrader consortia, and the Celluclast preparation were found to contain sugars. In order to remove these, we used Amicon® Ultra-15 Centrifugal Filter Devices (Millipore, Amsterdam, The Netherlands) on 12 ml of both. Following a spin at 5000xg for 60 min, solutes were recovered from the resulting preparations and stored at -20°C and the protein concentrations were determined.

**Enzymatic activity assays**

Enzymatic activity micro-assays were carried out to evaluate the effects of the selected (hemi) cellulolytic enzyme mixes. Three experiments (1, 2 and 3) were performed, in accordance with critical choices made along the work. The first experiment had 13 enzyme combinations, the second one 10 and the third one 5 (in each case, combining novel enzymes and/or secretomes with Celluclast). Thus, we analyzed the effect of: (1) treated versus non-treated wheat straw (2) the use of two secretomes, and effect of their origin, and (3) four selected enzymes with different activities. All treatments consisted of three replicates and were controlled with two negative controls (each substrate plus buffer). The independent variables were P1, P2, P5 and P6 (four levels), secretomes SNT and SSE (two levels), substrates NTWS and SEWS (two levels) and the enzymatic cocktail Celluclast 1.5L. The protein concentrations were set at 1 mg of proteins g substrate$^{-1}$ for all proteins, secretomes and the Celluclast. The reaction mixtures contained 20 mg of each substrate plus the enzyme combinations (Table 1) in sodium phosphate buffer (0.1 M, pH 6.0) in a final volume of 1.5ml. With respect to the experimental

design, '-' corresponds to the absence of the enzyme/secretome and '+' to its presence. All assays were done at 50°C (24h, shaking at 250rpm). After incubation, the mixtures were centrifuged (12000xg for 15 min at 4°C), and the supernatants collected for subsequent analyses. The amount of reducing sugars in the supernatants was measured according to the dinitro salicylic acid (DNS) colorimetric [30] method, and in specific cases by high pH anion exchange chromatography. All data are presented as total sugars released (mg/mL).

**Analysis of released sugars by high pH anion exchange chromatography (HPAEC)**

Samples from the third experiment were subjected to analysis by HPAEC on a Dionex ICS-3000 system (Thermo Scientific) equipped with a CarboPac PA-1 guard column (2x50 mm) and a CarboPac PA-1 column (2x250 mm). An internal pulsed amperometric detector with an Au working electrode and an Ag/ AgCl pH reference electrode was used for detection. The system was run with 25 mM sodium hydroxide (NaOH) for 3 min, followed by a gradient up to 85 mM NaOH at 17 min and subsequently a gradient to 100 mM NaOH and 30 mM of sodium acetate at 22 min and finally a gradient of [100 mM NaOH] and 210 mM sodium acetate at 30 min. The flow rate of the system was 0.25 ml/min. All chromatograms were analyzed using Chromeleon 6.8 Chromatography data system software (Thermo Scientific).

To interpret the data, calibration curves encompassing arabinose, galactose, glucose (Sigma-Aldrich Chemie B.V. Zwijndrecht, Netherlands), xylose (Janssen chimica, Beerse, Belgium), xylobiose, cellobiose and xylotriose (Megazyme Inc, Illinois, USA) were made. Concentrations of the used standards ranged from 50 μM to 1 mM. Standards and samples had an injection volume of 5 μl. The calibration curves were used to quantify the amount of released saccharides from the wheat straw. The peaks at the end of the chromatogram (≤40%) probably represent decorated oligosaccharides (e.g. acetyl groups, ferulyl groups and glucoronic acid). We did not further take these oligosaccharides into account.

**Statistical treatment of the data**

All experiments were performed in triplicate. To assess the differences between treatments, we used permutational multivariate analysis of variance (PerMANOVA) [31]. Different factors were addressed in each experiment, and full permutation of the raw data was done with Monte–Carlo tests (accounting for type III error), where the fixed effects sum to zero with 9999 permutations. For the first experiment, the factors were the pre-treatment of the wheat straw (untreated versus steam explosion), the presence/absence of Celluclast, and the addition of secretomes and the addition of enzymes. For the second experiment, the factors were each enzyme separately. For all three experiments, statistical comparisons were also performed using one-way ANOVA (Tukey's test) to

compare pairwise differences between treatments.

## Results

### Steam explosion of wheat straw – effects

Steam explosion of milled (average particle dimension < 1 mm) dry wheat straw was performed in a laboratory-scale steam explosion system, using an in-house made pressure device. We first experimentally tested the effects of different steam explosion regimes, including variations in the time at maximal temperature (5, 10 and 15 min) and temperature (170, 175 and 180°C). Taking the data together, a regime of 15 min at 180°C turned out to produce optimal results (as evidenced using direct microscopy of the treated versus untreated materials, **Figure** 1a). Thus, this regime was selected for the further steam explosion processing. We produced a total of 300g of SE wheat straw, of which the structure was clearly modified as compared to the control (**Figure** 1a). In addition, the liquid remaining after steam treatment (after filtering) had a raised amount of released sugars (1.2 mg/ml), as compared to the liquid from untreated wheat straw (NTWS; 0.2mg/ml) (**Figure** 1b).

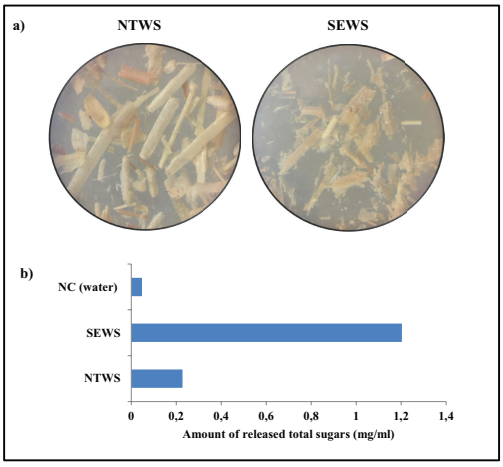



**Figure 1. a)** Steam-explosion: untreated (NTWS), treated (SEWS) b) Dinitro salicylic acid (DNS) colorimetric assay

### Metasecretomes produced from two biodegrader consortia – yield and purity

The filtered supernatants of the NTWS and SEWS consortia were concentrated using Amicon centrifugal devices, giving 0.3 and 0.1 µg of protein/µl, respectively. As shown in a previous study [32] the presence of enzymes belonging to glycosyl hydrolase families GH3, GH10, GH43, GH51, GH67 and GH95 was predicted (**Figure** 2), thus the existence of (hemi)cellulolytic and other carbohydrate-active proteins in the liquid was plausible.

### Selection of four enzymes - yield and purity

The enzymes encoded P1, P2, P5 and P6 were all successfully produced from the freshly-grown *E. coli* cultures. For all enzymes, we obtained maximally (approximately 80%) pure enzyme preparations after his-tag purification. The protein concentrations were: P1 (0.4 µg/µl), P2 (0.7 µg/µl), P5 (0.5 µg/µl) and P6 (0.7 µg/µl). Using enzymatic activity assays with *p*NP-beta-D-galactopyranoside (*p*NPGal), *p*NP-beta-D-xylanopyranoside (*p*NPXyl), *p*NP -alpha-L-

arabinofuranoside and *p*NP-alpha-D-glucopyranoside (*p*NPGlu), we confirmed the biochemical activities of the four enzymes as (thermoalkaliphilic) β-galactosidase (P1), β-xylosidase and α-arabinosidase (P2), and α-glucosidase (P5 and P6) (data not shown).

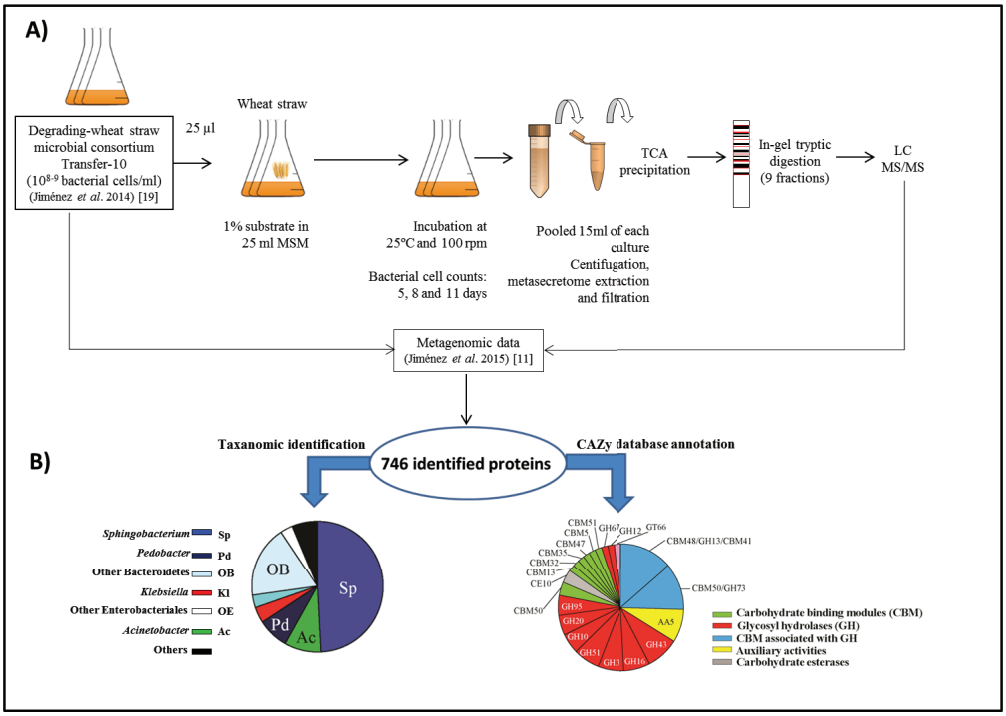



**Figure 2. a)** Schematic representation of the methodology used in this study **b)** Taxonomic assignment of detected proteins in the metagenome database, and functional assignment of proteins detected in metasecretome.

## Steam-exploded versus untreated wheat straw – enzymatic hydrolysis

Whereas Celluclast served as the control, the additional effects of different combinations of enzymes P1, P2, P5 and P6, and of secretomes SNT and SSE, were evaluated on both NTWS and SEWS. As extra controls, *i)* buffer, substrate (NTWS and/or SEWS) and sodium azide, and *ii)* only buffer, were used. Thus, initially 13 combinations of enzymes (Table 1) were used (**Figure** 3a; 3b) and released sugars measured. An overall statistical analysis of the data revealed that significantly higher amounts of sugars were released, across all enzyme preparations, from SEWS than from NTWS ($P<0.05$). First, Celluclast alone (treatment 8; Table 1) released significantly ($p=0.001$) higher amounts of reducing sugars from SEWS (2.0mg/ml) than from NTWS (0.7mg/ml). With respect to the Celluclast modifications, the addition of either of the two secretomes (Table 1) to Celluclast slightly but insignificantly enhanced the amount of released sugars in

comparison to the Celluclast alone (**Figure** 3a; 3b; p>0.05). Remarkably, the addition to Celluclast of the enzymes P1, P2, P5 and P6 (treatment 9, Table 1) strongly and significantly (p=0.018) enhanced the hydrolytic activity on SEWS, with the released sugars amounting to 3.42 mg/ml as compared to 1.86 mg/ml in the Celluclast-only treatment (**Figure** 3b). Thus, the degradation of SEWS was greatly aided by the supplementation of the four-enzyme mix to the Celluclast, significantly enhancing the release of sugars. Given this strong effect of the four enzymes (P1, P2, P5 and P6), in particular on SEWS, we further analyzed the potential role of each of these enzymes with Celluclast.

**Table 1.** Experiment 1: Different combinations of four enzymes, two secretomes, and the commercial enzyme mix Celluclast

| No | Substrate (a) | Substrate (b) | Sodium azide | Secretome / 20µg | Protein 1/ 20µg | Protein 2/ 20µg | Protein 5/ 20µg | Protein 6/ 20µg | Celluclast/ 20µg | Buffer (final volume) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | - | - | - | - | - | Upto 1,5 mL |
| 2 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | - | - | - | - | - | Upto 1,5 mL |
| 3 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | + | - | - | - | - | Upto 1,5 mL |
| 4 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | - | + | - | - | - | Upto 1,5 mL |
| 5 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | - | - | + | - | - | Upto 1,5 mL |
| 6 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | - | - | - | + | - | Upto 1,5 mL |
| 7 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | + | + | + | + | - | Upto 1,5 mL |
| 8 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | - | - | - | - | + | Upto 1,5 mL |
| 9 | 20 mg NTWS | 20 mg SEWS | 10 uL | - | + | + | + | + | + | Upto 1,5 mL |
| 10 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | - | - | - | - | + | Upto 1,5 mL |
| 11 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | - | - | - | - | + | Upto 1,5 mL |
| 12 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | + | + | + | + | + | Upto 1,5 mL |
| 13 | 20 mg NTWS | 20 mg SEWS | 10 uL | + | + | + | + | + | + | Upto 1,5 mL |

**Selection of optimal enzyme mixture for release of sugars from SEWS**

Based on the results obtained from the initial experiment, we selected the SEWS substrate and (different combinations of) each of the four enzymes (P1, P2, P5 and P6) for further analyses of their potential as Celluclast enhancers. Table 2 lists the experimental conditions, whereas the data from this second experiment are shown in **Figure** 4a. The addition of the P1 preparation to the Celluclast significantly (p=0.05) enhanced the release of sugars, i.e. a ~20% increase was found compared to the Celluclast alone treatment (3.90 mg/ml versus 3.27 mg/ml). In contrast, none of the other individual enzymes, i.e. P2, P5 and P6, significantly enhanced the activity of Celluclast. Remarkably, the combination of enzyme P1 with both P5 and P6 significantly (p=0.001) enhanced the Celluclast activity (Table 2: Treatment 8) by ~22%, with respect to the amount of released sugars (**Figure** 4a). Specifically, 3.98 mg/ml was released versus 3.27 mg/ml in the control. We thus concluded that a synergistic action of enzymes P1 with P5 and/or P6 was at the basis of the activity enhancement, and selected these three enzymes to further unravel their roles as Celluclast enhancers (Table 3). In the resulting experiment, the three-enzyme mix (P1+P5+P6) added to Cellucalst (~3.7

mg/ml of released sugars) again revealed a significant (roughly 25%) increase over the Celluclast-alone treatment (~ 2.95 mg/ml of released sugars). Remarkably, the P1+P5 and P1+P6 mixes added to the Celluclast produced even more sugars (~ 4.45 mg/ml) than the Celluclast-alone (**Figure** 4b), amounting, in both cases, to 50 and 51% increases. These increases were highly significant (p=0.0001).

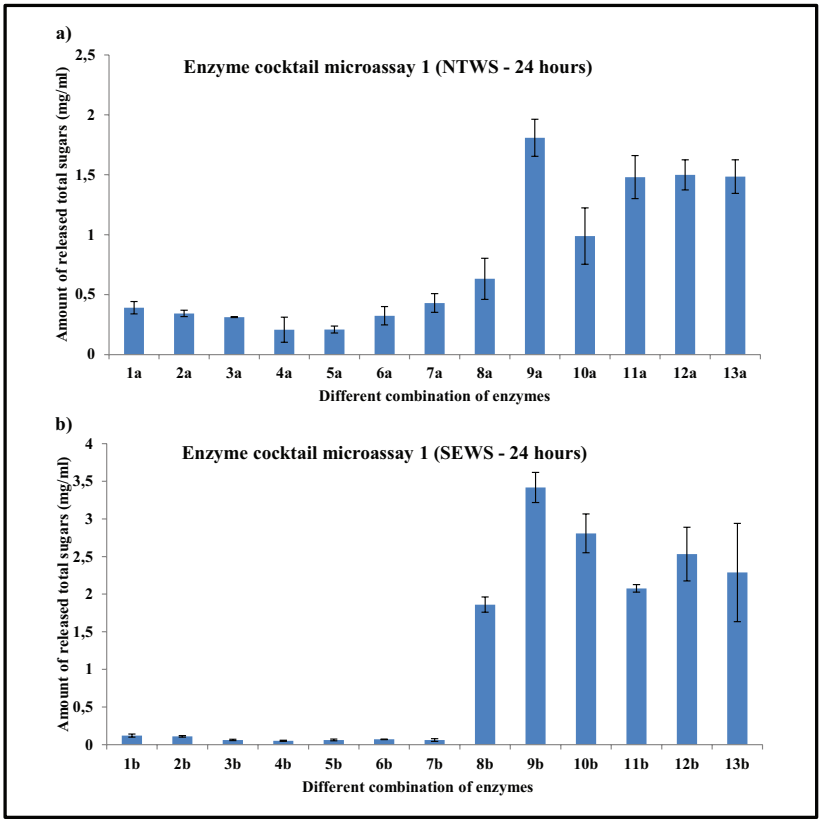



**Figure 3.** Experiment 1, Dinitro salicylic acid (DNS) colorimetric analysis a) with NTWS b) with SEWS

## Analysis of released sugars by HPAED

We then analyzed the nature of the released sugars as a result of the treatments, with a focus on the SEWS. The Celluclast-only treatment (Treatment1; Table 3) was found to release considerable amounts of arabinose (0.3 μg), glucose (1.1 μg) and xylose (0.5 μg), but no cellobiose and xylotriose. In contrast, treatments 2 (P1+P5+P6+Celluclast; Table 3), 3 (P1+P5+Celluclast; Table 3), 4 (P1+P6+Celluclast; Table 3) and 5 (P5+P6+Celluclast; Table 3) revealed the presence of increased levels of the former three sugars, next to other ones (i.e. Oligosaccharides and xylobiose) (**Figure** 5). Moreover, the release of xylose and arabinose was significantly higher in treatments 2 to 5 than in the control treatment 1. In particular, treatment 3 produced higher amounts of xylose (2.1 μg)

and 1.5 μg of arabinose versus 0.5 μg and 0.3 μg in the control treatment 1. In addition, the presence of glucose (1.25, 1.21 and 1.28 μg) was found to be statistically raised ($p<0.01$) in treatments 3 to 5 when compared to control treatment 1 (1.1 μg). This suggested that, whereas the Celluclast likely cleaved primarily cellulose moieties, the additional enzymes worked mainly as (hemi)cellulases, thus increasing the conversion rates. Overall, the HPAEC data (**Figure** 5) confirmed that the addition of the enzyme P1 preparation together with P5 and P6 significantly enhanced the Celluclast activity by enhancing the release of the monosaccharides xylose, glucose and arabinose from the (hemi)cellulose matrix.

**Discussion**

Given the fact that current enzyme cocktails are as-yet suboptimal on LCB, an enhancement of the hydrolysis rates of such cocktails is desirable [33,34]. In this regard, the development of hemicellulases for the enzymatic hydrolysis of plant biomass is not as advanced as that of cellulases, because current commercial preparations have been primarily used on pretreated biomass from which the hemicellulose part was partially removed before saccharification [35]. Moreover, current fungal-derived cellulases tend to have only weak hemicellulolytic activity and so are not adequate for the complete conversion of LCB when it comes to the hemicellulose moiety. Therefore, the development of low-cost commercial hemicellulases that work synergistically with cellulases is a prime goal of current research [36]. We here tested two supplementary enzyme sources [four (hemi)cellulolytic enzymes (P1, P2, P5 and P6), and the secretomes of NTWS and SEWT bred microbial consortia], in conjunction with the commercially available Celluclast 1.5L, to increase the (untreated versus SE) wheat straw degradation efficiency [37]. The SE pre-treatment process was thought to unlock the different (cellulose, hemicellulose, lignin) moieties constituting the wheat straw fibers, thus facilitating enzyme access [38]. Indeed, our steam-exploded material had clear signs of structural modifications, as predicted [25]. Therefore, we presumed that SE was highly important as a physical mechanism that disrupted bonds in some of the fibers, thus allowing (hemi)cellulose fractions to detach from the rigid parts of cellulose and lignin. Our subsequent sugar release data clearly indicated that the SEWS, as compared to the NTWS, had enhanced susceptibility to enzymatic attack, enabling the release of significantly higher amounts of sugars (**Figure** 3a). As optimal degradation of LCB depends on the synergistic action of different enzymes [36] that can be found in lignocellulose-grown microbial consortia, we included two metasecretomes produced from microbial consortia grown on wheat straw (NTWS and SEWS). As hemicellulolytic enzymes, i.e. endo-1,4-beta-xylanases (GH10), beta-xylosidases (GH43), alpha-L-arabinofuranosidases (GH43 or GH51) and alpha-L-fucosidases (GH95), may critically aid cellulases in the degradation of wheat straw [11,32], secretomes were selected which predictably

had (hemi)cellulolytic enzymes of families GH43, GH16, GH3, GH51, GH10, GH20, GH95, GH12 and GH67 [32]. However, none of the two secretomes was a clear Celluclast enhancer, although SNT and SSE did show some minor effects.

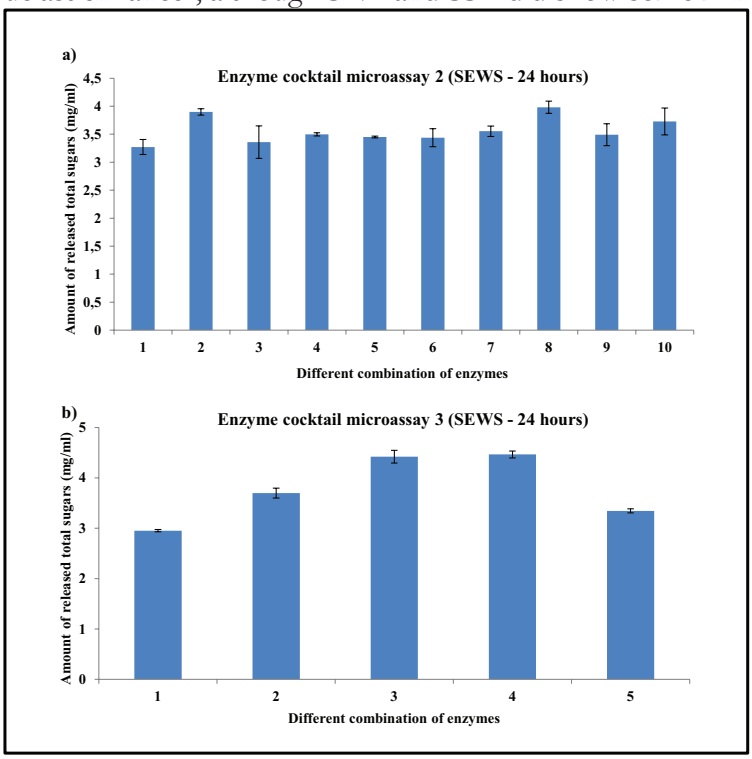



**Figure 4.** Experiment 2 and 3, Dinitro salicylic acid (DNS) colorimetric analysis a) Enzymes (P1, P2, P5 and P6) cocktail assay with SEWS b) Enzymes (P1, P5 and P6) cocktail assay with SEWS

**Table 2.** Enzyme cocktail microassay 2: Different combinations of purified enzymes and commercial enzyme celluclast

| No | Substrate | Sodium azide | Protein 1/ 20µg | Protein 2/ 20µg | Protein 5/ 20µg | Protein 6/ 20µg | Celluclast/ 20µg | Buffer (final volume) |
|---|---|---|---|---|---|---|---|---|
| 1 | 20 mg SEWS | 10 uL | - | - | - | - | + | Upto 1,5 mL |
| 2 | 20 mg SEWS | 10 uL | + | - | - | - | + | Upto 1,5 mL |
| 3 | 20 mg SEWS | 10 uL | - | + | - | - | + | Upto 1,5 mL |
| 4 | 20 mg SEWS | 10 uL | - | - | + | - | + | Upto 1,5 mL |
| 5 | 20 mg SEWS | 10 uL | - | - | - | + | + | Upto 1,5 mL |
| 6 | 20 mg SEWS | 10 uL | + | + | + | - | + | Upto 1,5 mL |
| 7 | 20 mg SEWS | 10 uL | + | + | - | + | + | Upto 1,5 mL |
| 8 | 20 mg SEWS | 10 uL | + | - | + | + | + | Upto 1,5 mL |
| 9 | 20 mg SEWS | 10 uL | - | + | + | + | + | Upto 1,5 mL |
| 10 | 20 mg SEWS | 10 uL | + | + | + | + | + | Upto 1,5 mL |

**Table 3.** Enzyme cocktail microassay 3: Different combinations of selected three purified enzymes and commercial enzyme celluclast

| No | Substrate | Sodium azide | Protein 1/ 20µg | Protein 5/ 20µg | Protein 6/ 20µg | Celluclast/ 20µg | Buffer (final volume) |
|---|---|---|---|---|---|---|---|
| 1 | 20 mg SEWS | 10 uL | - | - | - | + | Upto 1,5 mL |
| 2 | 20 mg SEWS | 10 uL | + | + | + | + | Upto 1,5 mL |
| 3 | 20 mg SEWS | 10 uL | + | - | + | + | Upto 1,5 mL |
| 4 | 20 mg SEWS | 10 uL | + | + | - | + | Upto 1,5 mL |
| 5 | 20 mg SEWS | 10 uL | - | + | + | + | Upto 1,5 mL |

Secretomes of biodegrader microbial consortia are complex mixtures of several secreted proteins, including a suite of different glycosyl hydrolases. However, other microbial proteins may act as antagonists or degraders of the secreted enzymes, and so the actual activities of key glycosyl hydrolases may have been hampered. Given that the combination of secretome with Celluclast induced the conversion rate of SEWS up to 12%, increasing the concentration of particular secretome proteins in the Celluclast might increase the LCB degradative activity. Another limiting factor may have been the adsorption of the enzymes to the SEWS-exposed lignin and cellulose moieties [39]. The catalytic functions of the enzymes involved are controlled by structural conformation [40], and conformational modifications may come about as a result of the binding of small molecules, such as amino acids or nucleotides. For example "allosteric inhibition", i.e. the binding of such a regulatory molecule, by changing the conformation of the protein structure, can swap the shape of the catalytic site and affect catalytic activity [40,41]. Such factors, although unstudied, may have reduced the activities of the secretomes.

Collectively, our findings highlight that key gains in enzymatic activity of a commercial enzyme preparation like Celluclast can be obtained by an educated guess with respect to additional enzymatic activities that may be required. Thus, the P1, P5 and P6 enzymes, when added to Celluclast, significantly increased the hydrolytic activity on SE wheat straw. Importantly, enzyme P1, which belongs to glycosyl hydrolase family 2 (GH2), was predicted to exert β-galactosidase activity, and P5 (and P6) were predicted to have α-glucosidase activities [42]. Interestingly, the β-galactosidase activity of P1, in concert with the Celluclast, appeared to spur the release of extra glucose, but not of galactose. It might be that the raw wheat straw has very low levels (e.g. 1-2%) of galactose, or that the prior steam explosion had removed it [43]. A previous study showed that Celluclast constitutes an effective enzyme mixture for cellulose and hemicellulose degradation, exhibiting synergistic interaction that fosters the release of arabinose and xylose from soluble

wheat arabinoxylan [44]. Here, we hypothesized that the steam explosion treatment of the wheat straw reduced the traffic of cellulose moieties; thus Celluclast would easily attack the 1-4 linkage of D-glucose, releasing glucose. In addition, Celluclast may have interacted with the hemicellulose moiety, e.g. the xylan, on the b-1-4 link of D-xylose (main chain), thus releasing xylose as well as arabinose from the xyloglucan backbone (side chain). The additive effect of enzyme P1 (β-galactosidase) might be explained by its entering the hemicellulose moiety, in which side chains of xylose-glucose (of xyloglucan) were broken, increasing the release of xylose and glucose. Along with it, enzymes P5 and P6 (α-glucosidases) might have cleaved the region of 1→4 linked a-glucose residues to release a-glucose molecules. Therefore, the above findings indicate that our (hemi)cellulases assisted the Celluclast activity, enabling the release of enhanced sugars by cleaving hemicellulose parts. On the other hand, the untreated wheat straw consists of 19 to 21% xylan, and so (next to glucose) we recovered more xylose (and arabinose) in the combination of P1 with P5 or P6 (**Figure** 5).

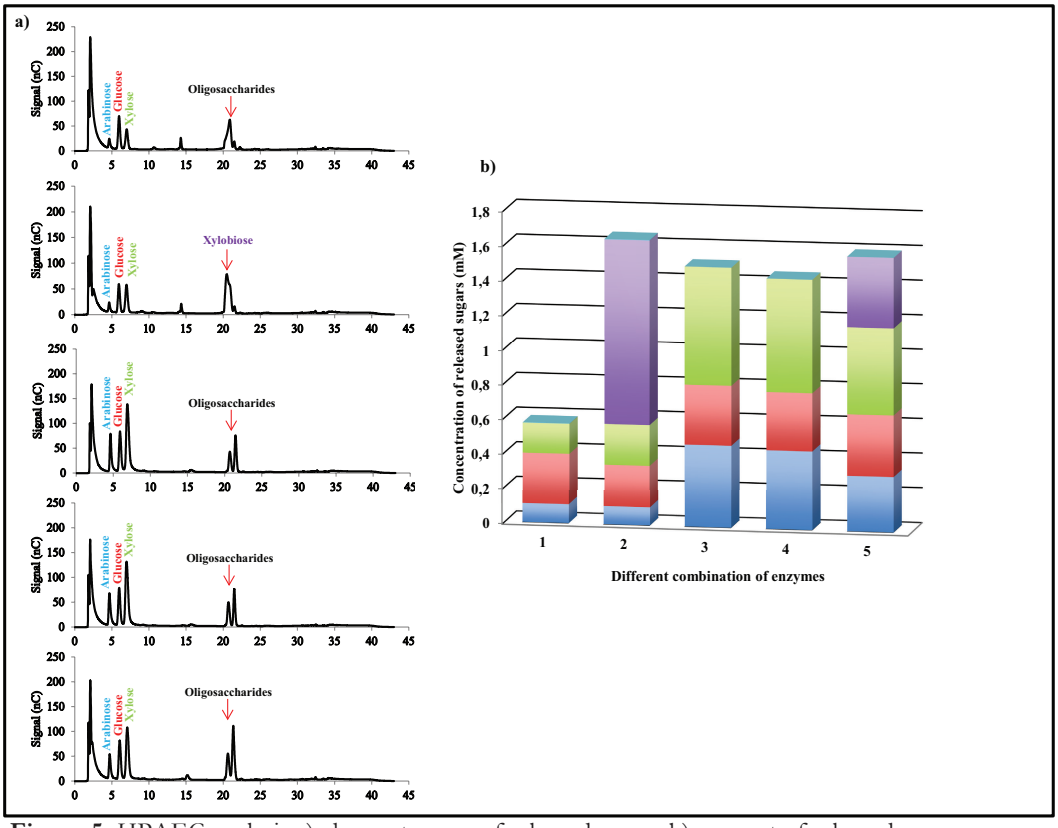



**Figure 5.** HPAEC analysis a) chromatogram of released sugars b) amount of released sugars

## Conclusion

In this study, we searched for enzymes that would be able to significantly enhance

the activity of the commercial enzyme mixture Celluclast on untreated versus steam-exploded wheat straw. Clearly, three enzymes denoted P1 (β-galactosidase, Tm 50°C, pH- 8.0, highly stable for 120 min), P5 (α-glucosidase, Tm 50°C, pH-10.0, highly stable for 120 min), and P6 (α-glucosidase, Tm 40°C, pH-10.0, highly stable for 120 min), significantly enhanced the degradation of SE wheat straw under the conditions (50 C; 24 h) used here. Thus the addition of the enzyme mixes P1+P5 and/or P1+P6 to the Celluclast enhanced the release of xylose (303%), arabinose (343%) and glucose (13%) from SE wheat straw, as compared to Celluclast alone.

**Authors' contributions**

MM designed, constructed the experiments and drafted the manuscript. MJL Brossi did enzyme cocktails assay. JDvE conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We thanks to H. Ruijssenaars and R. van Kranenburg for scientific support. Further thanks to G.A. ten Kate for HPAEC experimental support. This work was supported by the Netherlands Ministry of Economic Affairs and the BE-Basic partner organizations (http://www.be-basic.nl).

**References**

1. Goldbeck R, Damásio ARL, Gonçalves TA, Machado CB, Paixão DAA, Wolf LD, et al. Development of hemicellulolytic enzyme mixtures for plant biomass deconstruction on target biotechnological applications. Appl Microbiol Biotechnol. 2014;98:8513–8525.
2. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol. 2013;135:513–522.
3. Dashtban M. Fungal Bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci. 2009;578–595.
4. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol. 2013;79:1545–1554.
5. Brown ME, Chang MC. Exploring bacterial lignin degradation. Curr Opin Chem Biol. 2014;19:1–7.
6. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, et al. How biotech can transform biofuels. Nature Biotechnol. 2008;26:169–172.
7. Banerjee G, Scott-Craig JS, Walton JD. Improving enzymes for biomass conversion: A basic research perspective. BioEnergy Res. 2010;3:82–92.
8. Jiménez DJ, Korenblum E, Van Elsas JD. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl Microbiol Biotechnol. 2014;98:2789–2803.
9. Hui W, Jiajia L, Yucai L, Peng G, Xiaofen W, Kazuhiro M, et al. Bioconversion of un-pretreated lignocellulosic materials by a microbial consortium XDC-2. Bioresour Technol. 2013;136:481–487.
10. Yang H, Wu H, Wang X, Cui Z, Li Y. Selection and characteristics of a switchgrass-colonizing

microbial community to produce extracellular cellulases and xylanases. Bioresour Technol. 2011;102(3):3546-3550.
11. Maruthamuthu M, Jiménez DJ, Stevens P, Van Elsas JD. A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. BMC Genomics. 2016;17:86.
12. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels. 2010;3:22.
13. Sørensen HR, Pedersen S, Meyer AS. Synergistic enzyme mechanisms and effects of sequential enzyme additions on degradation of water insoluble wheat arabinoxylan. Enzyme Microb Technol. 2007;40:908–18.
14. Kumar R, Wyman CE. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. Bioresour Technol. 2009;100:4203–4213.
15. Jiménez DJ. Plant biomass-degrading microbial consortia. 2016.
16. Gusakov A V., Salanovich TN, Antonov AI, Ustinov BB, Okunev ON, Burlingame R, et al. Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose. Biotechnol Bioeng. 2007;97:1028–1038.
17. De Lima EA, Machado CB, Zanphorlin LM, Ward RJ, Sato HH, Ruller R. GH53 endo-beta-1,4-galactanase from a newly isolated bacillus licheniformis cbmai 1609 as an enzymatic cocktail supplement for biomass saccharification. Appl Biochem Biotechnol. 2016;179:415–426.
18. Karnaouri A, Matsakas L, Topakas E, Rova U, Christakopoulos P. Development of thermophilic tailor-made enzyme mixtures for the bioconversion of agricultural and forest residues. Front Microbiol. 2016;7:177.
19. Lima MS, Damasio AR de L, Crnkovic PM, Pinto MR, Da Silva AM, Da Silva JCR, et al. Co-cultivation of *Aspergillus nidulans* recombinant strains produces an enzymatic cocktail as alternative to alkaline sugarcane bagasse pretreatment. Front Microbiol. 2016;7:583.
20. Méndez Arias J, Modesto LFA, Polikarpov I, Pereira N. Design of an enzyme cocktail consisting of different fungal platforms for efficient hydrolysis of sugarcane bagasse and synergism studies. Biotechnol Prog. 2016;32(5):1222-1229.
21. Vasconcellos VM, Tardioli PW, Giordano RLC, Farinas CS. Addition of metal ions to a (hemi)cellulolytic enzymatic cocktail produced in-house improves its activity, thermostability, and efficiency in the saccharification of pretreated sugarcane bagasse. Nature Biotechnol. 2016;33:331–337.
22. Gao D, Chundawat SPS, Krishnan C, Balan V, Dale BE. Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. Bioresour Technol. 2010;101:2770–2781.
23. Barr CJ, Mertens JA, Schall CA. Critical cellulase and hemicellulase activities for hydrolysis of ionic liquid pretreated biomass. Bioresour Technol. 2012;104:480–485.
24. Szijártó N, Siika-aho M, Sontag-Strohm T, Viikari L. Liquefaction of hydrothermally pretreated wheat straw at high-solids content by purified *Trichoderma* enzymes. Bioresour Technol. 2011;102:1968–1974.
25. Billard H, Faraj A, Lopes Ferreira N, Menir S, Heiss-Blanquet S. Optimization of a synthetic mixture composed of major *Trichoderma reesei* enzymes for the hydrolysis of steam-exploded wheat straw. Biotechnol Biofuels. 2012;5:9.
26. Jiménez DJ, Korenblum E, Van Elsas JD. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl Microbiol Biotechnol. 2014;98:2789–2803.
27. Maruthamuthu M, Van Elsas JD. Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library. Biotechnol Biofuels. 2017;10:142.

28. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248–254.
29. Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V, et al. Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. Biosci Biotechnol Biochem. 2012;76:1075–1084.
30. Miller GL. Use of Dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem. 1959;31(3):426–428
31. Clarke KR, Gorley RN. PRIMER version 6: user manual/tutorial. Prim Plymouth. 2006;192.
32. Jiménez DJ, Maruthamuthu M, Van Elsas JD. Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. Biotechnol Biofuels. 2015;8:199.
33. Rocha GJM, Gonçalves AR, Oliveira BR, Olivares EG, Rossell CEV. Steam explosion pretreatment reproduction and alkaline delignification reactions performed on a pilot scale with sugarcane bagasse for bioethanol production. Ind Crops Prod. 2012;35:274–279.
34. Canettieri EV, Rocha GJ de M, de Carvalho JA, de Almeida e Silva JB. Optimization of acid hydrolysis from the hemicellulosic fraction of Eucalyptus grandis residue using response surface methodology. Bioresour Technol. 2007;98:422–428.
35. Meyer AS, Rosgaard L, Sørensen HR. The minimal enzyme cocktail concept for biomass processing. J Cereal Sci. 2009;50:337–344.
36. Mohanram S, Amat D, Choudhary J, Arora A, Nain L. Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries. Sustain Chem Process. 2013;1:15.
37. Kovacs K, Macrelli S, Szakacs G, Zacchi G. Enzymatic hydrolysis of steam-pretreated lignocellulosic materials with *Trichoderma atroviride* enzymes produced in-house. Biotechnol Biofuels. 2009;2:14.
38. Bragatto J, Segato F, Cota J, Mello DB, Oliveira MM, Buckeridge MS, et al. Insights on how the activity of an endoglucanase is affected by physical properties of insoluble celluloses. J Phys Chem B. 2012;116(21):6128-6136.
39. Rosgaard L, Pedersen S, Meyer AS. Comparison of different pretreatment strategies for enzymatic hydrolysis of wheat and barley straw. Appl Biochem Biotechnol. 2007;143:284–296.
40. Cooper GM. Regulation of protein function. Sinauer Associates. 2000.
41. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Protein function. Mol Biol Cell. 2002.
42. Maruthamuthu M, Van Elsas JD. Molecular cloning, expression and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library. Biotechnol Biofuels. 2017;10:142.
43. Lee D, Owens VN, Boe A, Jeranyama P. Composition of herbaceous biomass feedstocks. Cellulose. 2007;16.
44. Sørensen HR, Meyer ABS, Pedersen S. Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between alpha-L-arabinofuranosidases, endo-1,4-beta-xylanases, and beta-xylosidase activities. Biotechnol Bioeng. 2003;81(6):726-731.

# Chapter 6

# General discussion, conclusion and future prospective

Mukil Maruthamuthu

## 1. Introduction

The objective of this study was to examine existing wheat straw (WS) degradation strategies on the basis of traditional methods coupled to those developed on the basis of metagenomics approaches. In order to do so, a first focus needed to be placed on WS quality, as this was thought to strongly determine the type of degradation strategy that might be optimal. Moreover, variation in WS quality was deemed to be a key factor. In the following section, I will discuss the relevance of WS quality and variability therein, with the implications for its degradability. Then, I discuss my achievements with respect to physical pretreatment of WS, metagenomics and the possibilities of the newly found enzymes.

## 2. Wheat straw – Quality, structure, composition and implications for degradation

Wheat straw is produced in billions of tons every year throughout the world. It has good utilization potential but has not been fully explored. Raw wheat straw (RWS) includes cellulose (30-50%), hemicellulose (15-33%) and lignin (5-26%) fractions, next to other materials (such as pectin, sugars, acids and proteins) [1]. Here, I opted for the use of WS harvested from a Dutch field as the target material for degradation improvement.

### *2.1. Variation in chemical composition*

WS has generally a high cellulose content, similar to rice straw (RS) and poplar. The hemicellulose content of WS is approximately similar to that of rice straw but 1.45 times higher than that of poplar [2]. As is obvious from the ranges given in the foregoing, a huge variation exists in the chemical composition of WS, as driven by various factors such as cultivation conditions, cultivar type, climate, soil type and location, as well as soil treatment (Table 1). For instance, WS from tropical summer wheat is rich in cellulose and hemicellulose moieties (37-50% and 15-33%, respectively) that are often harvested (at well-matured conditions) between summer and autumn. On the other hand, winter WS of different wheat cultivars showed variation in chemical composition even when grown under the same climatic condition but at different geographical locations (around Denmark). In the latter WS, lignin and ash content was different between pretreated and unpretreated straw (Table 1; [3]). Moreover, it contained high (19-27%) lignin contents, which provides structural strength and mechanical support to the plant stem and leaf needed for often windy winter conditions. In comparison, summer WS had a low level (5.6%) of lignin [4]. In addition, the physical parts of wheat plants (internodes, leaf-sheath, leaf-blade, nodes and grains) also showed large variation in the WS chemical composition [5]. Here, I opted for the use of a local WS, which was predicted to have high (hemi)cellulose contents, but also nearly 22% of lignin. Such raised lignin contents are often difficult to remove from the (hemi)cellulose matrix, impairing the degradation processes. Thus, to efficiently

degrade the (hemi)celluloses in our local WS, special treatment strategy had to be designed. Overall, in the light of the huge variation and complexity of the composition of WS, we are far from a thorough understanding of what would constitute generally optimal WS degradation strategies.

**Table 1:** Variation in WS chemical composition

| No | WS chemical Composition | | | | | Country/Type | Sampling season | Plant parts | Ref |
|---|---|---|---|---|---|---|---|---|---|
| | Cellulose | Hemicellulose | Lignin | Ash | Others | | | | |
| 1 | 45.5 | 31.3 | 22.2 | ND | 1 | The Netherlands | Winter | Straw | [43] |
| 2 | 31.6 | 16.2 | 18.5 | 17.2 | 16.5 | China | Summer | Straw | [44] |
| 3 | 41.3 | 28.2 | 21.4 | ND | 9.1 | Denmark | Autumn | Straw | [45] |
| 4 | 41.6 | 18.7 | 22.2 | ND | 17.5 | Germany | Winter | Straw | [46] |
| 5 | 31.2 | 21.7 | 23.4 | 8.1 | 15.6 | China | Summer | Straw | [47] |
| 6 | 37.1 | 23.5 | 15.8 | 8 | 15.6 | Estonia | Summer | Straw | [48] |
| 7 | 40.7 | 25.2 | 21.7 | 7.6 | 4.8 | China | Summer | Stem | [49] |
| 8 | 35.1 | 27.5 | 14.2 | 11.3 | 11.9 | China | Summer | Leaf | [49] |
| 9 | 43.3 | 31.7 | 17.3 | ND | 7.7 | China | Summer | Straw | [50] |
| 10 | 30.6 | 32.3 | 16.8 | 5 | 15.3 | Portugal | Summer | Straw | [51] |
| 11 | 43.1 | 19.2 | 25 | 1.9 | 10.8 | China | Summer | Straw | [52] |
| 12 | 35.2 | 22.2 | 22.1 | 7.4 | 13.1 | Spain | Summer | Straw | [53] |
| 13 | 50.8 | 15.9 | 5.6 | 12.5 | 15.2 | Mexico | Summer | Straw | [4] |
| 14 | 35.4 | 21.6 | 22.6 | 4.4 | 16 | Cary, NC | Summer | Straw | [54] |
| 15 | 38.5 | 31.6 | 20.8 | ND | 9.1 | South Korea | Summer | Straw | [55] |
| 16 | 36.4 | 33.5 | 11.2 | 4.3 | 14.6 | China | Summer | Straw | [56] |
| 17 | 32.4 | 20.9 | 17.7 | ND | 29 | Boulder, CO | Summer | Straw | [57] |
| 18 | 39.3 | 20.7 | 16.2 | ND | 23.8 | France | Winter | Straw | [58] |
| 19 | 35.9 | 23.9 | 19.3 | 4.1 | 16.8 | Denmark | Summer | Straw | [59] |
| 20 | 32.6 | 24.2 | 26.5 | 4.6 | 12.1 | Sweden | Winter | Straw | [60] |
| 21 | 37.1 | 27.6 | 20.9 | 3.7 | 10.7 | cv Ambition,Denmark | WWS | Straw | [3] |
| 22 | 35.6 | 27.6 | 19.9 | 4.7 | 12.2 | cv Hereford, Denmark | WWS | Straw | [3] |
| 23 | 36.4 | 27.2 | 19.9 | 5.2 | 11.3 | cv Skalmeje, Denmark | WWS | Straw | [3] |
| 24 | 36.6 | 28.4 | 19.3 | 4.2 | 11.5 | cv Smuggler, Denmark | WWS | Straw | [3] |
| 25 | 35.9 | 27 | 20.1 | 4.6 | 12.4 | cv Frument, Denmark | WWS | Straw | [3] |

cv: Cultivar; WWS: winter wheat straw

## *2.2. Structural interconnectedness of WS chemical moieties*

In WS, cellulose is coupled to hemicellulose by hydrogen bonds, whereas hemicellulose is attached to lignin via a covalent bond. The most abundant homo-polymer sugar is a linear chain of several hundred to ten thousand β-(1→4) linked D-glucose units. The cellulose chains group together to form microfibrils and these microfibrils together form cellulose fibers. Cellulose fibers have different morphologies based on the variation in fibril structure [6]. The next major moiety, hemicellulose, is a heteropolymer consisting mainly of xylans, xyloglucans, glucomannans and mannans. Xylose is the main sugar unit for the first two polymers, whereas mannose is the main backbone sugar for the latter two. The xylose backbone is joined by β-1, 4 glycosidic bonds, often branched with other sugars such as arabinose, glucose, galactose and fructose in β-1 or β-6 positions. Hemicellulose covers the microfibrils of cellulose, and therefore a significant amount of hemicellulose has to be digested for efficient cellulose digestibility. A third polymer, pectin, forms a dynamic structure of the wheat cell wall located between the cellulose microfilaments. It is composed of hetero-polysaccharides such as galacto uronic acid residues substituted by methoxyl esters. Pectin is soluble and can be easily digested by enzymes. Finally, the WS lignin moiety is a

multifaceted, hard, insoluble and heterogeneous mixture of polymers linked via a variety of C-C and C-O ether linkages that protects the aforementioned polysaccharides from microbial degradation. In order to obtain the desired end products for bio-refineries, the removal or 'loosening' of lignin is an important key step.

## 2.3. A combination of physical and enzymatic methods is needed for wheat straw degradation

The variation in WS chemical composition combined with the structural interconnectedness highly influences the degradation process. Therefore, selecting a suitable WS degradation method is key to efficiently degrade any new WS. Most currently used WS degradation methods involve a first (physical) pretreatment followed by enzymatic hydrolysis. Table 2 extensively lists and compares the efficiencies of different WS degradation methods, as described in the literature.

**Table 2:** Wheat straw: different pre-treatments

| No | Pre-treatment (PT) methods | PT Conditions | Enzyme hydrolysis (EH) | EH Condition | Glucose & Xylose yield | EH rate (%) | Ref |
|---|---|---|---|---|---|---|---|
| 1 | Steam explosion (SE) | 210°C and 2.5 min and (4% (v/ v) $H_2SO_4$, 120°C and 30 min) | NS50013, NS50010, endoxylanase (XlnC) | 50 °C, 24 h | 28.6 & 14.7 g/L | 67/ 82 % | [8] |
| 2 | PHP (phosphoric acid plus hydrogen peroxide) | 70.2 % $H_3PO_4$(*w*/*w*), 40 °C, 2.0 h | Celluclast 1.5L, Novozyme 188 | 50 °C, 72 h | 946.2 mg/g | 95% | [44] |
| 3 | Hydrothermal pretreatment | 180 °C, 20 min | Cellic CTec2, Cellic HTec2 | 50 °C, 96 h | 24.9 & 7.4g | 72% | [54] |
| 4 | Organosolv | 0 % (*w*/*w*) aqueous EtOH, 210 °C, 0 min | Accellerase 1500 | 50 °C, 72 h | 814.9 mg/g | 82% | [61] |
| 5 | Dilute alkaline | 2 % (*w*/*v*) NaOH, 121 °C, 30 min | NS50013, NS50010, NS50030 | 50 °C, 72 h | 775 mg/g | 78% | [62] |
| 6 | Wet explosion | 35 % (*v*/*v*) $H_2O_2$, 180–185 °C, 15 min | Celluclast 1.5L, Novozyme 188 | 50 °C, 72 h | 741.6 mg/g | 74.20% | [63] |
| 7 | Steam explosion | Soaked for 18 h at 45 °C in 0.9 % (*w*/*w*) $H_2SO_4$ or water, 180 °C, 10 min | Celluclast 1.5L, Novozyme 188 | 50 °C, 72 h | 761.6 mg/g | 76% | [64] |

PT: Pre-treatment; EH: Enzyme hydrolysis;

### *2.3.1. Pretreatment*

A major hindrance in WS degradation is due to the problem of 'recalcitrance', i.e. the natural resistance of the plant cell wall to microbial and enzymatic deconstruction. To overcome this recalcitrance, pre-treatment is highly required, in order to *i)* disrupt the cell wall structure, *ii)* remove the lignin moieties and *iii)* improve the access of the enzymes to the sugar polymers. Pre-treatment specifically breaks down the macroscopic rigidity of WS and decreases the physical barriers to mass transport. Various pre-treatment methods have been concocted, which are biological, physical (mechanical comminution, extrusion), chemical (alkali, acid, ozonolysis, organosol, ionic liquids) and physico-chemical (steam explosion, liquid hot water, ammonia fiber explosion, wet oxidation, microwave pretreatment, ultrasound, CO2 explosion) in nature. An extensive overview of different studies revealed up to five methods to be promising for WS (Table 3). The optimal pre-treatment method has to elevate the sugar yields of the subsequent enzymatic hydrolysis. Also, it has to minimize the degradation level of

sugars as well as the formation of inhibitory compounds (5-HMF and furfural). I summarized the findings in Table 2. From the collective data, and placing a focus on WS, I distill that steam explosion can be superior in many aspects (Table 3). Moreover, it is performed with no chemicals and is economic and fairly energy-efficient [7]. The in-house steam explosion method that we developed was found to clearly disrupt many of the cell walls of the WS material and enhanced the break-down of cellulose fibres. This was evident when released sugars were quantified, as the steam-treated WS yielded 6 times more sugar than untreated WS (1.2 mg/ml versus 0.2 mg/ml) **(Chapter 5).** At the same time, our in-house method still revealed reduced levels of total sugars in the liquid fraction (1.2 g/L) when compared to a previous study (5.8 g/L) [8]. This suggested that major fractions of steam exploded WS material remained in the solid fraction, which was to be subjected to enzymatic hydrolysis treatment. However, a well-known problem is the generation of inhibitory compounds during WS steam explosion that affect any subsequent enzymatic treatment. Therefore, there is an increasing demand for new enzymes to effectively degrade the (hemi)cellulose moieties as well as tolerant to inhibitory compounds.

**Table 3:** Different pre-treatments and their effect on lignocellulosic biomass (wheat straw)

| | SE | Acid | Alkaline | Oxidative | AFEX | Lime |
|---|---|---|---|---|---|---|
| **Increases accessible surface area** | H | H | H | H | H | H |
| **Cellulose decrystallization** | - | - | - | n.d. | H | n.d. |
| **Hemicellulose solubilization** | H | H | L | - | M | M |
| **Lignin removal** | M | M | M | M | H | H |
| **Generation of toxic compounds** | H | H | L | L | L | M |
| **Lignin structure alteration** | H | H | H | H | H | H |

H: High); M: Medium; L: Low; n.d: not determined; -: not detected; SE: Stem explosion; AFEX: ammonia fibre expansion

### *2.3.2. Enzymatic hydrolysis*

Following steam explosion, enzymes are required to efficiently decompose the loosened up WS moieties into their primary components. Most likely, enzymes in the classes cellulases, hemicellulases, putative accessory enzymes and even laccases/ligninases are essential. Cellulases involve endo-1, 4-β-glucanases (EG) that cleave internal bonds and exo-1,4-β- glucanases that cleave two to four units from the ends of cellulose strands. Furthermore, cellobiohydrolase cleaves the disaccharide cellobiose into two glucose moieties. As discussed above, given its intimate linkage to cellulose, enhanced attack on hemicelluloses may offer major improvements in total degradation success. Various enzymes may be needed for this. Thus, de-branching enzymes like β-xylosidase, arabinofuranosidase, feruloyl esterase, acetylxylan esterase, and a-glucuronidase may be required to degrade side chains. When the side chains are degraded, the main xylan backbone is unlocked

for xylanases to act. The unlocked xylan will be hydrolyzed to xylobiose moieties that can be attacked by β-xylosidase, thus releasing xylose monomers. It also cleaves the end of the xylo-oligosaccharide (xylan backbone) to release more xylose [9]. Finally, the lignin moiety appears to pose the main obstacle to degradation due to its aromatic nature and highly branched polymer network. The key issue here is the optimal loosening of the lignin from the other moieties, allowing enzymatic access to the latter. Several commercial enzyme cocktails (Celluclast 1.5L, CTec2, Spezyme CP, Novozymes 188) are currently available for the enzymatic hydrolysis of WS. These commercial enzymes are complex mixtures of approximately 80 to 200 proteins [10], with the specific function of each individual protein in the complex mixture being unclear [11]. In general, such enzymes are mined from microorganisms such as bacteria and fungi. For example, Celluclast 1.5L has been produced from the biodegradative fungus *Trichoderma reesei*. It contains five different endoglucanases (EG; Cel7B, Cel5A, Cel12A, Cel61A, and Cel45A, respectively), two cellobiohydrolases (CBH; CBHI (Cel7A) and CBHII (Cel6A)) and several xylanases [12]. In our work, steam-treated WS with commercial enzyme cocktails (Celluclast 1.5L) yielded relatively increased amount of glucose by ~180-200 % than treated WS alone [13]. Therefore, Celluclast 1.5L cocktail is best in recovering almost 80% of glucose (cellulose moiety) but less efficient in extracting (hemicellulose moiety) xylose (56%) [14]. Thus, more opportunities exist with respect to the exploration of efficient hemicellulases by metagenomics approach.

## 3. Metagenomics to get improved WS degrading enzymes

Metagenomics involves the direct extraction of DNA from microbiomes from environmental samples, often producing a metagenomics library. Such DNA encompasses genes that may be impossible to get at by cultivation-based methods. The metagenomic DNA thus obtained is sequenced and analyzed for genes of interest. Thus, selected genes for, for instance, biocatalysts can be singled out from the metagenomic DNA and cloned to obtain 'clean' production systems. Subsequently, the new biocatalytic enzyme activity will be tested. The overall method is defined as 'functional or function-directed metagenomics'.

### 3.1. *Functional screening of metagenomic libraries for identification of genes encoding novel enzymes*

Metagenomics applied to WS degrader consortia has great potential to uncover genes for (hemi)cellulolytic enzymes [15]. The fed-batch set up with RWS used by us was supposed to select bacterial communities that enhance the prevalence of genes that encode key target enzymes [16] **(Figure 1)**. Here, fosmid-based libraries were produced from 10-fold selected WS degrading microbial consortia. Our functional screening relied on the ready expression of the gene of interest detected by activity screening with specific substrate. Thus, activity-based screening was

successfully used as described by Ekkers et al (2012) [17] and van Elsas et al (2008) [18]. The screening method primarily focussed on the degradation of a specific substrate (mimicking bonds in WS moieties) which resulted in a colour change of a substrate, **(Figure 2)** finally allowing us to pre-select (hemi)cellulolytic clones.

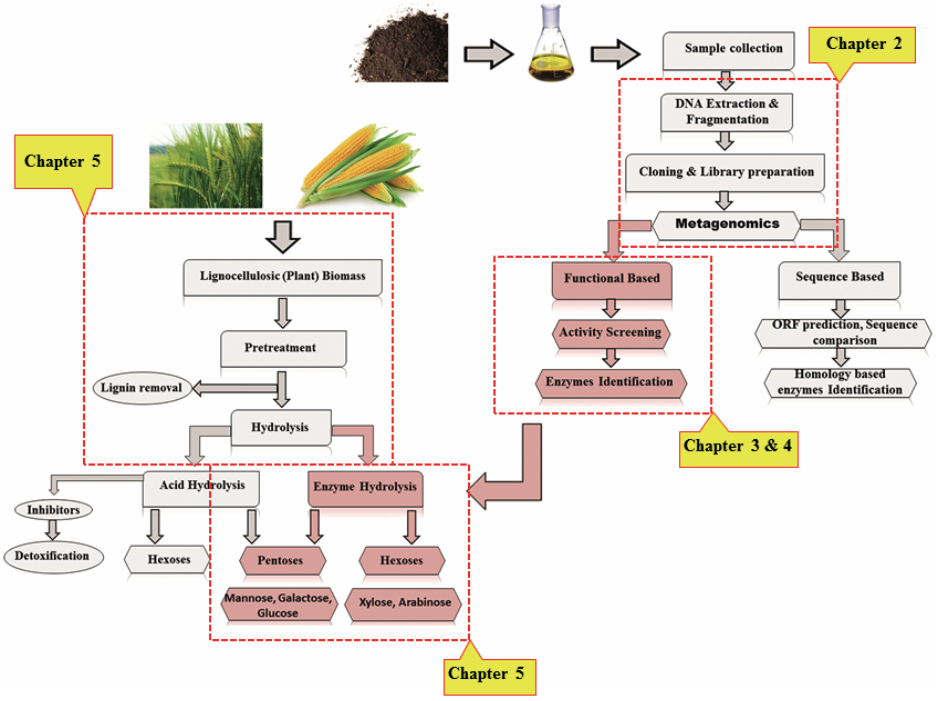



**Figure 1:** Strategies used in this thesis: Screening enzymes to impact in wheat straw degradation

***3.1.1. Substrate specificity*** - To address glycosidase activities, more than 15 distinct chromogenic and fluorimetric substrates were employed; varying levels of success were found [19]. For example, bromo-4-chloro-3-indolyl-β-D-galactopyranoside, the common substrate for β-galactosidases, produced the lowest positive hit rate of 1:700,000 [20] while azurine hydroxyethyl cellulose (for endo-celluloses) yielded high hit rates of 1 per 108 clones screened [21]. In **Chapter 2**, we give full details of the screens for genes encoding (hemi)cellulolytic enzymes in a fosmid library. Briefly, we adopted two screening strategies:

1) High-throughput screening using mixed chromogenic substrates, and,
2) Detection of enzyme activity in cell lysates (crude extracts).

The innovative approach consisted of an integration of a maximum of six substrates into the screening medium, thus enhancing hit rates as compared to the use of single substrates. This methodology yielded a total of 71 positive hits (fosmid clones), being 50 from RWS (untreated wheat straw) and 21 from TWS (heat-treated wheat straw). This corresponds to, respectively, 1 hit per 440-

screened clones (RWS), and 1 hit per 1,047 screened clones (TWS). To date, most similar screens revealed consistently low incidence rates with respect to lignocellulolytic enzymes [22]. Such hit rates are dependent on multiple contributing factors, including the enzyme activity being sought **(Chapter 3)**, the metagenomics source and host organism, the DNA extraction method and the choice of cloning vector **(Chapter 3 to 5)**. In addition, the sensitivities of the substrates and the screening conditions (such as pH, Tm, stability and addition of ions) also play vital roles in the approach **(Chapter 2)**.

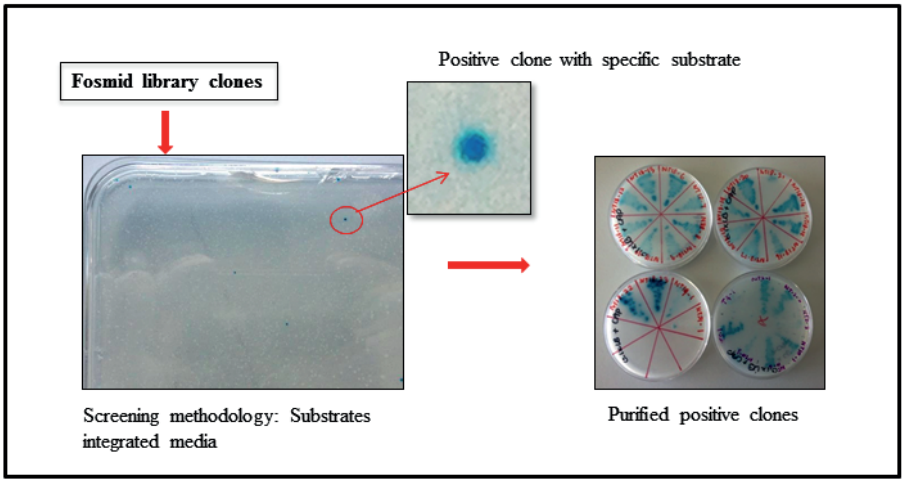



**Figure 2:** Activity screening of fosmid clones using chromogenic substrate

### *3.2. Targeted synthetic metagenomics for enzyme production*

Genes cloned directly from metagenomes may not be optimally suited for expression in the cloning host *E. coli* due to codon usage incompatibilities [23]. Thus, the codon usage of a gene of interest may have to be adjusted. However, codon swapping has possible disadvantages that are related to a fast depletion of the precise cognate tRNAs in the host organism. An exact "smart mix" of major and minor codons appears to be necessary, which may be gene- and host-dependent. However, it is often unclear to what extent a novel protocol can optimize gene expression. We applied synthetic gene technology using a preconceived optimal codon mix for the expression of gene XylM1989 in *E. coli*. In **chapter 4,** I report this novel bi-functional hemicellulolytic enzyme produced in *E. coli*. The homology-modelled structure of XylM1989 displayed an architecture that was predicted to have a catalytic triad and a substrate-binding site, which are responsible for the two activities shown for the enzyme (β-xylosidase/ α-arabinosidase). This type of constellation of active sites resulting in two different catalytic activities has been described for enzymes similar to XylM1989 [24]. The

bi-functional enzyme is thus predicted to work on the hemicellulose moiety of the WS, specifically attacking side chains [25]. It is likely that such bi-functional enzymes will find their way into innovative WS degraded strategies.

**4. Selection and biochemical characterization of enzymes**

I here described the identification of several genes for novel proteins with catalytic function. A good biocatalyst must meet particular criteria to be suitable for biotechnological industry applications. In particular, the pH characteristics, Tm, stability, kinetics, response to additives and general condition tolerance have to be known and optimal [26]. In this thesis, I mined seven selected fosmids that were predicted to encode diverse (hemi)cellulases (chapter 2). From these, eight glycosyl hydrolase family genes were selected. Four of the predicted enzymes (P1, P2, P5, and P6) and one bi-functional hemicellulase XylM1989 (Chapter 4) showed significant hemicellulolytic activities (Chapter 3), and were thus taken on for further exploratory work.

The enzymes P1 and P2 worked efficiently under thermo-alkaliphilic conditions. Such thermo-alkaliphilic tolerance (without any additives) recently improved industrial needs [27] and so these enzymes are highly recommended, for instance for pulp bleaching processes. In this respect, enzymes P1 and P2 showed high homologies with protein (1) a β-galactosidase (74%, *Enterobacter hormaechei*) and protein (2) a β-xylosidase (84%, *Enterobacter mori*); they were identified as responsible for the thermo-alkaliphilic glycosidase activities of extracts of cultures of *E. coli* carrying the respective fosmids **(chapter 2)**. P1 showed elevated enzymatic activity with increased Tm, pH and was highly stable when compared with previously known β-galactosidase (BglA); it released high amounts of p-nitrophenol (58.7 U/mg) [28]. I foresee that enzymes like P1 and P2 may play vital roles in the food industry, hydrolysing for instance lactose. P2 showed an interesting bi-functional (β-xylosidase/ α-arabinosidase) activity (pH9.0; Tm 50°C), which was found to be higher than that of other β-xylosidase enzymes (Rubgx1 and XynB5) [29,30].

### *4.1. Unexpected enzymatic activity in proteins P5 and P6, initially predicted to be a diguanylate cyclase and an aquaporin*

Proteins P5 and P6 were anticipated as a diguanylate cyclase (79% with *Enterobacter cloacae*) and an aquaporin (60%, *Hyphomonas neptunium*), respectively. Remarkably, both showed similar activities on pNP-α-D-glucopyranoside at pH 4.0 – 6.0 under the same conditions. They were also highly stable for 120 min. Maximal activity was revealed with a release of 107.63 U/mg of p-nitrophenol at pH 10.0. This enhanced activity under alkaline conditions might relate to the excess $OH^-$ ions altering the enzyme shape or active site. Thus alkaline conditions might alter the pattern of substrate binding, in consistency with the "induced-fit" theory of enzyme action. The theory states that the active site of an enzyme does not

necessarily match the shape complementary to that of substrate; rather, when the substrate binds to the enzyme, the active site of the enzyme assumes its shape.

### *4.1.1 Newly identified protein P5*

Remarkably, protein P5, predicted to have diguanylate cyclase activity, showed α-glucosidase activity. This in spite of the fact that no such activity was detected from its mother fosmid clone 10BT (chapter 2). In the presence of multiple substrates, this fosmid clone had shown some activity, presumably with *p*NP-xyl. The expression of the α-glucosidase encoding gene may have been suppressed at the (fosmid) genomic level, but might be expressed from the inducible promoter of the used vector expression system (chapter 3). Also, protein P5 was found to clearly belong to the "GGDEF" domain protein family [31]. This domain was first identified in a response regulator involved in cell differentiation of *C. crescentus* [32]. It was also observed in *Salmonella enterica* enzymes involved in cellulose biosynthesis and biofilm formation [33]. For more than 20 years, all GGDEF domain family enzymes have been classified as diguanylate cyclases and/or phosphodiesterases [34]. The former enzyme produces cyclic di-GMP (cdiG), a messenger that regulates the bacterial lifestyle transition from a motile to a sessile, biofilm-forming state [35]. Most bacterial genomes have genes that encode a range of GGDEF domain proteins [33], allowing functional diversity across them. As yet, the function of most GGDEF domain proteins has not been experimentally demonstrated [36]. Surprisingly, the GGDEF domain family protein P5 revealed 100 % α-glucosidase activity in the context of (hemi)cellulose degradation. Detailed structural studies should be performed that address the dynamics of such glycoside hydrolase activities at varying pH values in WS degradation processes. Moreover, protein P5 was found to have a motif "GDSL" by CDART prediction.

### *4.1.2 Newly identified protein P6*

Remarkably, protein P6 (presumptively identified as an aquaporin) was identified to have two different conserved motifs, denoted NPA and GDSL. The NPA motif is a key structural feature of proteins that plays a crucial role in water transport across membranes. Surprisingly, the GDSL hydrolase family domain (encompassing esterases/lipases) of protein P6 had no shared sequence homology with any of the glycosyl hydrolases. In general, aquaporins are partially-hydrophobic integral transmembrane proteins involved in a wide variety of biological processes including photosynthesis, respiration, signal transduction, molecular transport (water), and – finally – also catalysis [37]. In fact, membrane proteins often cause experimental problems, as they are unstable and difficult to express in aqueous solution. However, P6 was successfully expressed and was soluble, with good stability (95% of activity/120 min/ $50^0$C); this stands in contrast to the predicted properties of aquaporins. Therefore, our findings suggest that P6 might not be a typical trans-membrane protein but belong to a new family

of glycosyl hydrolases with α-glucosidase activity. A closer scrutiny of the available information revealed that this GDSL family subgroup was categorized as a so-called SGNH hydrolase. Thus, our protein P6 might belong to this subgroup. The SGNH family of hydrolytic enzymes involves a wide range of catalytic functions such as lipase, protease, carbohydrate esterase, (thio)esterase, aryl esterase and acyltransferase activities [38]. The newly-identified carbohydrate esterase family 3 (CE3) gene axe2 product (acetyl xylan esterase) has a GDSL motif, which removes acetyl groups from acetylated xylan [39]. Although untested, such activity might likely enable protein P6 to hydrolyze substituents on the xylan backbone, aiding in the degradation of hemicellulose [40].

### *4.2. Enzymes showed high tolerance to inhibitors (5-HMF and furfural)*

The tolerance to inhibitory compounds, such as 5-HMF and furfural, that are often generated during WS steam explosion, is not often studied [41]. According to a previous study, the amount of 5-HMF and furfural released from WS during steam explosion was 0.2 and 0.7%, respectively [8]. I here tested the activities of the selected enzymes against different concentration of 5-HMF and furfural. Interestingly, enzymes P1, P5, P6 and XylM1989 all showed 40 to 60% tolerance to 0.5% (w/v) 5-HMF and furfural. Unfortunately, enzyme P2 showed 100% inhibition. To conclude, three of the novel enzymes showed approximately 50% tolerance. Therefore, we speculate that these are superior accessory enzymes to include in enzyme cocktails for LCB degradation processes.

## 5. Application of novel enzymes to improve the LCB degradation activity of existing (commercial) enzyme preparations

In recent studies, activities of commercial cocktails have been enhanced by supplementation with accessory enzymes. For example, supplementing AFCel (*A. fumigatus* cellulase) with Spezyme CP on corn straw consistently increased the sugar yield when compared with Spezyme CP alone [42]. Celluclast 1.5L is a so-called "efficient enzyme cocktail" that is widely used for cellulose degradation. Here, we speculate that admixture to it of our novel accessory enzymes enhances WS hydrolytic activities.

### *5.1 Do the new enzymes act as a complement for commercial cocktails to break (hemi)cellulose moieties?*

We evaluated the synergistic activity of Celluclast 1.5L and our hemicellulases P1, P2, P5, and P6 on SE WS (in different combinations) by measuring the final sugar yield (**chapter 5)**. Noticeably, the addition of enzymes P1+P5 and/or P1+P6 to the Celluclast 1.5L produced significantly raised amounts of xylose (303%), arabinose (343%) and glucose (13%) when compared to Celluclast 1.5L alone. This clearly showed that the added enzymes, with Celluclast 1.5L, worked extensively on the hemicellulose moieties, increasing the yield of xylose and arabinose. We

speculate that the action of the Celluclast 1.5L/enzyme mixture on the hemicellulose moiety might be on the xylan, the main chain of β-1-4 link of D-xylose. Thus, xylose and arabinose were primarily released from the xyloglucan backbone (Chapter 1, Figure 3). Also, the effects of enzyme P1 (β-galactosidase) might have been on the side chains of xylose-glucose (of xyloglucan), thus increasing the release of xylose and glucose. Later, enzymes P5 and P6, which are α-glucosidases, might have cleaved 1→4 linked α-glucose residues to release α-glucose moieties. To conclude, in a presumably joint and spatially-explicit action, our (hemi)cellulases were – together - effective in enhancing the activity on SE WS of the commercial Celluclast 1.5L.

## 6. Outlook and recommendations for further work

This study applied metagenomics to lignocellulolytic microbial communities bred (from soil) on WS, in order to produce a suite of novel enzymes with biotechnological potential. Clearly, several new (hemi)cellulases were enriched, selected, produced and characterized from the WS degrading microbial consortia. When microbial populations undergo enrichments on wheat straw as the single driving carbon and energy source, the (spatial) conditions in the system are modulated so as to favour those organisms that function well under wet and well-mixed conditions. With the deliberate choice of this strategy, the organisms and enzymes that come to prevail may actually be less relevant in their natural soil habitat. Clearly, the approach in my thesis was tuned to biotechnological applications via increasing the prevalence of promising organisms that might serve as the best potential sources of LCB degrading enzymes. However, finding the best enzymes is still fraught with uncertainties that derive from choices in the technologies used. For instance, in the construction of metagenomic library, *Escherichia coli* is a commonly used, but not always optimal, host. . Transfer of cloned DNA from *E. coli* to alternative hosts systems offers a promising perspective in future efforts to enhance frequencies. However, this was beyond the scope of the current study. In future studies one might adopt broad-host-range plasmid systems for cloning and subsequent spread across different hosts, in order to maximize the potential of successful expression and detection of the genes being targeted.

### *6.1. Improvement of enzyme cocktails*

As discussed above, a key prerequisite for most plant biomass saccharification processes is LCB pre-treatment. Here, we opted for SE pretreatment of WS. I found that sugars released during SE entered the liquid fractions of the substrate. In future endeavors, this released sugar has to be captured and included in the preparation, by adjusting the steam explosion process. Though our newly identified P5 and P6 enzymes worked effectively on the SE WS hemicellulose moieties, we still ignore their exact action on the substrate. Hence, I posit here that

crystallization structure analysis is important to understand the catalytic mechanism involved in the reaction. Optimal characterization of the enzyme characteristics and its mode of action is always mandatory in order to sustain our understanding of efficient enzyme cocktails for complex substrates. Therefore, the required dosage of our enzyme has to be understood. It will then need to be standardized to supplement the commercial cocktails. Furthermore, the degree of enzyme loading is important in industrial processes. Less enzyme loading at a similar degradation rate practically implies that these enzymes are active and effective in WS degradation. In this thesis, we adopted only one commercial enzyme cocktail (Celluclast 1.5L) to test our enzymes. For future work, I suggest to evaluate different commercial cocktails with our enzymes (with different dosage), in order to improve the global efficiency of cocktails for (hemi)cellulose hydrolysis. The Dutch (summer) wheat used was found to have high lignin content (22.2%), whereas for maximal degradation of hemicellulose, lignin is mainly a blocking module. Though the steam explosion applied was able to alter the lignin structure, the amount of lignin loosened up or removed is still unknown. Therefore, a robust evaluation of lignin removal might fine tune the enzyme dosage that has to be applied for effective degradability. Moreover, the generation of enzyme inhibitors during steam explosion is a hindrance to subsequent enzymatic hydrolysis approaches. Therefore, measuring the concentration of inhibitors provides valuable information regarding the types of enzymes required, and their dosage. The enzyme biochemical characteristics such as stability, tolerance to inhibitors and recovery from the conversion reaction, are important parameters for industrial purposes. In addition, other 'conditions' such as medium ionic strength, pH, temperature and buffer strength, also play important roles in LCB conversion, and therefore the effects of these factors have to be characterized.

## 7. References:


1. Lee D, Owens VN, Boe A, Jeranyama P. Composition of herbaceous biomass feedstocks. Cellulose. 2007.
2. Chen HZ, Liu ZH. Multilevel composition fractionation process for high-value utilization of wheat straw cellulose. Biotechnol Biofuels. 2014; 7:137.
3. Lindedam J, Bruun S, Jørgensen H, Felby C, Magid J. Cellulosic ethanol: interactions between cultivar and enzyme loading in wheat straw processing. Biotechnol Biofuels. 2010; 3:25.
4. Rojas-Rejón OA, Sánchez A. The impact of particle size and initial solid loading on thermochemical pretreatment of wheat straw for improving sugar recovery. Bioprocess Biosyst Eng. 2014; 37:1427–1436.
5. Collins SR, Wellner N, Martinez Bordonado I, Harper AL, Miller CN, Bancroft I, et al. Variation in the chemical composition of wheat straw: the role of tissue ratio and composition. Biotechnol Biofuels. 2014; 7:121.
6. Liu R, Yu H, Huang Y. Structure and morphology of cellulose in wheat straw. Cellulose. 2005; 12:25–34.
7. Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. Bioresour Technol.

2010; 101:4851–4861.
8. Alvira P, Tomás-Pejó E, José Negro M, Ballesteros and M. Strategies of xylanase supplementation for an efficient saccharification and cofermentation process from pretreated wheat straw. Biotechnol Prog. 2011; 27:944–950.
9. De Souza WR. Microbial degradation of lignocellulosic biomass in sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization. InTech; 2013.
10. Goldbeck R, Damásio ARL, Gonçalves TA, Machado CB, Paixão DAA, Wolf LD, et al. Development of hemicellulolytic enzyme mixtures for plant biomass deconstruction on target biotechnological applications. Appl. Microbiol Biotechnol. 2014; 98:8513–8525.
11. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels. 2010; 3:22.
12. Vinzant TB, Adney WS, Decker SR, Baker JO, Kinter MT, Sherman NE, et al. Fingerprinting *Trichoderma reesei* hydrolases in a commercial cellulase preparation. Appl Biochem Biotechnol. 2001; 91–93:99–107.
13. Rosgaard L, Pedersen S, Meyer AS. Comparison of different pretreatment strategies for enzymatic hydrolysis of wheat and barley straw. Appl Biochem Biotechnol. 2007; 143:284–296.
14. Gao D, Chundawat SPS, Krishnan C, Balan V, Dale BE. Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. Bioresour Technol. 2010; 101:2770–2781.
15. Deangelis KM, D'Haeseleer P, Chivian D, Simmons B, Arkin AP, Mavromatis K, et al. Metagenomes of tropical soil-derived anaerobic switchgrass-adapted consortia with and without iron. Stand Genomic Sci. 2013; 7:382–398.
16. Jiménez DJ, Korenblum E, van Elsas JD. Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion. Appl Microbiol Biotechnol. 2014; 98:2789–2803.
17. Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD van. The great screen anomaly-a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol. 2012; 93:1005–1020.
18. Van Elsas JD, Costa R, Jansson J, Sjöling S, Bailey M, Nalin R, et al. The metagenomics of disease-suppressive soils - experiences from the metacontrol project. Trends Biotechnol. 2008; 26(11):591–601.
19. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina O V., Golyshin PN. Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. Microb Biotechnol. 2016; 9(1):22-34
20. Wang S di, Guo G shan, Li L, Cao L chuang, Tong L, Ren G hui, et al. Identification and characterization of an unusual glycosyltransferase-like enzyme with β-galactosidase activity from a soil metagenomic library. Enzyme Microb Technol. 2014; 57:26–35.
21. Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V, et al. Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. Biosci Biotechnol Biochem. 2012; 76:1075–1084.
22. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R, et al. Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. Biotechnol Biofuels. 2013; 6:78.
23. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One; 3(5):e2189
24. Maruthamuthu M, Jiménez DJ, van Elsas JD. Characterization of a furan aldehyde-tolerant *β*-xylosidase/*α*-arabinosidase obtained through a synthetic metagenomics approach. J Appl Microbiol. 2017; 123:145–158.
25. Vrzheshch P V. Steady-state kinetics of bifunctional enzymes. Taking into account kinetic

hierarchy of fast and slow catalytic cycles in a generalized model. Biochemistry. 2007; 72(9):936-943.
26. Lin Z-X, Zhang H-M, Ji X-J, Chen J-W, Huang H. Hydrolytic enzyme of cellulose for complex formulation applied research. Appl Biochem Biotechnol. 2011; 164:23–33.
27. Vester JK, Glaring MA, Stougaard P. Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb Cell Fact. 2014; 13:72.
28. Nakagawa T, Fujimoto Y, Ikehata R, Miyaji T, Tomizuka N. Purification and molecular characterization of cold-active beta-galactosidase from *Arthrobacter psychrolactophilus* strain F2. Appl Microbiol Biotechnol. 2006; 72:720–725.
29. Zhou J, Bao L, Chang L, Liu Z, You C, Lu H. Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. Lett Appl Microbiol. 2012; 54:79–87.
30. Justo PI, Corrêa JM, Maller A, Kadowaki MK, da Conceição-Silva JL, Gandra RF, et al. Analysis of the xynB5 gene encoding a multifunctional GH3-BglX β-glucosidase-β-xylosidase-α-arabinosidase member in *Caulobacter crescentus*. Antonie Van Leeuwenhoek. 2015; 108:993–1007.
31. Cotter PA, Stibitz S. c-di-GMP-mediated regulation of virulence and biofilm formation. Curr Opin Microbiol. 2007; 10:17–23.
32. Hecht GB, Newton A. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. J Bacteriol. 1995; 177:6223–6229.
33. García B, Latasa C, Solano C, García-del Portillo F, Gamazo C, Lasa I. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol Microbiol. 2004; 54:264–277.
34. Römling U, Galperin MY, Gomelsky M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev. 2013; 77:1–52.
35. Hallberg ZF, Wang XC, Wright TA, Nan B, Ad O, Yeo J, et al. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3′, 3′-cGAMP). PNAS. 2016; 113:1790–1795.
36. Galperin MY, Nikolskaya AN, Koonin E V. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 2001; 203:11–21.
37. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol. 2008;18:581–586.
38. Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F. GDSL family of serine esterases/lipases. Prog Lipid Res. 2004; 43:534–552.
39. Alalouf O, Balazs Y, Volkinshtein M, Grimpel Y, Shoham G, Shoham Y. A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from *Geobacillus stearothermophilus*. J Biol Chem. 2011; 286:41993–42001.
40. Kabel MA, Yeoman CJ, Han Y, Dodd D, Abbas CA, de Bont JAM, et al. Biochemical characterization and relative expression levels of multiple carbohydrate esterases of the xylanolytic rumen bacterium *Prevotella ruminicola* 23 grown on an ester-enriched substrate. Appl Environ Microbiol. 2011; 77:5671–5681.
41. Van der Pol EC, Bakker RR, Baets P, Eggink G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl Microbiol Biotechnol. 2014; 98(23):9579-9593.
42. Wang D, Sun J, Yu H-L, Li C-X, Bao J, Xu J-H. Maximum saccharification of cellulose complex by an enzyme cocktail supplemented with cellulase from newly isolated *Aspergillus fumigatus* ECU0811. Appl Biochem Biotechnol. 2012; 166:176–186.
43. De Lima Brossi MJ, Jiménez DJ, Cortes-Tolalpa L, van Elsas JD. Soil-derived microbial consortia enriched with different plant biomass reveal distinct players acting in lignocellulose degradation. Microb Ecol. 2016; 71:616–627.
44. Qiu J, Wang Q, Shen F, Yang G, Zhang Y, Deng S, et al. Optimizing phosphoric acid plus

hydrogen peroxide (PHP) pretreatment on wheat straw by response surface method for enzymatic saccharification. Appl Biochem Biotechnol. 2017; 181:1123–1139.
45. Djajadi DT, Hansen AR, Jensen A, Thygesen LG, Pinelo M, Meyer AS, et al. Surface properties correlate to the digestibility of hydrothermally pretreated lignocellulosic poaceae biomass feedstocks. Biotechnol Biofuels 2017; 10:49.
46. Yan Q, Wang Y, Rodiahwati W, Spiess A, Modigell M. Alkaline-assisted screw press pretreatment affecting enzymatic hydrolysis of wheat straw. Bioprocess Biosyst Eng 2017; 40:221–229.
47. Qi G, Xiong L, Wang B, Lin X, Zhang H, Li H, et al. Improvement and characterization in enzymatic hydrolysis of regenerated wheat straw dissolved by LiCl/DMAc solvent system. Appl Biochem Biotechnol 2017; 181:177–191.
48. Merali Z, Ho JD, Collins SRA, Gall G Le, Elliston A, Käsper A, et al. Characterization of cell wall components of wheat straw following hydrothermal pretreatment and fractionation. Bioresour Technol. 2013; 131:226–234.
49. Jiang B, Wang W, Gu F, Cao T, Jin Y. Comparison of the substrate enzymatic digestibility and lignin structure of wheat straw stems and leaves pretreated by green liquor. Bioresour Technol. 2016; 199:181–187.
50. Sun FF, Wang L, Hong J, Ren J, Du F, Hu J, et al. The impact of glycerol organosolv pretreatment on the chemistry and enzymatic hydrolyzability of wheat straw. Bioresour Technol. 2015; 187:354–361.
51. Silva-Fernandes T, Duarte LC, Carvalheiro F, Marques S, Loureiro-Dias MC, Fonseca C, et al. Biorefining strategy for maximal monosaccharide recovery from three different feedstocks: eucalyptus residues, wheat straw and olive tree pruning. Bioresour Technol. 2015; 183:203–212.
52. Zhang J, Liu W, Hou Q, Chen J, Xu N, Ji F. Effects of different pre-extractions combining with chemi-thermomechanical treatments on the enzymatic hydrolysis of wheat straw. Bioresour Technol. 2015; 175:75–81.
53. Toquero C, Bolado S. Effect of four pretreatments on enzymatic hydrolysis and ethanol fermentation of wheat straw. Influence of inhibitors and washing. Bioresour Technol. 2014; 157:68–76.
54. Ertas M, Han Q, Jameel H, Chang H. Enzymatic hydrolysis of autohydrolyzed wheat straw followed by refining to produce fermentable sugars. Bioresour Technol. 2014; 152:259–266.
55. Barman DN, Haque MA, Kang TH, Kim MK, Kim J, Kim H, et al. Alkali pretreatment of wheat straw (Triticum aestivum) at boiling temperature for producing a bioethanol precursor. Biosci Biotechnol Biochem. 2012; 76:2201–2207.
56. Qiu W, Chen H. Enhanced the enzymatic hydrolysis efficiency of wheat straw after combined steam explosion and laccase pretreatment. Bioresour Technol. 2012; 118:8–12.
57. Sathitsuksanoh N, Zhu Z, Zhang Y-HP. Cellulose solvent- and organic solvent-based lignocellulose fractionation enabled efficient sugar release from a variety of lignocellulosic feedstocks. Bioresour Technol 2012; 117:228–233.
58. Heiss-Blanquet S, Zheng D, Ferreira NL, Lapierre C, Baumberger S. Effect of pretreatment and enzymatic hydrolysis of wheat straw on cell wall composition, hydrophobicity and cellulase adsorption. Bioresour Technol. 2011; 102:5938–5946.
59. Kaparaju P, Serrano M, Thomsen AB, Kongjan P, Angelidaki I. Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. Bioresour Technol. 2009; 100(9):2562-2568
60. Linde M, Jakobsson E, Galbe M, Zacchi G. Steam pretreatment of dilute H2SO4-impregnated wheat straw and SSF with low yeast and enzyme loadings for bioethanol production. Biomass and Bioenergy. 2008; 32:326–332.
61. Wildschut J, Smit AT, Reith JH, Huijgen WJJ. Ethanol-based organosolv fractionation of wheat straw for the production of lignin and enzymatically digestible cellulose. Bioresour Technol. 2013; 135:58–66.

62. McIntosh S, Vancov T. Optimisation of dilute alkaline pretreatment for enzymatic saccharification of wheat straw. Biomass and Bioenergy. 2011; 35(7): 3094-3103
63. Georgieva TI, Hou X, Hilstrøm T, Ahring BK. Enzymatic hydrolysis and ethanol fermentation of high dry matter wet-exploded wheat straw at low enzyme loading. Appl Biochem Biotechnol. 2008; 148:35–44.
64. Ballesteros I, Negro MJ, Oliva JM, Cabanas A, Manzanares P, Ballesteros M. Ethanol production from steam-explosion pretreated wheat straw. Appl Biochem Biotechnol. 2006; 130:496–508.

# SUMMARY

Mukil Maruthamuthu

**Summary**

Microbiomes in different ecosystems, like soil, are diverse and highly complex. They have served, and continue to serve, as one of the largest and useful sources for multifold enzymes. The genetic information present in soil microbiomes has long remained cryptic as a result of the lack of cultivability of many microbes. Hence, establishing the methods necessary to unlock uncultivable microbes is of vital importance to foster applications of novel enzymes. Here, metagenomics-based approaches are appropriate to unlock the bio-catalytic potential of microbial genomes. Multi-species microbial consortia grown on target carbon sources offer key sources of such biocatalysts. In this study, soil-derived microbiomes that encode essential enzymes required for the degradation of wheat straw were explored. **In chapter 2,** I screened the system for (hemi)cellulolytic genes using fosmid libraries and applying functional metagenomics. Further, I developed a novel functional screening (multi-substrate) method to investigate the specific (hemi)cellulolytic clones; thus the hit rate of the screens was enhanced when compared to previous studies. This method, in combination with classical screening (using single substrate) and metagenomic sequencing, revealed the presence of 18 enzyme-encoding predicted genes. These enzymes belonged to twelve different glycosyl hydrolase (GH) families. Most of the genes were associated with similar ones from *Klebsiella*- related organisms. Three clones were identified that encoded putative thermoalkaliphilic enzymes (xylanase, galactosidase and glucosidase). **In chapter 3 and 4**, I characterized the activities of selected (hemi)cellulolytic enzymes based on parameters such as temperature, pH, stability and tolerance to nocious substances. One of the selected GHs (Protein P2), which was previously believed to have single β–xylosidase activity showed dual, i.e. β –xylosidase and a–arabinosidase activities. In addition, two more GH family enzymes (Proteins P5 and P6) showed thermotolerant a–glucosidase activities. *In silico* structure prediction revealed that these two enzymes belong to a new family of a-glucosidases. Another gene, denoted xylM1989, was also retrieved from the wheat straw-degrading microbial consortia. **In chapter 4,** I investigated the function of xylM1989 in detail, produced by targeted synthetic metagenomics. XylM1989 showed β-xylosidase and α-arabinosidase (GH43 family) activities and resembled a *Sphingobacterium sp.* protein. Furthermore, five features make XylM1989 a good candidate for industrial plant biomass saccharification: *i)* high activity at alkaline pH, *ii)* high reaction speed *($V_{max}$)*, *iii)* tolerance to lignocellulosic hydrolysate derived inhibitors (i.e. furfural and 5-hydroxymethylfurfural), *iv)* tolerance to high concentration of xylose in the presence of $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$ salts and *v)* release of sugars from complex polysaccharides such as xylan from beechwood, oat spelt xylan, and arabinoxylan. Finally, I addressed the effect of the novel enzymes as adjuvants to commercial enzyme cocktails in wheat straw degradation assays. **Chapter 5** describes the efficiency of the enzymes in these assays. Here, selecting the proper plant biomass pretreatment method, as well as

the composition of the enzyme cocktail play a determining role in the efficiency of enzymatic hydrolysis. Here, I opted for an in-house-developed steam explosion method for the pretreatment of wheat straw. The three (hemi)cellulases (P1, P5 and P6) identified in **chapter 3** efficiently enhanced the activity of a commercially available cellulase cocktail.

# SAMENVATTING

**Mukil Maruthamuthu**

## Samenvatting

Microbiomen in verschillende ecosystemen, zoals de bodem, zijn divers en zeer complex. Zij vormen één van de belangrijkste bronnen voor een scala aan enzymen. De genetische informatie die aanwezig is in bodemmicrobiomen is lang cryptisch gebleven als gevolg van het gebrek aan kweekbaarheid van veel microben. Derhalve is het nu van vitaal belang dat metagenomica-gestoelde methoden voor het ontgrendelen van de onkweekbare microben worden toegepast om de ontdekking van nieuwe enzymen te bevorderen. Multi-species microbiële consortia gekweekt op specifieke (plant-afgeleide) substraten offreren belangrijke bronnen van dergelijke biokatalysatoren. In deze studie werden de consortia die essentiële enzymen coderen nodig voor de afbraak van tarwestro, onderzocht. In hoofdstuk 2 heb ik het systeem gescreend op (hemi) cellulolytische genen in fosmid-bibliotheken, via een functionele metagenomica aanpak. Verder ontwikkelde ik een nieuwe functionele (mulitsubstraat) screeningsmethode om de specifieke (hemi) cellulolytische klonen te onderzoeken. Zo werd de 'hit rate' verbeterd in vergelijking met die van eerdere studies. Deze methode, in combinatie met klassieke screening (met behulp van enkel substraat) en metagenoom sequentiebepaling, onthulde de aanwezigheid van 18 enzymencoderende genen. Deze enzymen behoren tot twaalf verschillende glycosylhydrolase (GH) families. De meeste genen waren geassocieerd met soortgelijke genen uit *Klebsiella*-gerelateerde organismen. Drie klonen werden geïdentificeerd die vermeende thermoalkalifiele enzymen (xylanase, galactosidase en glucosidase) coderen. In hoofdstuk 3 en 4 heb ik de activiteiten van geselecteerde (hemi) cellulolytische enzymen gekenmerkt op basis van parameters als temperatuur, pH, stabiliteit en tolerantie voor toxische stoffen. Een van de geselecteerde GH's (Proteïne P2), die eerder werd geacht enkelvoudige β-xylosidase activiteit te hebben, vertoonde twee, d.w.z. β-xylosidase- en α-arabinosidase activiteiten. Daarnaast vertoonden twee andere GH familie enzymen (Proteïnen P5 en P6) thermotolerante α-glucosidase activiteiten. In de in silico structuurvoorspelling bleek dat deze twee enzymen tot een nieuwe familie van α-glucosidasen behoren. Een ander gen, xylM1989, werd ook gevonden in de microbiële consortia gekweekt vauit tarwestro. In hoofdstuk 4 heb ik de functie van xylM1989 gedetailleerd onderzocht, via gerichte synthetische metagenomica. XylM1989 vertoonde β-xylosidase- en α-arabinosidase-activiteiten (GH43-familie) en lijkt op een *Sphingobacterium* sp. eiwit. Vijf eigenschappen maken XylM1989 een goede kandidaat voor industriële plantbiomassa saccharificatie: i) hoge activiteit bij alkalische pH, ii) hoge reactiesnelheid (Vmax), iii) tolerantie voor lignocellulose hydrolysaat afgeleide remmers (dwz furfural en 5-hydroxymethylfurfural), iv) tolerantie voor een hoge concentratie xylose in aanwezigheid van $Ca^{2+}$, $Mg^{2+}$ en $Mn^{2+}$ zouten en v) het vrijmaken van suikers uit complexe polysacchariden zoals xylan uit beukenhout, haverspelt xylaan en arabinoxylaan. Tenslotte heb ik het effect van de nieuwe enzymen getest, als adjuvantia in commerciële enzymcocktails, in de afbraak van tarwestro. Hoofdstuk 5 beschrijft de efficiëntie van de enzymen in deze analyses. Hierbij spelen het kiezen van de juiste voorbehandelingsmethode, evenals de samenstelling van de enzymcocktail een bepalende

rol in de efficiëntie van de enzymatische hydrolyse. Hier heb ik gekozen voor een in-house stoomexplosiemethode voor het voorbehandelen van de tarwestro. De drie (hemi) cellulasen (P1, P5 en P6) geïdentificeerd in hoofdstuk 3 verhoogden de activiteit van een commercieel verkrijgbare cellulase cocktail efficiënt.

## *Acknowledgement*

*"I am no longer afraid of becoming lost, because the journey back always reveals something new, and that is ultimately good for the next."*
*— Billy Joel*

It is my great pleasure to thank people who supported me throughout my doctoral studies and during my stay in Netherlands.

First, I would like to thank my parents (Maruthamuthu and Kaliyammal) for motivating and encouraging me to realize my own potential in choosing my life path. Their support in every aspect of my life is the greatest gift that I have ever had in my life. I would like to thank my wife Kavitha for her endless support, encouragement, patience and unwavering love, which are undeniably the bedrock upon which my past seven years of my life has been built. Her tolerance of my occasional stressful moods is a testament in itself of her firm love and care. I thank all these three persons for leading me in an unexplored path, lifting me when I am falling, protecting me when I am exposed, loving me when I am hated, encouraging me when I am defeated, and for everything you gave me.

The sweetest and precious god's gift is Aaradhana, my daughter, who made my life's journey pleasant for now and forever. I also would like to thank Sivakami (and Mani) who was here all the way from India to Netherlands to take-care of my daughter. Apart from this, I would like to thank all my family members who always have a question for me, "When are you guys coming to India?".

A very special gratitude goes to my mentor Prof. Jan Dirk van Elsas who allowed me to explore my interests and aspirations in scientific career. Dirk, you gave me a complete freedom to think, ask questions, design experiments, execute them, and analyze the results. I am grateful for all your support and encouragement throughout my PhD. I also thank BE-BASIC foundation for the financial support they provided for my doctoral studies. My special thanks to the reading committee members Prof. dr. J.T.M. Elzenga, Prof. dr. Dick B. Janssen and Prof. dr. Hans van Veen for approving my thesis.

Prof. Lim Jae_Hwan (Andong National University, Andong, South Korea), I was very fortunate to have you in my scientific career. The best person whom I loved to discuss science and I never forgets our casual chats about epigenetics, which always fascinates me to work in this field. You had a great influence in shaping up my scientific career.

I take this chance to acknowledge my lovable colleagues Pilar, Miaozhi, Rashid, Stephanie, Sylvia, Irshad, and Pu Yang. I thank the other lab members including Maryam, Miao Wang, Larrissa, Elisa, Sasha, Maju, Adjie, Diego, Francisco, Patricia and Victor for their help and casual chat. Finally, I thank Joana, Joukje and Jolanda for their kind help in my PhD journey.

My stay in Netherlands was very pleasant and unforgettable. I would like to thank Manoj (paranymph) and Vijay for the sweet memories and the fun that we had in Groningen.

Indeed, I don't want to give an end by just saying thanks to these lovable friends. I hope our journey continues.

My stay in Netherlands wouldn't be pleasant without you friends. The first Indian family whom I met in Groningen is Vinod, Kavitha Vinod, Somu and Anupama Somu. The time that we spent together and the party that we had are the sweetest memories of Groningen. The weekends were very nice because of Mohan (apparom boss, intha week enna plan) and Lalitha Mohan. My special thanks to Vinod and Mohan (Groningen gangsters & Co) and I never forget our wonderful trip to Munich for beer festival. I would like to thank my Groningen friends Ranjith (mixture mama), Nithya Ranjith, Sunil, Smitha Sunil, Senthil, Priya Senthil , Veera, Elan and Selvambigai.

Finally, I would like to thank the University of Groningen for this wonderful opportunity with great environment to work. I really had lots of fun in Groningen and it gave nice feelings like my hometown and for sure I will miss Groningen and the weather.

## List of publications

1. **Maruthamuthu M***, Jiménez DJ*, Stevens P. van Elsas JD. (2016). A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. BMC Genomics. 17(1):86. *equal contributors

2. Jiménez DJ*, **Maruthamuthu M***, van Elsas JD. (2015). Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. Biotechnology for Biofuels. 8:199. *equal contributors

3. **Maruthamuthu M***, van Elsas JD. Cloning, expression, purification, and properties of novel thermoalkaliphilic enzyme cocktails. Biotechnology for Biofuels, 2017; 10: 142.

4. **Maruthamuthu M***, Jiménez DJ*, van Elsas JD. Characterization of a furan aldehydes-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic-metagenomic approach. J Appl Microbiol. 2017 May 10. doi: 10.1111/jam.13484. *equal contributors

5. **Maruthamuthu M***, van Elsas JD. Novel (hemi)cellulase enzyme cocktails for lignocellulose hydrolysis in biorefineries. (Submitted).