Accepted Manuscript

Title: Cell-free Pipeline for Discovery of Thermotolerant Xylanases and Endo-1,4-β-Glucanases

Authors: Dejan Gagoski, Zhenyu Shi, Lars Keld Nielsen, Claudia E. Vickers, Stephen Mahler, Robert Speight, Wayne A. Johnston, Kirill Alexandrov

Please cite this article as: Gagoski, Dejan, Shi, Zhenyu, Nielsen, Lars Keld, Vickers, Claudia E., Mahler, Stephen, Speight, Robert, Johnston, Wayne A., Alexandrov, Kirill, Cell-free Pipeline for Discovery of Thermotolerant Xylanases and Endo-1,4-β-Glucanases.Journal of Biotechnology http://dx.doi.org/10.1016/j.jbiotec.2017.07.014

Cell-free Pipeline for Discovery of Thermotolerant Xylanases and Endo-1,4-β-Glucanases

Short running title: "Cell-free pipeline for discovery of enzymes"

Dejan Gagoski[1], Zhenyu Shi[2], Lars Keld Nielsen[2], Claudia E. Vickers[2], Stephen Mahler[2], Robert Speight[3], Wayne A. Johnston[1] and Kirill Alexandrov[1*]

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, 4072, QLD, Australia

[2]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, 4072, QLD, Australia

[3]Queensland University of Technology (QUT), Brisbane, 4001, QLD, Australia

*to whom correspondence should be addressed:

 Kirill Alexandrov; Address: 306 Carmody Rd, Building 80, St Lucia, QLD, 4072

Australia; Phone: +61 7 3346 2017; e-mail: k.alexandrov@uq.edu.au

Highlights:

- Putative enzymes from gene databases are becoming accessible due to low cost of gene synthesis.

- Cell-free protein expression allows the expression of identified putative genes.

- Novel glucanase and xylanase assays are compatible with cell-free produced proteins.

- A streamlined pipeline integrating the steps described above has been established.

- Putative glucanases and xylanases were characterised using this pipeline.

Abstract:

The rapid expansion in the number of sequenced genomes and metagenomes provides an exceptional resource for mining of the enzymes with biotechnologically relevant properties. However, the majority of protein production and analysis methods are not sufficiently cost-efficient and scalable to experimentally verify the results of computational genomic mining. Here, we present a pipeline based on *Leishmania tarentolae* cell-free system that was used to characterise 30 putative thermostable endo-1,4-β-glucanases and xylanases identified in public genomic databases. In order to analyse the recombinant proteins without purification, novel high-throughput assays for glucanase and xylanase activities were developed. The assays rely on solubilisation of labelled particulate substrates performed in multiwell plates. Using this approach both acidophilic and thermophilic enzymes were identified. The developed approach enables rapid discovery of new biotechnologically useful enzymes.

2

Introduction

Lignocellulosic biomass is the most abundant renewable resource for sustainable production of food, fuels, chemicals, feed and fibre. Cellulose and xylan are the major polysaccharides of biomass, and endo-1,4-β-glucanases and xylanases are the enzymes that can degrade these polymers to sugar monomers for their use in nutrition or in industrial fermentation (Garvey *et al.*, 2013; Park *et al.*, 2015; Passos *et al.*, 2015). Natively produced by soil and ruminant microbes, these enzymes find use in paper and natural fibre processing, biofuel production, fodder processing, laundry detergents, and as additives to feed for ruminant livestock (Kuhad *et al.*, 2011; Juturu and Wu, 2012).

Naturally occurring enzymes frequently require adaptation for optimal use in manufacturing and industrial processes, where conditions differ from those found in nature. For example, a feed additive enzyme must have optimal activity at the pH and temperature of the animal digestive system, while being able to withstand the high temperatures of the feed pelleting process. Although advances in directed evolution approaches have greatly improved enzyme engineering, prospecting naturally occurring variants remains critical to effectively explore the vast solution space (Heux *et al.*, 2015).

Screening environmental samples for the desired enzymatic activity is generally cumbersome and inefficient. Xylanase or endo-1,4-β-glucanase activity of microorganisms can be identified colorimetrically by cultivating them on agar-plates with labelled substrate (Farrow and Arnold, 2011). However, only a fraction of

environmental microorganisms can be readily cultivated (Puspita *et al.*, 2012; Vester *et al.*, 2015) thus creating a strong bias in gene selection. Recent advances in metagenomic sequencing have dramatically increased the number of available gene candidates (Fernández-Arrojo *et al.*, 2010). The abundance of potential candidate genes places greater emphasis on selection of most promising candidates. It is not yet possible to identify best lead candidates from thousands of potential candidates using bioinformatic analysis. However, it is possible to reduce the set of candidates while ensuring maximum diversity (Uchiyama and Miyazaki, 2009; Fernández-Arrojo *et al.*, 2010).

With decreasing cost of commercial *de novo* gene synthesis (Kosuri and Church, 2014), synthesis of dozens to hundred of genes is easily achievable. The synthesised gene candidates are expressed and in some cases may be assayed directly in the lysate of the expression host (Mewis *et al.*, 2011). In an alternative approach, genes may be expressed *in vitro*, thus bypassing cell cultivation and disruption steps. Cell-free protein expression reduces the time needed to obtain protein product considerably and is amenable to multiplexed operation. In the current study, we adapted the *Leishmania tarentolae* cell-free system (LTE) for this purpose. We chose the LTE system because it produces recombinant proteins of quality sufficient for their direct analysis without purification (Gagoski, Polinkovsky, *et al.*, 2015).

In order to realise the full potential of high throughput screening (HTS), we developed a novel continuous semi-quantitative assay for both xylanases and endo-1,4-β-glucanases. The classical endpoint assay based on dinitrosalycilic acid (DNS) for reducing sugars (Miller, 1959) is time consuming and laborious even when streamlined methods of bacterial expression of proteins in 96-well plates and a

4

multiplexed DNS assay are implemented. The new assay is based on substrates (xylan and cellulose) labelled with a colorimetric dye that are stable at neutral and acidic conditions as well as at higher temperatures, facilitating characterization of enzyme activity at different temperatures and pH values. We demonstrate that the assays are suitable for analysis of recombinant proteins expressed in LTE.

Materials and Methods

Constructs

The amino acid sequences of the catalytic domains of 15 xylanases and endo-1,4-β-glucanases and the control xylanase (XynA) and endo-1,4-β-glucanase (EglS) were obtained from online databases (Table 1). The sequences were reverse translated and codon optimized with the Gene Designer 2.0 software (DNA 2.0) using the codon usage table for *L. tarentolae* (www.kazusa.or.jp/codon). The restriction sites for NcoI and NotI were omitted from the sequences by using synonymous codons. These two restriction sites were than introduced on the 3' and 5' end of the sequence, respectively. A stop codon was also introduced upstream of the NotI site. The optimized DNA sequences were synthesised by GeneScript and ProteinCT and cloned into the pLTE EGFP vector (Gagoski, Mureev, *et al.*, 2015) using the NcoI and NotI restriction sites. The resulting plasmids were prepared using the NucleoBond® Xtra Midi / Maxi kit (Macherey-Nagel). In addition the control xylanase (*xynA*) and endo-1,4-β-glucanase (*eglS*) genes were cloned into the pET21a vector using the NdeI and XhoI restriction sites.

Protein expression, purification and characterization

The xylanase (*xynA*) and endo-1,4-β-glucanase (*eglS*) cloned into the pET21a vector were expressed in *Escherichia coli* and purified on a 1 mL HisTrap Crude column (GE Healthcare). The activity of the purified enzymes was determined by the DNS-based colorimetric assay for reducing sugars (Miller, 1959) using birchwood-xylan (Sigma) and carboxymethyl cellulose (Sigma). The enzymes cloned in the pLTE EGFP vector, were expressed in the *Leischmania tarentolae*-based cell-free system (LTE). The LTE lysate and the feeding solution for cell-free expression were prepared and used as described previously (Johnston and Alexandrov, 2014). The final concentration of template DNA in the expression reaction was 10 nM. Test expression reactions were further supplemented with 1:200 v/v BODYPI-Lysine-tRNA (FluoroTect™ GreenLys, Promega) in order to fluorescently label the synthesis products. The reactions with the fluorescently labelled products were mixed 1:1 v/v 4x NuPAGE sample buffer (Life Technologies) and denatured at 95 °C for 5 minutes before loading them on a NuPAGE Novex 4-12% Bis-Tris protein gel (Life Technologies, Australia). Gels were imaged for fluorescence in a ChemiDoc MP Systems (Bio-Rad, Australia) gel documentation system. The fluorescent bands on the gels were quantified with ImageJ, in order to obtain the relative expression levels. The LTE-expressed eGFP and a dilution series of purified eGFP of known concentration were also loaded on a separate gel as described above. This gel was then blotted onto an Osmonics NitroBind™ Nitrocellulose Transfer Membrane (GE LifeSciences) in Tris-Glycine transfer buffer using an XCell II™ Blot Module (Life Technologies). The membranes were blocked overnight with 1 x Casein Blocking Buffer (Sigma-Aldrich) in 1 x PBS and 0.1% (v/v) Tween-20 (Sigma-Aldrich). The detection of the purified eGFP and the cell-free expressed eGFP was performed with a 1:2000 dilution

6

of GFP (D5.1) XP® Rabbit mAb (Cell Signaling) primary antibody and a 1:2500 diluition of IRDye® 800CW Goat anti-Rabbit IgG (LI-COR) secondary antibody. Both antibodies were diluted in 1 x Casein Blocking Buffer (Sigma-Aldrich). The fluorescent secondary antibody was detected by scanning the blot with the Odyssey Infrared Imager (LI-COR) at 800 nm. The detected bands on the Western blot were quantified with ImageJ and the expression levels of eGFP in the LTE cell-free reaction was calculated using the the purified eGFP standards using GraphPad Prism 6 software. This interpolated value was used to estimate the expression levels of the enzyme constructs by multiplying it with their relative expression levels recorded from the BODYPI-Lysine-tRNA co-expression.

Enzymatic assay

The enzymatic assays for the xylanases and endo-1,4-β-glucanases were performed with azurine-labeled (AZCL-) hydroxyethyl-cellulose and birchwood xylan (Megazyme). Both substrates were sieved through a 100 μm Falcon cell strainer (BD) before producing stock suspensions in water or 10 mM HEPES (pH 7.6). The substrate and corresponding enzyme (purified or in a cell-free reaction) were mixed and 10 mM HEPES (pH 7.6) was added up to 200 μL. Commercially available endo-1,4-β-glucanase (ECONASE CEP) and xylanase (ECONASE HCP) resuspended in PBS as well as purified and cell-free produced EglS and XynA were used for the assay calibration experiments. For the novel enzyme screening the cell-free expression reactions containing the enzymes were first diluted 1:1 to 1:10 (v/v) in PBS, incubated at different temperatures for 1 minute. 20 μL of the heat-treated sample was then mixed with 20-40 μL of labelled substrate and the reactions were supplemented with phosphate-citrate buffer (pH: 7.2; 6.2; 5.15 or 4.1) to a total

volume of 200 μL. The increase of optical density at 590 nm (OD_{590}) was measured with the Synergy 4 – BioTek plate reader (Millennium Science, Australia) with 10s-long plate shaking intervals between reads. The plate reader temperature was set to 40° C in order to simulate the body temperature of non-ruminant animas, and the xylanase reactions were measured for 90 minutes and the endo-1,4-β-glucanase reactions for 180 minutes. The maximum slope of the read was evaluated with the Gen5 software (Bio-Tek) and further analyses of the data were performed with GraphPad Prism 6. The assay plates were incubated over night at room temperature after the kinetic read. If any of the reactions exhibited visibly higher absorption at 590 nm that was previously not visible after the kinetic run, and this was the case for all triplicate reactions, then that reaction was marked with a red dot in the results. Assay plates, where there was increase of absorption at 590 nm in all of the wells were deemed as contaminated and were therefore repeated in order to verify the conditions at which the candidate shows low activity.

Results and Discussion

Continuous xylanase and endo-1,4-β-glucanase activity assay

Characterization of specific xylanase and endo-1,4-β-glucanase activities first became possible with the establishment of the DNS-based assay for detecting reducing sugars (Miller, 1959). This is a laborious end point assay as it requires the sample to be incubated with the DNS-reagent at 100°C. In an alternative approach, labelled xylan and HE-cellulose are used for "on plate" discovery of microorganisms that secrete xylanase and endo-1,4-β-glucanase. Here enzymatic degradation of insoluble

8

substrate leads to formation of a clear halo around the secreting colony (Meddeb-Mouelhi *et al.*, 2014; Qu *et al.*, 2015).

We tested whether this principle can be utilised to build scalable multiplexed assays for xylanase and endo-1,4-β-glucanase activities. Increase in OD_{590} from the release of azurine dye (AZCL) was observed for both labelled substrates in a multi-well plate containing the mixture of the particulate, labelled xylan or HE-cellulose and the solution of the corresponding enzyme. However, the residual granular, insoluble substrate found on the bottom of the 96-well plate and the inhomogeneous diffusion of the solubilized oligosaccharides introduced large variations into the observed signal. Agitation of the plate between individual reads not only led to more uniform diffusion of the solubilized dye species, but also concentrated of the insoluble grains on the sides of the well, thus clearing the light path and reducing the noise. This mixing enabled a reliable recording of optical density over time ($\Delta OD_{590}/\Delta t$), which is essential for continuous, multiplexed and semi-quantitative analysis of xylanase and endo-1,4-β-glucanase activities.

The dynamic range of the assay was determined by titrating substrates into the reaction mixture at a constant enzyme concentration (Figure 1A, 1B). While the heterogeneity of the substrates caused some variation between replicates, the mean values correlated well with the increase in substrate concentration. The endo-1,4-β-glucanase assay demonstrated a robust linear response in the 200 to 700 μg/mL range of AZCL-cellulose with low background, while the xylanase assay linear range was between 200 and 450 μg/mL of AZCL-xylan. The enzyme activity in both assays for these substrate concentrations was recorded within the linear dynamic range of the plate-reader (OD 0.0-2.5). To further asses the robustness of the assay we incubated the labelled substrates at 100 °C and in phosphate-citrate buffer with ranging from pH

9

7.2 to 3.0. In both cases, there was no detectable increase in optical density of the soluble fraction. This observation demonstrates that the assay is suitable for semi-quantitative characterisation of thermophilic and acidophilic enzymes.

Adaptation of the developed assay to in vitro expressed enzymes

The semi-quantitative enzymatic assay allowed us to assess the activity of the xylanases and endo-1,4-β-glucanases in high throughput mode, which shifted the rate-limiting step to enzyme production. Therefore, we investigated whether *in vitro* translated enzymes could be analysed using the developed assays. We first confirmed that the LTE cell-free system itself does not generate a strong background signal in the assay (Figure S1). This was expected as the LTE cell-free system is based on the intracellular parasite of lizards *L. tarentolae*, which is unlikely to require activity of oligosaccharide degrading enzymes. Next we compared the activity of LTE spiked with recombinant endo-1,4-β-glucanase EglS and xylanase XynA and LTE expressing the same enzymes (Figure 1C, 1D). We concluded that 3-10 μL of enzyme-expressing crude cell-free reaction mixture contains sufficient amounts of enzyme to enable an assay response comparable to the one produced by a purified enzyme. The activity of 3-6 μL cell-free expressing EglS corresponded to 100-200 mU of the purified enzyme. Similarly 1-10 μL of XynA expressing LTE corresponds to 10-200 mU of the purified enzyme. The relatively small volume of enzyme expressing lysate required for the assay and the inexpensive substrate facilitated comprehensive characterization of xylanases and endo-1,4-β-glucanases in a higher throughput manner compared to previous systems.

10

Bioinformatic identification of xylanase and endo-1,4-β-glucanase candidates from genomic databases

Having developed a pipeline suitable for high-throughput activity screening we wanted to identify from the available genomic data xylanase and endo-1,4-β-glucanase proteins that would have an increased likelihood of displaying high thermostability. Using an in-house bioinformatics protocol we collected all publically available sequences annotated as encoding xylanases (1658 genes) or endo-1,4-β-glucanases (4154 genes). We listed them together with the optimal growth temperatures of their hosts and their predicted thermostability score obtained by an in-house scoring algorithm. This score was dependent on the number of prolines as well as on predicted surface charge and hydrophobicity of the core obtained by amino acid alignment of the putative enzymes with the available crystal structures of the related enzymes. Using these selection criteria we selected fifteen putative endo-1,4-β-glucanases and twelve putative xylanases as well as three xylanases proven to exist on a protein level (Table 1). One of the putative xylanases (X10) is identified through metagenomic sequencing whereas the other putative xylanases and endo-1,4-β-glucanases are identified by sequence homology from genomic data of cultured organisms. All of the selected xylanases showed sequence similarity to one characterized fold (β-Jelly roll) of the glycoside hydrolase family 11 (Lombard *et al.*, 2014). The identified endo-1,4-β-glucanases formed a structurally more heterogeneous group. They belong to six glycoside hydrolase families (5, 6, 8, 9, 35, 44) and align to three characteristic folds: $(\beta/\alpha)_8$, $(\beta/\alpha)_n$ or $(\alpha/\alpha)_6$.

The amino acid sequences of the candidate genes were aligned and the portions of the respective sequences, flanking the catalytic domain consensus, were removed (Table

11

S1). The sequences were commercially synthesized *de novo* and cloned into the pLTE vector (Gagoski, Mureev, *et al.*, 2015). The selected enzymes were expressed in the LTE-based cell-free system in presence of BODYPI-Lysine-tRNA to facilitate visualization and quantification of expressed products (Figure 2). Expression levels differed 9-fold among the candidate endo-1,4-β-glucanases and an even greater difference (28-fold) was observed for xylanases, in spite of the more uniform size and fold. The variance may be attributed to overrepresentation of rare codons in individual sequences or secondary structures in the mRNA remaining despite sequence optimisation.

Screening for pH-dependent activity of xylanase and endo-1,4- β-glucanase

As feedstock additives are required to display high enzymatic activity in the acidic environment of the digestive tract of livestock, we performed a pH profiling of the constructed xylanase and endo-1,4-β-glucanase libraries (Figure 3). Of 15 selected endo-1,4-β-glucanases only five were active under the chosen conditions. This lack of activity may be a result of the absence of the catalytic domains' flanking sequences, but also is likely to be a consequence of the structural heterogeneity across the enzyme class that led to erroneous identification and annotation and is indirectly supported by the observation that the structurally uniform xylanase library contained only three inactive candidates. One of these xylanases originates from a thermophilic organism and has been previously characterized to be active at temperatures above 50 °C (Mäntylä *et al.*, 2007). Despite the fact that the volume of the cell-free expressed enzyme was not adjusted to the expression levels, we observed no correlation between the expression levels and the activity. This lack of correlation indicates that there are

12

significant differences in the activity levels among selected enzymes. The enzymes also show different levels of activity at different pH values.

All of the endo-1,4-β-glucanases and xylanases express as soluble proteins in the LTE-based cell-free expression system. However, for some of the investigated candidates there was visible precipitation in the activity assay reactions with low pH values. Although majority of this precipitation most likely consists of the LTE-protein (especially for the assay reactions where higher volume of cell-free protein expression reaction was added), it suggests that some of the endo-1,4-β-glucanases and xylanases exhibit no activity due to instability at low pH values or co-precipitation.

Characterization of the endo-1,4-β-glucanase library

Next, we pre-incubated five active endo-1,4-β-glucanases at different temperatures and analysed their activity at four different pH (Figure 4). The low variance in expression levels of the active candidates enabled us to adjust the volumes of the enzyme-expressing lysate to directly compare the activity of different proteins. The reaction plate was imaged after an overnight incubation at room temperature, which helped to reveal very low activities that became apparent only after long incubations. Two of the candidates exhibited significantly lower activity than the others; G02 is not thermostable and only active at neutral pH, while G15 is active enough so that the activity can be detected. However this low activity of G15 is observed at lower pH values after exposure at elevated temperatures (60 °C – 75 °C) as well as at neutral pH levels after exposure at high temperatures. This activity profile of G15 needs to be tested at a higher assay temperature than the one used for this investigation (40 °C), at which it can exhibit higher activity, as it comes from a thermophilic organism

13

(*Spirochaeta thermophila*, temperature growth range 40°C- 73°C) (Aksenova *et al.*, 1992). The G12 candidate is more active than the previous two, but neither thermostable nor acidophilic. Finally G03 and G04 both show higher activity than the EglS that was used as a standard. Although the G03 is the most active endo-1,4-β-glucanase it does not tolerate exposure to temperatures above 60 °C, whereas the G04 is stable up to 75 °C. At the lowest pH observed in the gastrointestinal tract of poultry, the gizzard pH 3.0-3.5 (Mabelebele *et al.*, 2014), no activity was detected for any of the candidate enzymes. Although the screen did not yield an endo-1,4-β-glucanase that could be used as animal feed supplement immediately, it uncovered five previously uncharacterized enzymes with favourable properties. The G15 endo-1,4-β-glucanase showed the most favourable properties and is consequently a candidate for enzyme engineering.

Characterization of the xylanase library

We repeated the procedure described above for cell-free lysates expressing active xylanases and the control lysate expressing XynA. However, due to the larger variation in expression levels the reactions were not adjusted to the relative expression levels, as this would introduce pipetting errors. Consequently, we could only directly compare the activity for each candidate at different pH values and after exposure at different temperatures (Figure 5). A number of the xylanases (X01; 05; 07; 08 and 10) remained active when exposed to temperatures of up to 70 °C. Further, the XynA as well as the X02 and X12 displayed detectable activity even after exposure to 95 °C. However, both groups are active only between pH 6.7 and 4.1. In contrast, X03, X04 and X15 were active at pH as low as 3.0. The X15 can be

classified as acidophilic xylanase due to its activity optimum between pH 5.2 and 3.0. This group of more acidophilic enzymes is however not stable at temperatures higher than 56-60 °C. Unlike X15, X11 is not active at pH 3.0, but can withstand heat treatment at 95 °C. Finally, the X14 xylanase is the only one exhibiting activity at pH 3.0 after being heat-treated at 95 °C. Remarkably during the kinetic measurement, this xylanase exhibited activity after being exposed to 40, 50, 85.4 and 95 °C but only very low or undetectable activity at 56, 60, 70 and 75 °C. The reactions at these temperatures (56-75 °C) were identified as positive after a day of incubation at room temperature. These observations suggest that the X14 refolds into the active form after denaturation at 85-95 °C, but the exposure to 56-75 °C induced a different and inactive conformation. In addition, this xylanase is more active at pH 5.2-3.0, thus making it potentially suitable for use as a feed supplement. To get a better insight into how the activity levels compare among the xylanases the maximal slope of the reactions was divided by the relative expression levels (Figure S2). Relative to the XynA the xylanases X01, 03 and 10 show very low activity, whereas the X04, 05, 07, 14 and 15 are two to fourfold more active. Finally, X02, 08, 11 and 12 are more than ten-fold more active than the XynA. Hence, all of the three most promising candidates, the thermostable X11, the acidophilic X15 and acidophilic re-folder X14 show higher activity levels than the control xylanase when the relative expression levels are taken into account. This result demonstrates the suitability of the pipeline for discovery novel xylanases from a computationally selected library.

Conclusions

In this report we describe an integrated pipeline for discovery of novel xylanases and endo-1,4-β-glucanases. The developed enzymatic assay offers a real time

measurement of enzyme activity and produces semi-quantitative data in a multiplexed format. Furthermore, the stability of the labelled substrates used in this assays makes it amenable to investigation of the thermophilic and acidophilic properties of the enzymes. Since the *Leishmania tarentolae*-based cell-free protein expression system does not produce a significant background signal with the assay, its integration with the enzymatic assay builds the foundation for all in vitro pipeline for enzyme activity screening. We used this pipeline to discover and characterize xylanases and endo-1,4-β-glucanases suitable for animal feed applications. Fifteen candidates for each enzyme were selected from the publically available sequences with an in-house algorithm used to enrich for thermostable enzymes (manuscript in preparation). Most of the candidate genes were putatively assigned as xylanases and endo-1,4-β-glucanases and have not been biochemically characterized before. Furthermore, one of the xylanase candidates was identified from an uncultured organism through metagenomic sequencing. In the pilot screen reported here we chose to synthesize and clone fifteen xylanases and endo-1,4-β-glucanases to showcase the screening pipeline. DNA synthesis eliminated the need for culturing of the source organisms and cloning the gene for each enzyme. Using the cell-free system for the expression of the enzymes and the continuous activity assay we were able to rapidly characterize both sets of enzymes. The preliminary activity screen showed that although all open reading frames were expressed, there was a clear difference in frequency of occurrence of active candidates between the two enzyme groups. This difference of 5/15 active endo-1,4-β-glucanases to 12/15 active xylanases can be attributed to the structural diversity, reflected in the number of glycoside hydrolase families. Activity characterization of the endo-1,4-β-glucanases identified one thermostable candidate, whereas three xylanases out of fifteen displayed favourable activity profiles. One

xylanase was highly active and thermostable, another acidophilic and the third most interesting xylanase (X14) was both active at low pH and capable of refolding into an active conformation after exposure to high temperatures. These features make the last enzyme a promising candidate for further development towards application as a feed supplement. With the development of the enzymatic assay in combination with genomic database mining, commercial DNA synthesis and cell-free protein expression we have established a time- and cost-efficient pipeline for discovery of novel xylanases and endo-1,4-β-glucanases suited for specific applications such as feed supplements.

Acknowledgments:

All of the authors declare that they do not have any conflict of interest in regard to this manuscript.

References:

Aksenova, H.Y., Rainey, F.A., Janssen, P.H., Zavarzin, G.A., and Morgan, H.W. (1992) Spirochaeta thermophila sp. nov., an Obligately Anaerobic, Polysaccharolytic, Extremely Thermophilic Bacterium. *Int. J. Syst. Bacteriol.* **42**: 175–177.

Farrow, M.F. and Arnold, F.H. (2011) High throughput screening of fungal endoglucanase activity in Escherichia coli. *J. Vis. Exp.*

Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A., and Ferrer, M. (2010) Metagenomic era for biocatalyst identification. *Curr. Opin. Biotechnol.* **21**: 725–33.

Gagoski, D., Mureev, S., Giles, N., Johnston, W., Dahmer-Heath, M., Škalamera, D., et al. (2015) Gateway-compatible vectors for high-throughput protein expression in pro- and eukaryotic cell-free systems. *J. Biotechnol.* **195**: 1–7.

Gagoski, D., Polinkovsky, M.E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y., and Alexandrov, K. (2015) Benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.*

Garvey, M., Klose, H., Fischer, R., Lambertz, C., and Commandeur, U. (2013) Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol.* **31**: 581–93.

Heux, S., Meynial-Salles, I., O'Donohue, M.J., and Dumon, C. (2015) White biotechnology: state of the art strategies for the development of biocatalysts for biorefining. *Biotechnol. Adv.*

Johnston, W.A. and Alexandrov, K. (2014) Production of eukaryotic cell-free lysate from Leishmania tarentolae. *Methods Mol. Biol.* **1118**: 1–15.

Juturu, V. and Wu, J.C. (2012) Microbial xylanases: engineering, production and industrial applications. *Biotechnol. Adv.* **30**: 1219–27.

Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**: 499–507.

Kuhad, R.C., Gupta, R., and Singh, A. (2011) Microbial cellulases and their industrial applications. *Enzyme Res.* **2011**: 280696.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**: D490–5.

Mabelebele, M., Alabi, O.J., Ng`ambi, J.W., Norris, D., and Ginindza, M.M. (2014) Comparison of Gastrointestinal Tracts and pH Values of Digestive Organs of Ross 308 Broiler and Indigenous Venda Chickens Fed the Same Diet. *Asian J. Anim. Vet. Adv.* **9**: 71–76.

Mäntylä, A., Paloheimo, M., Hakola, S., Lindberg, E., Leskinen, S., Kallio, J., et al. (2007) Production in Trichoderma reesei of three xylanases from Chaetomium thermophilum: a recombinant thermoxylanase for biobleaching of kraft pulp. *Appl. Microbiol. Biotechnol.* **76**: 377–86.

Meddeb-Mouelhi, F., Moisan, J.K., and Beauregard, M. (2014) A comparison of plate assay methods for detecting extracellular cellulase and xylanase activity. *Enzyme Microb. Technol.* **66**: 16–9.

Mewis, K., Taupp, M., and Hallam, S.J. (2011) A high throughput screen for biomining cellulase activity from metagenomic libraries. *J. Vis. Exp.*

Miller, G.L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **31**: 426–428.

Park, S.-H., Ong, R.G., and Sticklen, M. (2015) Strategies for the production of cell wall-deconstructing enzymes in lignocellulosic biomass and their utilization for biofuel production. *Plant Biotechnol. J.*

Passos, A.A., Andrade, C., Phillips, C.E., Coffey, M.T., and Kim, S.W. (2015) Nutrient value of spray field forages fed to pigs and the use of feed enzymes to enhance nutrient digestibility. *J. Anim. Sci.* **93**: 1721–8.

Puspita, I.D., Kamagata, Y., Tanaka, M., Asano, K., and Nakatsu, C.H. (2012) Are uncultivated bacteria really uncultivable? *Microbes Environ.* **27**: 356–66.

Qu, W., Xue, Y., and Ding, Q. (2015) Display of fungi xylanase on Escherichia coli cell surface and use of the enzyme in xylan biodegradation. *Curr. Microbiol.* **70**: 779–85.

Uchiyama, T. and Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* **20**: 616–22.

Vester, J.K., Glaring, M.A., and Stougaard, P. (2015) Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. *Extremophiles* **19**: 17–29.

Figure captions:

Figure 1: Characterisation of a semi-quantitative glucanase and xylanase assay with purified and cell-free expressed enzymes. Enzyme activity of commercial endo-1,4-β-glucanase (A) and xylanase (B) measured as change in OD590 over time plotted against the substrate concentration. The titrations were performed using total reaction concentration of 1 U/mL of endo-1,4-β-glucanase and 5 U/mL of xylanase. Titration of purified and LTE expressed enzymes to the assay mixture containing fixed amounts of substrates. (C) Purified EglS endo-1,4-β-glucanase of known activity and EglS –expressing LTE lysate were titrated into 200 μL of the reaction volume containing 500 μg/mL AZCL HE-cellulose. (D) Purified XynA xylanase of known activity as well as XynA expressing LTE lysate were titrated into 400 μg/mL AZCL-xylan. The ΔOD590/Δt were determined as the maximum slope of 20 consecutive measuring points from the kinetic read. Error bars represent the Standard Deviation.
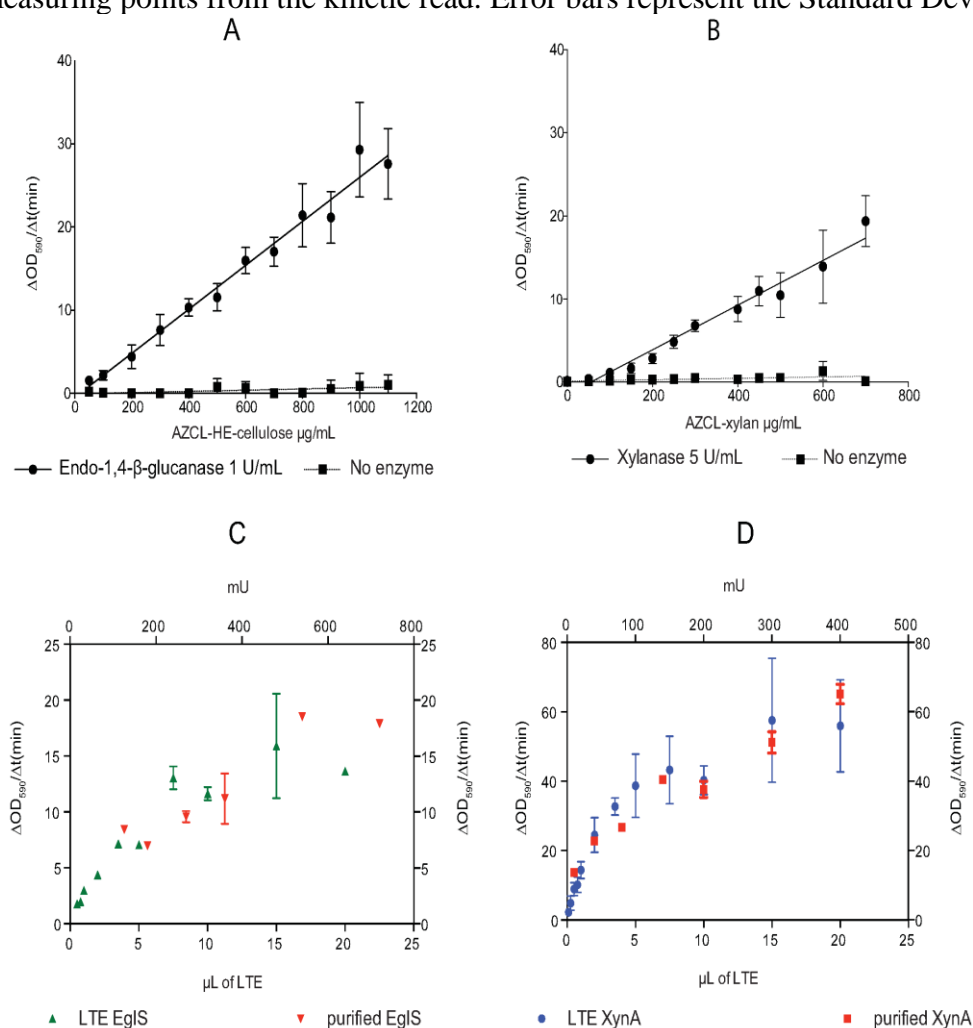
Figure 2: SDS-PAGE analysis of endo-1,4- β -glucanases and xylanases expressed with LTE-based cell-free expression system. SDS-PAGE analysis of 10 μL endo-1,4- β -glucanases and xylanases expressed in the presence of BODYPI-Lysine-tRNA in LTE cell-free system (A). (B) The expression levels for each enzyme were estimated by comparing the fluorescence band intensities to the one of 10 μL eGFP co-expression with BODYPI-Lysine-tRNA in LTE-based cell-free system, after dividing the band intensities by their respective number of lysines. The expression level of LTE-based cell-free expressed eGFP in was derived by interpolating the western blot signal of 10 μL expression reaction to the signals of purified eGFP dilution series. This estimation was performed on the total expression reaction.
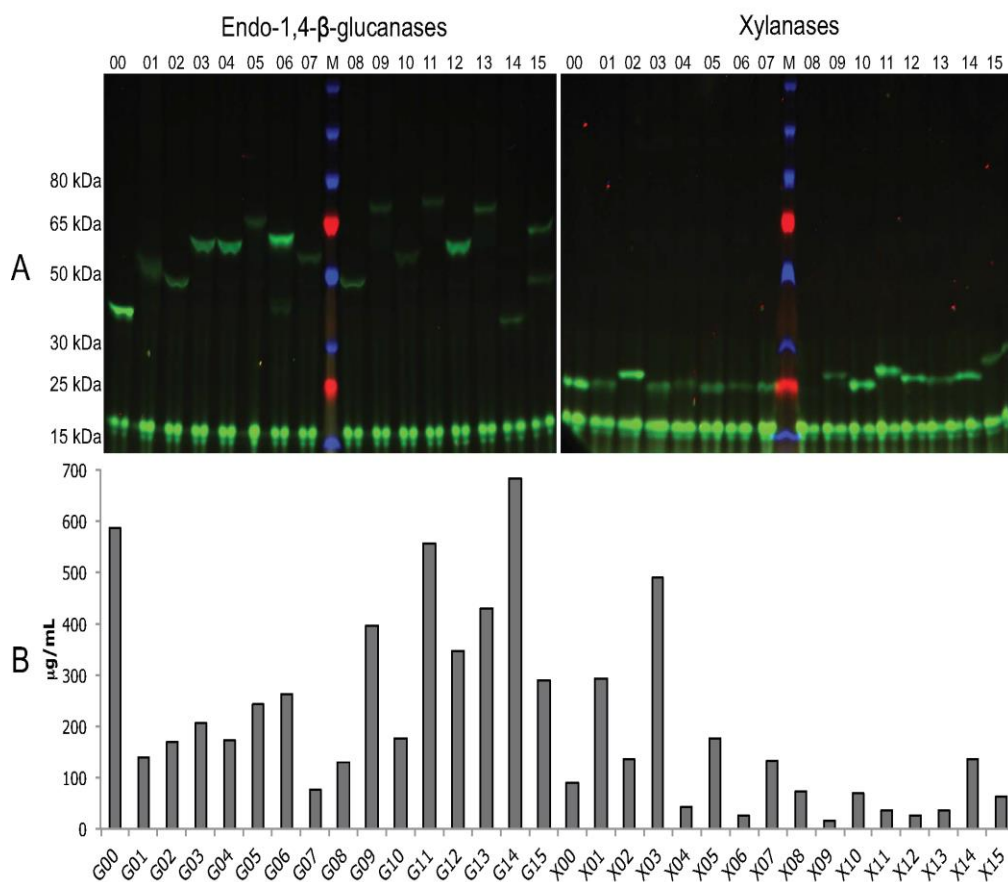
Figure 3: Activity screening of endo-1,4-β-glucanase (left) and xylanase (right) candidates at four pH values and 40 $^\circ$ C. The ΔmOD590/Δt values for all candidates were compared, where the highest slope is represented as 100% activity (red) and no change in OD590 is represented as 0% (blue).

Endo-1,4-β-glucanases	pH 6.7	pH 5.2	pH 4.1	pH 3.0
G00 (eglS)	39%	53%	23%	0%
G01	0%	0%	0%	0%
G02	2%	0%	0%	0%
G03	73%	100%	29%	0%
G04	31%	38%	1%	0%
G05	0%	0%	0%	0%
G06	0%	0%	0%	0%
G07	0%	0%	0%	0%
G08	0%	0%	0%	0%
G09	0%	0%	0%	0%
G10	0%	0%	0%	0%
G11	0%	0%	0%	0%
G12	28%	11%	14%	0%
G13	0%	0%	0%	0%
G14	0%	0%	0%	0%
G15	0%	0%	1%	0%

Xylanases	pH 6.7	pH 5.2	pH 4.1	pH 3.0
X00 (xynA)	39%	28%	4%	2%
X01	1%	4%	2%	1%
X02	98%	71%	16%	0%
X03	7%	14%	6%	1%
X04	6%	22%	4%	1%
X05	25%	14%	1%	0%
X06	0%	0%	0%	0%
X07	25%	15%	2%	0%
X08	21%	12%	2%	0%
X09	0%	0%	0%	0%
X10	17%	13%	0%	0%
X11	75%	100%	44%	0%
X12	88%	61%	6%	0%
X13	0%	0%	0%	0%
X14	1%	81%	67%	19%
X15	0%	51%	19%	3%

Figure 4: Relative activity levels of candidate endo-1,4-β-glucanase at four pH values and after exposure at eight different temperatures. The highest recorded activity is taken as the maximal activity (100%, red) for all of the investigated endo-1,4-β-glucanases. Reactions that did not produce a signal above the noise within the kinetic read (0%, blue), but resulted in increased OD590 after an overnight incubation at room temperature are designated with a red dot. These reactions need to be repeated with higher enzyme concentrations in order to be validated. The values are averages of triplicates. Relative to a commercial endo-1,4-β-glucanase the highest measured activity level (100%) is estimated to 9 ± 1 U.

G00 (eglS)

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	17%	27%	8%	0%
50°	17%	23%	3%	0%
56°	18%	28%	8%	0%
60°	21%	30%	6%	0%
70°	14%	19%	5%	0%
75°	10%	21%	0%	0%
85.4°	● 0%	● 0%	0%	0%
95°	● 0%	● 0%	0%	0%

G02

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	2%	0%	0%	0%
50°	0%	0%	0%	0%
56°	0%	0%	0%	0%
60°	0%	0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

G03

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	60%	97%	37%	0%
50°	63%	100%	35%	0%
56°	62%	91%	57%	0%
60°	64%	79%	42%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

G04

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	42%	41%	3%	0%
50°	45%	40%	3%	0%
56°	43%	45%	6%	0%
60°	45%	48%	● 0%	0%
70°	38%	48%	● 0%	0%
75°	39%	42%	● 0%	0%
85.4°	● 0%	● 0%	0%	0%
95°	0%	0%	0%	0%

G12

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	22%	6%	0%	0%
50°	19%	4%	0%	0%
56°	● 0%	● 0%	0%	0%
60°	● 0%	● 0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	● 0%	● 0%	0%	0%

G15

1 min	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	2%	2%	2%	0%
50°	1%	2%	2%	0%
56°	2%	2%	1%	0%
60°	2%	2%	3%	0%
70°	2%	2%	3%	0%
75°	3%	2%	2%	0%
85.4°	2%	2%	● 0%	0%
95°	2%	● 0%	● 0%	0%

Figure 5: Relative activity levels of xylanases at four pH values and after exposure to eight different temperatures. Due to the different expression levels of the enzymes the relative activities were compared for the different conditions for each enzyme. The highest recorded activity is taken as the maximal activity (100%, red) for the respective xylanase. Reactions that did not produce a signal above the noise within the kinetic read (0%, blue) but resulted in increased OD590 after an overnight incubation at room temperature are designated with a red dot. These reactions need to be repeated with higher enzyme concentrations in order to be validated. The compared values are averages of triplicates. Relative to a commercial xylanase the highest measured activity levels (100%) for the respective xylanases are estimated to: X00: 4.6 ± 0.5 U; X01: 0.1 ± 0.05 U; X02: 29.6 ± 6 U; X03: 0.8 ± 0.2 U; X04: 2 ± 0.3 U; X05: 3.2 ±0.2 U; X07: 2.4 ± 0.2 U; X08: 3 ± 0.3 U; X10: 1.1 ± 0.2 U; X11: 34.1 ± 2 U; X12: 27.9 ± 2 U; X14: 25.5 ± 2.5 U; X15: 5.3 ± 0.7 U.

X00 (xynA)

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	97%	54%	10%	4%
50°	91%	51%	10%	1%
56°	88%	40%	7%	2%
60°	56%	34%	4%	0%
70°	9%	1%	0%	0%
75°	8%	5%	0%	0%
85.4°	10%	2%	0%	0%
95°	12%	5%	0%	0%

X01

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	19%	80%	65%	0%
50°	0%	0%	0%	0%
56°	0%	0%	0%	0%
60°	0%	0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X02

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	83%	54%	5%	0%
50°	84%	43%	3%	0%
56°	85%	50%	4%	0%
60°	81%	44%	3%	0%
70°	3%	● 0%	0%	0%
75°	2%	0%	0%	0%
85.4°	1%	0%	0%	0%
95°	1%	0%	0%	0%

X03

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	42%	74%	40%	7%
50°	34%	73%	26%	13%
56°	6%	13%	14%	7%
60°	0%	0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X04

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	19%	90%	22%	4%
50°	5%	49%	11%	3%
56°	3%	5%	6%	0%
60°	0%	0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X05

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	96%	45%	5%	0%
50°	92%	38%	10%	0%
56°	88%	50%	7%	0%
60°	60%	30%	6%	0%
70°	4%	2%	0%	0%
75°	2%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	● 0%	0%	0%	0%

X07

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	94%	57%	4%	0%
50°	64%	35%	6%	0%
56°	7%	6%	2%	0%
60°	2%	1%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X08

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	95%	46%	3%	0%
50°	67%	36%	2%	0%
56°	78%	29%	0%	0%
60°	11%	11%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X10

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	74%	76%	0%	0%
50°	60%	62%	0%	0%
56°	10%	6%	0%	0%
60°	0%	0%	0%	0%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

X11

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	56%	99%	15%	0%
50°	61%	72%	20%	0%
56°	68%	79%	23%	0%
60°	76%	81%	26%	0%
70°	35%	57%	7%	0%
75°	32%	54%	11%	0%
85.4°	27%	83%	3%	0%
95°	51%	56%	14%	0%

X12

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	92%	64%	5%	0%
50°	77%	57%	4%	0%
56°	89%	56%	5%	0%
60°	83%	60%	3%	0%
70°	2%	1%	0%	0%
75°	3%	0%	0%	0%
85.4°	1%	0%	0%	0%
95°	3%	● 0%	0%	0%

X14

1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	1%	92%	50%	14%
50°	0%	5%	11%	2%
56°	0%	● 0%	1%	● 0%
60°	0%	● 0%	1%	● 0%
70°	0%	● 0%	1%	● 0%
75°	0%	● 0%	1%	1%
85.4°	0%	1%	2%	0%
95°	● 0%	13%	24%	1%

X15

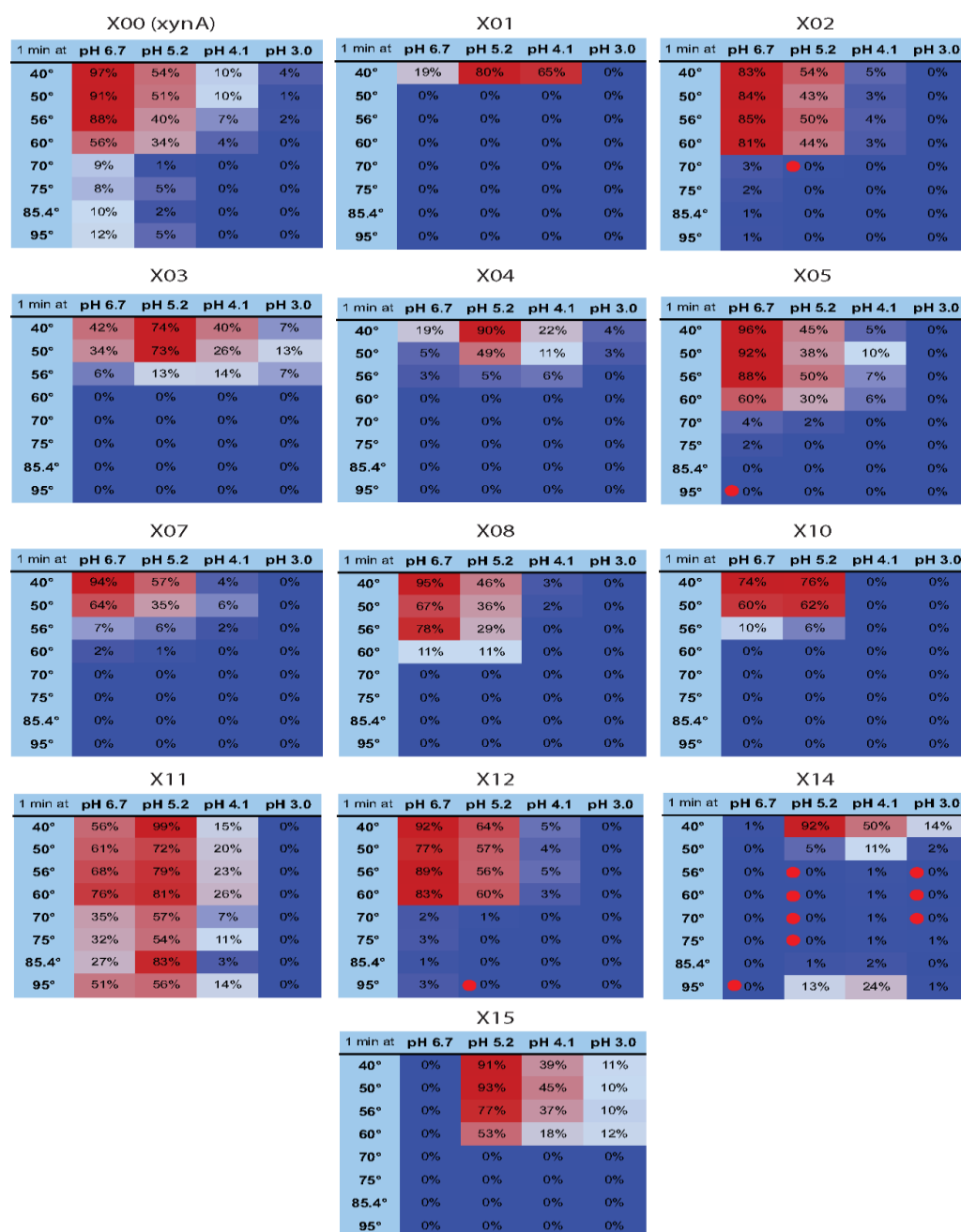| 1 min at	pH 6.7	pH 5.2	pH 4.1	pH 3.0
40°	0%	91%	39%	11%
50°	0%	93%	45%	10%
56°	0%	77%	37%	10%
60°	0%	53%	18%	12%
70°	0%	0%	0%	0%
75°	0%	0%	0%	0%
85.4°	0%	0%	0%	0%
95°	0%	0%	0%	0%

Table captions:

Table 1: List of endo-1,4-β-glucanases and xylanases used in this study.
List of endo-1,4-β-glucanases and xylanases selected from the publicly available genomic database analysis. The list shows to which glycoside hydrolase family each candidate belongs and the most probable fold. Evidence for existence on a protein level was obtained from www.uniprot.org.

<p align="center">Endo-1,4-β-glucanases</p>

Candidate	Organism	GH family	Fold	Evidence at protein level
G00 (eglS)	*Bacillus subtillis*	5	$(\beta/\alpha)_8$	yes
G01	*Medicago truncatula*	9	$(\alpha/\alpha)_6$	no
G02	*Bacillus licheniformis*	9	$(\alpha/\alpha)_6$	no
G03	*Clostridium cellulolyticum*	9	$(\alpha/\alpha)_6$	no
G04	*Melioribacter roseus*	9	$(\alpha/\alpha)_6$	no
G05	*Streptomyces enggensis*	6	$(\beta/\alpha)_n$	no
G06	*Eubacterium siraeum*	35	$(\beta/\alpha)_8$	no
G07	*Melissococcus plutonius*	5	$(\beta/\alpha)_8$	no
G08	*Coprinopsis cinerea*	5	$(\beta/\alpha)_8$	no
G09	*Streptomyces cinicus*	6	$(\beta/\alpha)_n$	no
G10	*Brevundimonas sp.*	5	$(\beta/\alpha)_8$	no
G11	*Halosimplex carlsbadense*	5	$(\beta/\alpha)_8$	no
G12	*Paenibacillus polymyxa*	44	$(\beta/\alpha)_8$	no
G13	*Streptomyces usniger*	6	$(\beta/\alpha)_n$	no
G14	*Castellaniella defragrans*	8	$(\alpha/\alpha)_6$	no
G15	*Spirochaeta thermophila*	5	$(\beta/\alpha)_8$	no

<p align="center">Xylanases</p>

Candidate	Organism	GH family	Fold	Evidence at protein level
X00 (xynA)	*Bacillus subtillis*	11	β-jelly roll	yes
X01	*Penicillium oxalicum*	11	β-jelly roll	no
X02	*Neocallimastix frontalis*	11	β-jelly roll	no
X03	*Thielavia terrestris*	11	β-jelly roll	no
X04	*Ustilago maydis*	11	β-jelly roll	yes
X05	*Humicola insolens*	11	β-jelly roll	no
X06	*Cellulophaga algicola*	11	β-jelly	no

			roll	
X07	*Setosphaeria turcica*	11	β-jelly roll	no
X08	*Fusarium verticillioides*	11	β-jelly roll	no
X09	*Ruminococcus flavefaciens*	11	β-jelly roll	no
X10	*Enviromental Bacteria*	11	β-jelly roll	no
X11	*Ruminococcus albus*	11	β-jelly roll	no
X12	*Neocallimastix frontalis*	11	β-jelly roll	no
X13	*Chaetomium thermophilum*	11	β-jelly roll	yes
X14	*Talaromyces stipitatus*	11	β-jelly roll	no
X15	*Bispora sp.*	11	β-jelly roll	yes