This work was assessed during the iGEM/PLOS Realtime Peer Review Jamboree on 23rd February 2018 and has been revised in response to the reviewers' comments. A transcript of these comments is available with the original article as a supplementary file.

Discovery and Characterization of Novel Lignocellulose-Degrading Enzymes from the Porcupine Microbiome

Authors

Mackenzie Thornbury* (1), Jacob Sicheri* (1), Caroline Guinard (1), David Mahoney (1), Francis Routledge (2), Matthew Curry (1), Mariam Elaghil (1), Nicholas Boudreau (1), Angela Tsai (3), Patrick Slaine (1), Emma Finlayson-Trick (1), Landon Getz (1), Jamie Cook (1), Dr. John Rohde (1), Dr. Craig McCormick (1)

1. Department of Microbiology and Immunology, Dalhousie University, Nova Scotia, Canada

2. Department of Neuroscience, Dalhousie University, Nova Scotia, Canada

3. Faculty of Science, Dalhousie University, Nova Scotia, Canada

*Co-Author

Author Contributions

Conceptualization: The entire team with the help and advice of mentors and supervisors

Parts Design: Jacob Sicheri, Mackenzie Thornbury, Mariam Elaghil; Supervision: Patrick Slaine, Landon Getz,

Bioinformatics: Nicholas Boudreau, Matthew Curry, and Jacob Sicheri; Supervision: Landon Getz

Cloning: Jacob Sicheri, Mackenzie Thornbury, Caroline Guinard; Supervision: Patrick Slaine, Landon Getz, Emma Finlayson-Trick

Expression Assay: Jacob Sicheri, Mackenzie Thornbury; Supervision: Jamie Cook, Landon Getz, Emma Finlayson-Trick

Activity Assay: Jacob Sicheri; Supervision: Jamie Cook, Patrick Slaine

Manuscript Writing: Mackenzie Thornbury, Jacob Sicheri, Matthew Curry, Francis Routledge, Mariam Elaghil, Caroline Guinard, David Mahoney, Emma Finlayson-Trick, Landon Getz

Manuscript Editing: The entire team with the help and advice of mentors and supervisors.

Abstract

Plant cell walls are comprised of cellulose, hemicellulose, and lignin, collectively known as lignocellulose. Microorganisms degrade these components to liberate sugars to meet metabolic demands. Using a metagenomic sequencing approach, we previously demonstrated that the microbiome of the North American porcupine (*Erethizon dorsatum*) is replete with novel lignocellulose-degrading enzymes. Here, we report the identification, synthesis and partial characterization of four genes from the porcupine microbiome encoding putative novel lignocellulose-degrading enzymes, including a β-xylanase, endoxylanase, β-glucosidase, and an α-L-arabinofuranosidase. These genes were identified via conserved catalytic domains associated with cellulose and hemicellulose degradation. We cloned the putative β-xylanase into the pET26b(+) plasmid, enabling inducible gene expression in *Escherichia coli* (*E. coli*) and periplasmic localization. We demonstrated IPTG-inducible accumulation of β-xylanase protein but failed to detect xylobiose degrading activity in a reporter assay. Alternative assays may be required to measure activity of this putative β-xylanase. In this report, we describe how a synthetic metagenomic pipeline can be used to identify novel microbial lignocellulose-degrading enzymes and take initial steps to introduce a hemicellulose-degradation pathway into *E. coli* to enable biofuel production from wood pulp feedstock.

Financial Disclosure

Funded by Springboard, Dalhousie Medical Research Foundation, Dalhousie University Office of the President, Dalhousie Faculty of Science, Dalhousie Faculty of Medicine, Genome Atlantic, and Dalhousie Department of Microbiology and Immunology.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Competing Interests

The authors have declared that no competing interests exist.

Ethics Statement
Not Applicable

Data Availability
Yes -- all data are fully available without restriction

Metagenomic raw data reads can be found at the NCBI Sequence Read Archive at Accession number SRP115632

Introduction

The gut microbiome comprises thousands of bacterial species encoding millions of genes with the potential to affect host physiology [1]. These microbes provide a genetic repository of enzymes that aid digestive processes, which could be repurposed for various bioengineering applications, including biofuel production. Lignocellulosic biofuel production can be achieved via synthetic microbial pipelines that degrade complex lignocellulose polymers into simple fermentable sugars [2,3]. Lignocellulosic biomass consists of cellulose, hemicellulose, and lignin in a 4:3:3 ratio [4]. Lignin has a complex chemical structure that impedes chemical and/or enzymatic hydrolysis of lignocellulose. Wi *et al*. recently demonstrated a new hydrogen peroxide pretreatment that improves downstream biocatalytic hydrolysis of lignocellulose by removing lignin [7]. Similar to cellulose, hemicellulose can be degraded to release monosaccharides that can be utilized for ethanol production, such as xylose [8]. Currently, hemicellulose sugars are not widely used for ethanol production, leading to losses in efficiency of biofuel production per input of lignocellulosic biomass.

As a hind-gut fermenter, the North American Porcupine, *Erethizon dorsatum*, has an enlarged cecum packed with microbes that aid digestion of lignified plants, coniferous (preferred) and deciduous cambium (inner bark), and flowers [5]. Using metagenomic sequencing, the 2016 Dalhousie iGEM team discovered that the porcupine microbiome is replete with microbial enzymes with putative lignocellulose-degrading properties, and that host diet influences gut microbial diversity and metabolic function [6]. Their study compared shot gun metagenomic and 16S sequencing of carnivore microbiomes to the microbiomes of herbivores such as the porcupine and beaver [6]. The results of this comparison revealed that herbivores have elevated levels of celluloytic genes in their microbiome versus carnivores [6]. This research is what inspired the 2017 team to continue to work with the porcupine microbiome with the long-term objective of harnessing the enzymatic potential of the porcupine microbiome to turn lignocellulosic biomass into biofuel. To do this, we created our own synthetic metagenomic pipeline to identify candidate cellulose- and hemicellulose-degrading enzymes from our existing datasets. We selected four novel candidate enzymes for synthesis and further study; One these previously uncharacterized enzymes, with 75% homology to a *Butyrivibrio sp*. β-xylanase (NCBI protein database: CDC35707.1), was cloned and transformed into *Escherichia coli* (*E. coli*) BL21(DE3) to enable IPTG-inducible protein expression and investigation of enzymatic activity.

Methods

All protocols can be found here.

Identification of Open Reading Frames

Metagenomic analysis of Illumina MiSeq data was conducted using our previously published protocols [6]. FASTQC and BowTie2 were used to inspect reads for overall quality and contaminants from sequencing. Reads were trimmed to 400 bp in length to remove low-quality terminal sequences from further analysis. MegaHIT alignment software processed reads in FASTq format and stitched reads into longer contigs by identifying overlapping coding regions [9]. Prodigal was used to identify open reading frames (ORFs) by searching sequences in six frames across both DNA strands [10]. A '-c' command modifier in Prodigal was used to ensure the program only detected ORFs with both start and stop codons present. Prodigal also searched for non-canonical start codons, as well as ribosome binding sites to identify all ORFs present in sequencing data. Non-canonical stop codons are relevant for gene searches as <10% of prokaryotic protein translation is initiated this way and these products are often overlooked in conventional searches [11,12].

In Silico Protein Function Predictions

pHmmer was used to identify putative function of protein domains [13,14]. Protein domains and possible functions were identified using the Research Collaboratory for Structural Bioinformatics Protein Data Bank [15]. e-values were calculated to compare domains identified in candidate proteins to known domains in the database [15]. Candidate proteins with the lowest e-values were queried against the Basic Local Alignment Search Tool (BLAST) database using pHmmer to identify proteins with major protein domain conservation. Selected candidate genes were codon-optimized for *E. coli* and synthesized by Integrated DNA technologies (IDT, Coralville, IA, USA) as gBlock gene fragments.

Gene Cloning

Candidate genes were PCR-amplified from IDT gBlock gene fragments with Phusion High-Fidelity DNA Polymerase according to manufacturer's instructions (New England Biolabs (NEB), Ipswich, MA, USA). PCR products were purified using the QIAquick gel extraction kit protocol (Qiagen Inc., Toronto, ON, Canada) (Table 1). The pET26b(+) expression plasmid was used as a backbone due to its pelB leader sequence that translocates fusion proteins to the periplasm, after which they can be secreted into the extracellular space [16]. Thus, by fusing our putative enzymes to pelB we increased the likelihood of secretion to the extracellular space to access substrates. This plasmid was generously donated to us by Dr. Zhenyu Cheng. Candidate genes and pET26b(+) were digested with restriction endonucleases (NEB) indicated in Table 1 for 1 hour at 37 °C. Digested DNA was subjected to agarose gel electrophoresis on a 0.8% agarose gel and purified using the QIAquick gel extraction kit according to manufacturer's instructions (Qiagen Inc.), then ligated with pET26b(+) plasmid DNA using T4 DNA ligase (NEB). Ligation products were transformed into chemically competent Stbl3 *E. coli* via standard heat-shock transformation method [17]. Specifically, 5 µL of ligation products were added to 50 µL of *E. coli* suspension in Luria-Bertani (LB) broth, and following heat-shock transformation, 250 µL of LB broth was added during the 1 hour recovery stage and the mixture was subsequently plated on LB agar + 25 µg/ml kanamycin. Plates were incubated at 37°C for 18-24 hours to allow the growth of transformants. Colonies were picked and inoculated into 5 mL of LB broth, grown overnight to saturation, and plasmid DNA was extracted via QIAprep Spin Miniprep Kit (Qiagen, Inc.). Plasmids were screened by restriction digestion, and processed for Sanger sequencing (Genewiz, South Plainfield, NJ, USA).

Table 1. Oligonucleotide primers for PCR amplification of candidate genes

Enzyme of Interest	Forward Primer	Reverse Primer	Restriction Enzymes used for insertion into pET26b(+) (5'-3')
β-glucosidase	GAATTCAAGGATCC AATGAACAAGACCCAC	AAAAAACTGCAGCA TAGCGGCCGCTGAG CTCTTAAGCTT	SacI/HindIII
β-xylanase	AAAAAAGGATCCTATG GCTAACCCGTACTTACCTG	AAAAAAGAGCTCAT TACTTGGTAAACACG AACGCC	BamHI/SacI
Endoxylanase	GAATTCTGGATCCAATGA TCAATAAATTACTGTCATC GGCATTGGCA	AAAAAACTGCAGCAT AGCGGCCGCTGAGC TCTTAAGCTT	SacI/HindIII
α-L-Arabinofuranosidase	GAATTCTGGATCCAATG GGTCTTGCTATCG	AAAAAACTGCAGCA TAGCGGCCGCTGAG CTCTTCTGAG	HindIII/XhoI

Inducible Protein Expression

pET26b(+)-β-xylanase plasmid was transformed into BL21(DE3) *E. coli* to enable inducible protein expression. Selected colonies were inoculated into 5 mL of LB broth and incubated at 37˚C in a shaker (220 RPM) overnight. The overnight culture was diluted 1:100 in 5 mL LB borth and incubated in a shaker until the OD_{600} of the culture reached 0.5-0.8. Once in log phase, 0.1 mM isopropylthio-β-galactoside (IPTG) (Thermo Fisher Scientific, Waltham, MA, USA) was added to induce protein expression. After 4 hours of shaking incubation at 37°C, protein lysates were harvested in 2x Laemmli. Empty vector pET26b(+) (EV) was used as a negative control. Proteins were separated by SDS-PAGE, fixed in methanol and acetic acid solution, and stained with Coomassie Brilliant Blue (Thermo Fisher Scientific).

Xylobiose Degradation Assay

The xylobiose degradation assay was modified from a previously described method [18]. BL21(DE3) *E. coli* transformed with β-xylanase-pET26b(+) vector were grown in LB broth supplemented with 25 µg/ml kanamycin to an OD_{600} of 0.6-0.8, and β-xylanase expression was induced using the allolactose mimetic IPTG. After 4 hours, 50 µL of culture was added to an opaque-walled 96-well plate (Thermo Fisher Scientific), followed by 50 µL of 200 µM of 4-methylumbelliferyl β-glycosides of xylobiose (CMU-X) in lysis buffer (1% Triton-X100, 50 mM Potassium Acetate at pH 7). The plate was incubated at 37°C for 18 hours, shaking at 220 RPM. After incubation, fluorescence was measured on a Tecan Infinite M200 PRO microplate reader with excitation at 365 nm, and emission at 450 nm. Each sample was normalized to LB broth alone. The unconjugated fluorophore, 4-methylumbelliferone, was used as a positive control and un-induced recombinant *E. coli* and pET26b(+) empty vector were used as negative controls.

Data

Identification and Cloning of Putative Microbial Enzymes via a Synthetic Metagenomic Pipeline

Using our metagenomic sequencing pipeline to investigate porcupine fecal samples (Figure 1A), we identified four microbial genes encoding putative cellulose- and/or hemicellulose-degrading enzymes (Table 2). These genes were identified by similarity of predicted primary amino acid sequences to conserved domains from known enzymes found in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (Figure 1B). Genes encoding predicted cellulose-/hemicellulose-degrading microbial enzymes from the porcupine microbiome were synthesized and cloned into the pET26b(+) vector to enable drug-inducible gene expression. The T7 promoter positioned upstream from the multiple cloning site in pET26b(+) enables IPTG-inducible gene expression. Successful cloning of the putative β-xylanase was confirmed by Sanger sequencing.

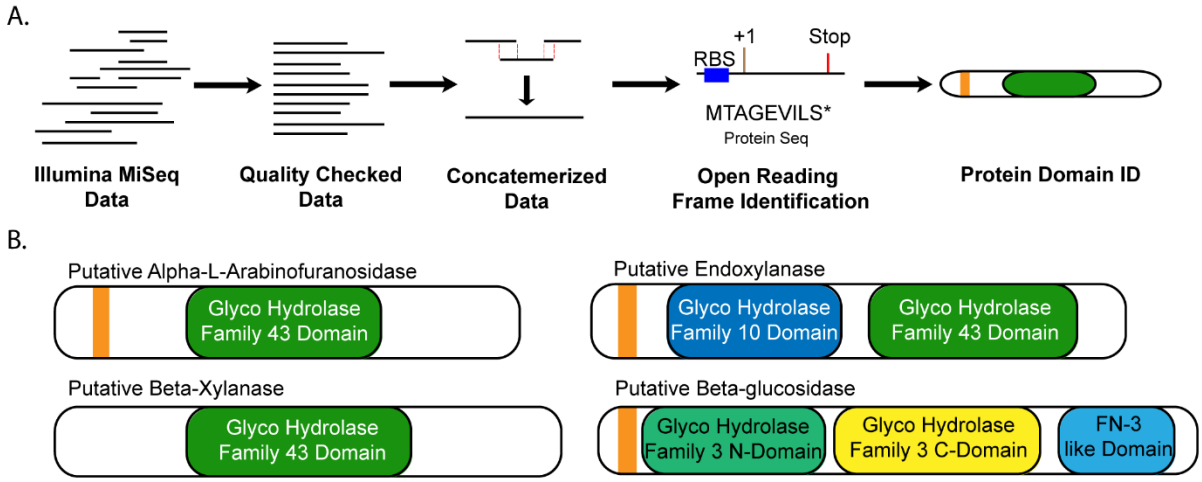

Figure 1: Identification of four genes from the porcupine microbiome with putative cellulose- and/or hemicellulose-degrading activity using a metagenomic sequencing pipeline. The bioinformatic pipeline (A) began with Illumina MiSeq data previously collected from a porcupine fecal DNA sequencing project [6]. Reads were checked for quality and trimmed, concatemerized via MegaHit, and open reading frames were identified using Prodigal. Protein sequences of interest were identified by pHMMER using various protein databases and were selected for matches of interest based on e-value selection. (B) Top candidate microbial enzymes identified by the metagenomic sequencing pipeline; putative signal sequences are shown in orange and predicted conserved protein domains are shown.

Gene (Putative Function)	Function	e-Value
Beta-Glucosidase	Cleaves beta-1,4 glycosidic bonds between two glucose monomers	2.1E-125
Beta-Xylanase	Cleaves beta-1,4 bonds between two xylose monomers	7.6E-157
Endoxylanase	Cleaves internal beta-1,4 bonds in xylose polymers	8E-94
Alpha-L-Arabinofuranosidase	Cleaves terminal non-reducing alpha-L-arabinofuranoside residues	3.5E-94

Table 2: Four candidate cellulose-/hemicellulose-degrading enzymes identified from the porcupine microbiome. The e-value is a measure of confidence, with lower values denoting higher confidence.

Inducible expression of a putative β-xylanase

To test inducible expression of the putative β-xylanase, IPTG was added to log-phase *E. coli* cultures transformed with pET26b(+)-β-xylanase plasmid or pET26b(+) vector control. After 4 hours, lysates were harvested and processed for SDS-PAGE. Proteins were visualized via Coomassie Brilliant Blue staining. IPTG treatment caused the accumulation of a distinct 51 kDa protein species in lysates from pET26b(+)-β-xylanase-transformed *E. coli,* consistent with the predicted molecular weight of the putative β-xylanase (Figure 2). This protein species was not observed in lanes containing negative control lysates (pET26b(+) empty vector, or no IPTG controls). Results were reproduced in biological replicates seen on the left and right side of the center molecular weight marker lane (M).

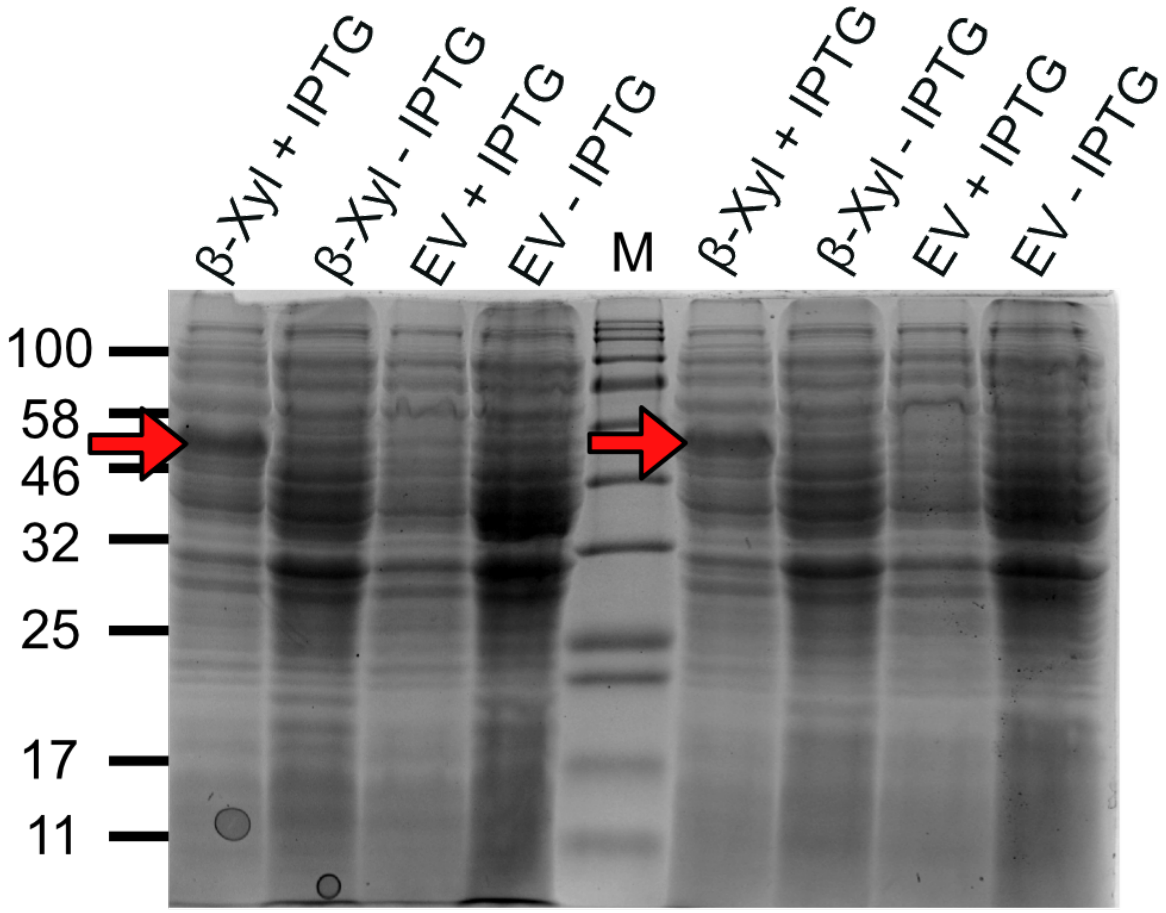

Figure 2. IPTG-inducible expression of a putative β-xylanase. Log-phase BL21(DE3) *E. coli* bearing pET26(+)-β-xylanase expression vector or empty vector control were treated with IPTG to induce T7-promoter-dependent transcription. After 4 hours, cells were lysed and processed for SDS-PAGE and stained with Coomassie Brilliant Blue to visualize proteins. Technical replicates are separated by a molecular weight marker. The red arrows indicate the ~51 kDa protein of interest, 'M' indicates molecular weight marker.

β-xylanase Enzyme Assay

β-xylanase activity was tested using substrates and a protocol developed by Hallam and Withers [18]. The assay employs CMU-X, comprised of a fluorophore conjugated to a xylose sugar via a β-1,4 glycosidic bond; β-xylanase activity cleaves this bond resulting in fluorescence emission at 450 nm. Expression of the putative β-xylanase was induced in BL21(DE3) *E. coli* as described above and cells were incubated with CMU-X in lysis buffer for 18 h, followed by measurement of fluorescence emission. Baseline fluorescence emission from control cultures was ~100 relative light units, consistent with previous observations, whereas the pure unconjugated fluorophore control was almost 100-fold higher (Figure 3). Expression of the putative β-xylanase did not increase hydrolysis of the CMU-X substrate compared to controls.

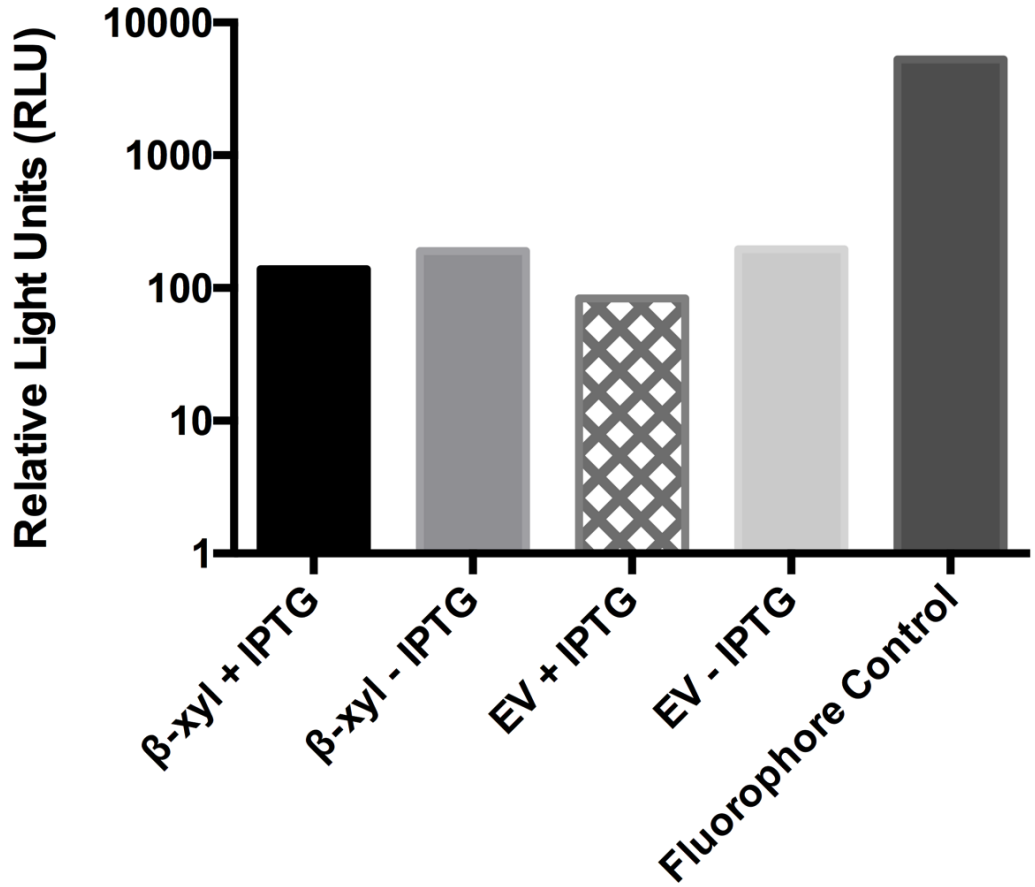

Figure 3. Candidate β-xylanase does not function in a xylobiose cleavage assay. Log-phase BL21(DE3) *E. coli* bearing pET26(+)-β-xylanase plasmid or empty vector control were treated with IPTG to induce T7 promoter-driven gene expression for 4 h, and then lysed and mixed with a solution containing CMU-X substrate for 18 h. Fluorescence emission was measured in a spectrophotometer and values expressed as Relative Light units (RLU).

Interpretation

Using a synthetic metagenomic pipeline, we discovered, synthesized and initiated characterization of four novel putative enzymes from the porcupine microbiome (Figure 1; Table 2). These candidate enzymes were selected based on predicted conservation of protein domains typically associated with cellulose- and/or

hemicellulose-degradation processes. Of these four candidate enzymes, we focused our attention an uncharacterized β-xylanase enzyme with 75% homology to a *Butyrivibrio sp.* β-xylanase (NCBI protein database: CDC35707.1) (Figure 4). Our synthetic metagenomic pipeline enabled rapid identification of novel putative proteins with desired properties. In the future, this pipeline will be useful for further mining of metagenomic sequencing datasets by others. For example, future iGEM teams may use our pipeline to discover novel enzymes with potentially useful properties for synthetic biology applications.

```
E. Ruminantium       MANPYLPNWEYIPDGEPRVFGDRIYVYGSHDTAGSDKFCDYVLKCWSAPVDDPGNWTCHG
Butyribrio Sp.       MANPYLPQWEYIPDGEPRVFGDRVYVYGSHDKAGADSFCDYVLKCWSAPVDNLNDWVCHG
iGEM2018             MANPYLPEWEYIPDGEPRVFGDRIYIYGSHDKAGSDSFCDYTLKCWSAPVSDPTNWVCHG
Ruminococcus Sp.     MSNPYLPKWEYIPDGEPRVFGDRIYVYGSHDNAGSDSFCDYVLKCWSAPLNDLSKWTCHG
RuminococcusCag:353 MSNPYLPKWEYIPDGEPRVFGDRIYVYGSHDNVGSDSFCDYVLKCWSAPLNDLSKWTCHG
                     *:*****:***************:*:*****.**:*.****.*******:.:  .*.***

E. Ruminantium       DIFRTQATRDYPADTDWTKDNNELYAPDVVEKDGKYYLYCYIINSIGCVAVSDRPCGPFK
Butyribrio Sp.       DIFHTLPDRDHAADTDWTNENNQLFAPDVVEKDGRYYLYAYIINARGCVGVSDRPEGPFK
iGEM2018             EIFRTVAENGRKADTDWTDRTNQLYAPDVVEKDGKYYLYAYIVNSIGCVAVSDKPEGPFK
Ruminococcus Sp.     EIFRTKAEKDRPANTDWTGENNQLFAPDVVEKDGKYYLYAYIINATGCVGVSDRPEGPFK
RuminococcusCag:353 EIFRTKAEKDRPANTDWTGENNQLFAPDVVEKDGKYYLYAYIINATGCVGVSDRPEGPFK
                     :**:*   .. *:****  .*:*:*********:****.**:*: ***.***:* ****

E. Ruminantium       LISKYKYTIPDEICANGWFIDPGVLVDDDGRVFIYCGYLRSFMAEVNPKNMYEILDNSCI
Butyribrio Sp.       LISQYKYTMPDEICCDGWFIDPGVFVDDDGRTYIYCGFERSFMAEVNSDNMYEIIDGSEI
iGEM2018             LLSKYKYTIPDEICCNGWFIDPGVLVDDDGRTYIYCGYERSFMAEINSDNMYEVLDGSCI
Ruminococcus Sp.     LLSKYKYTVPDEICCNGWFIDPGVLVDDDGRVYIYCGFERSFMAEVS-QNMYEISGDTVI
RuminococcusCag:353 LLSKYKYTVPDEICCNGWFIDPGVLVDDDGRVYIYCGFERSFMAEVS-QNMYEISGDTVI
                     *:*:****:*****.:********:******.:****: ******:. .****: ..: *

E. Ruminantium       EHFIPNEKTEKGGFTEEDSLFFEAASPRKVGDTYYMIYSPKRGSRLAYATSDSPTGPFTY
Butyribrio Sp.       EHIIPNEPVPERGFDTKDKLFFEACSMRKINGLYYLIYSPQQGSRLAYATSDKPTGPFTY
iGEM2018             EHIIPCEPREEGGFTERESLFFEACSLRKVNGTYYMIYSPKQGSRLAYATSDKPTGPFTY
Ruminococcus Sp.     EHIIPIVPDKEHGFAEEESLFFEACSPRKINGTYYLIYSPKRGSRLAYATSDKPTGPFTY
RuminococcusCag:353 EHIIPIVPDKEHGFAEEESLFFEACSPRKINGTYYLIYSPKQGSRLAYATSDKPTGPFTY
                     **:**     : **  .:.*****.* **:.. **:****::**********.*******

E. Ruminantium       RGYIVDSGVDYPGGNDHGSIACINGQWYVFYHRMTNNTIMSRRGCMEKIEILPDGTIPTV
```

```
Butyribrio Sp.       RGYIVDNGADYPGGNDHGSVACINGQWYIFYHRMTNGSIMSRRGCVEKIEILADGTIPTV
iGEM2018             RGYIVDNEVDYPGGNDHGSIICVNGQWYIFYHRMTNGTIMSRRGCVERIEILPDGTIPTV
Ruminococcus Sp.     RGYIVDNGVDYPGGNDHGSVACINGQWYIFYHRMTNGTIMSRRGCVERIEILPDGTIPPV
RuminococcusCag:353  RGYIVDNGVDYPGGNDHGSVACINGQWYIFYHRMTNGTIMSRRGCVERIEILPDGTIPPV
                     ******. .**********: *:*****:*******.:*******:*:**** ***** *

E. Ruminantium       ETTSLGFEEALNPYRETPAEIACVLTGGAFITEKNVFDRVIRNITDETVIGYRYFDFGDD
Butyribrio Sp.       EMTSLGFEDALSPYKVTPADIACVLKGEAFITEHNPFERVITNIKEGTVIGYKYYEFGDD
iGEM2018             EMTSLGFEESLNPYRITPADIACVLKNGGRITERNVFERYITDIFADTVIGYKYFDFGDD
Ruminococcus Sp.     EMTSLGFEETLNPYEITPADIACVLKGGATIIEKNVFERVIINITDGCEIGYKYYDFGDD
RuminococcusCag:353  EMTSLGFEETLNPYEITPADIACVLKGGATIIEKNVFERVIINITDGCEIGYKYYDFGDD
                     * ******::*.**. ***:*****.. . * *:* *:* * :*     ***:*::****

E. Ruminantium       FGSKTMEFAAKVRGMGTDCRMNIYIDGVDTADRLESVSTIGKKIGSVQIGRDDSIVSAVI
Butyribrio Sp.       YSSGTMEFAAQINGMGADCTLHILIDGED-----------GEEIGSVKIGHDSGMVNTVV
iGEM2018             FGGGDMGFSAQVGGTGCDCELHILIDGED-----------GEEIGVCSIGHDGGVVETSV
Ruminococcus Sp.     FTGKNMEFSADIIGCGCDCTVHIFADD--------------EEIGSFDVGRNGGIITCPV
RuminococcusCag:353  FTEKNMEFSADIIGCGCDCTVHIFADD--------------EEIGSFDVGRNGGIITCPV
                     :    * *:*.: * * ** ::*  *.             ::**  .:*::..::   :

E. Ruminantium       EAVTGRHAVYFTIETDYTGWTGDFFKGRSLFELESFVFMK
Butyribrio Sp.       KAVTGRHAVFFKVTTAYGGWTKTFFETKALFELKSFVFMK
iGEM2018             KCVTGRHAVYFKVTAGYSGWTGEYFKNRCLFELKAFVFTK
Ruminococcus Sp.     KCVTGRHAVYLKVSSPYSGWTGEYFKARDLFNLRSFVFMK
RuminococcusCag:353  KCVTGRHAVYLKVSSPYSGWTGEYFKARDLFNLRSFVFMK
                     :.*******::.: : * ***  :*: : **:*.:*** *
```

Figure 4. Alignment of amino acid sequences of *Eubacterium ruminantium* (*E. ruminantium)* xylan 1,4-beta-xylosidase (WP_103987394.1), *Butyrivibrio sp. CAG:318* (*Butyribio sp.*) beta-xylosidase (CDC35707.1), our putative β-xylanase protein (labeled: **iGEM2018**), uncultured *Ruminococcus sp.* (*Ruminococcus sp.)* xylosidase (SCI92110.1), and *Ruminococcus sp. CAG:353* (*RuminococcusCAG:353*) carbohydrate binding molecule (Family 6) (CDE7873.1). Alignment performed using ClustalO. Below the protein sequences is a key denoting conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations ( ).

Our team was the first to search the porcupine microbiome for enzymes that can be used in synthetic biology applications; our specific focus was on finding enzymes that could be harnessed to convert wood pulp waste into useful sugars for fermentation and biofuel production [6]. Over millions of years of evolution certain organisms, including termites and ruminids, and microbes, including fungi, have evolved multi-component enzyme systems to break down lignocellulosic biomass. Fungal species including *Aspergillus*, *Trichoderma* and *Acrophialophora* break down wood biomass using cellulases, ligninases, and auxiliary cellulase enzymes [19,20]. By contrast, termites and ruminids rely on their microbiomes to break down lignocellulosic biomass to liberate sugars [21, 22]. A study by Fibryanti *et al.* cultured four bacterial species from the gut of builder termites, two of which were able to be characterized as *Bacillus megaterium* and *Paracoccus yeei* [22]. Inspired by these examples, we chose the humble porcupine as our source for microbial cellulolytic enzymes. Like termites, fungi and ruminids, the diet of porcupines consists of cellulose-rich sources; for porcupines, these include young softwood buds and branches [23]. We hypothesized that the porcupine microbiome was a heretofore untapped source of lignocellulose-degrading enzymes that could be exploited for industrial processes. These enzymes could be produced by different microbes in the porcupine gut, with the common goal of breaking down complex lignocellulose chains, and for this reason they would have evolved to work optimally in the consistent temperature and pH of the hind-gut. We believe that careful study of the porcupine hind-gut environment will be required to guide design of bioreactors that utilize enzymes from the porcupine microbiome.

Our metagenomic sequencing pipeline identified a plethora of cellulose, hemi-cellulose, and auxiliary hemi-cellulose degradation enzymes, but we focused on four high-confidence genes for synthesis and further characterization. Significantly more work will be required to fully characterize these putative enzymes, and to start to incorporate them into pilot experiments in small-scale *E. coli* bioreactors. Beyond these four candidates, many more enzymes will be required to recreate the complete cellulose- and hemicellulose- degradation pathway in *E. coli*. Breakdown of cellulose into glucose monomers requires at least three enzymes: endoglucanase, exoglucanase, and β-glucosidase. Endoglucanase was cloned directly from *Ruminiclostridium thermocellum* cultured from porcupine fecal matter by the 2016 Dalhousie University iGEM Team [6]. Using our synthetic metagenomic pipeline, we identified and synthesized a putative β-glucosidase; more work will be required to identify an exoglucanase from our existing datasets. Hemicellulose is more complex than cellulose, necessitating a set of enzymes to hydrolyze the main polymer, and auxiliary enzymes to cleave side-chains [24]. Known hemicellulose degradation enzymes include endoxylanase and β-xylanase, which target the main polymer backbone, and α-glucuronidase and α-arabinofuranosidase that cleave side chains [24]. In this study, we identified endoxylanase, β-xylanase, and α-arabinofuranosidase enzymes, whereas α-glucuronidase remains elusive. Once we have identified, synthesized and validated a full set of cellulose- and hemicellulose-degrading enzymes, we plan to subclone these genes into operons to facilitate construction of a full cellulose-, and hemicellulose-degrading pathway in *E. coli*, wherein each enzyme will be shuttled to the periplasmic space for subsequent secretion. By using an *E. coli* with a mutant gluc1 importer, we will prevent import and utilization of glucose and xylose end-products in the bioreactor. Instead, these products will be restricted to the extracellular space, which will allow transfer to a second bioreactor where *S. cerevisiae* can ferment the sugars into bioethanol. The whole system will be assessed for efficiency and yield for comparison against current bioreactor systems.

Future Studies

Functional Metagenomics as an Alternative Discovery Method

Our synthetic metagenomic pipeline is a powerful tool for discovery, but because it infers functional relationships from homology to previously characterized sequences in a database, it will likely fail to identify greatly divergent proteins with desirable properties. It may also identify candidate proteins

that appear to have conserved functional domains, but lack the function predicted by homology. By contrast, functional metagenomic library screens rely on functional assays for gene discovery. Thus, functional metagenomics provides a convenient approach to new gene discovery that nicely complements sequence-based approaches, but with greater potential for discovery of truly novel genes that don't resemble those in existing databases. Recently, using functional metagenomics, Cheng, *et. al*. discovered three novel β-galactosidase enzymes, two of which had conserved domains, and one of which was part of a previously undiscovered enzyme family [25]. While less high-throughput, we believe creating a functional metagenomic library from porcupine microbiome DNA would be a viable way to continue to search for enzymes in the cellulose and hemicellulose degradation pathways.

Alternative Enzyme Activity Assays

Going forward, several other assays can be performed to measure activity of the putative enzymes described in this study. β-xylanase and β-glucosidase activity can be assessed by the Somogyi-Nelson method, which is based on measuring the reducing power of sugars in solution; this method relies on cupric ion reduction by a reducing sugar such as glucose and xylose [26]. β-xylanase and β-glucosidase work by cleaving the β-1,4 backbone of xylobiose and cellobiose, respectively. This cleavage results in the release of monomeric xylose or glucose, both of which can be reduced by the Somogyi-Nelson method. By contrast, endoglucanase activity can be measured by the agar-based Congo Red staining assay. As pH decreases, the colour of Congo Red darkens, which enables measurement of the reaction between carboxymethylcellulose and endoglucanase which yields a cellobiose molecule as well as a phosphate group [27]. The phosphate group lowers local pH in the semisolid medium allowing visualization of enzyme activity; this is a qualitative measure [27]. By contrast, high performance liquid chromatography (HPLC) using a column designed for carbohydrate analysis could facilitate measurements of cellulose/hemi-cellulose degradation products [28].

In our pilot experiments described in this report, all bacterial growth was carried out at 37℃, at a neutral pH, in aerobic conditions. While these conditions are optimal for growth of most laboratory *E. coli* strains, they do not match the anaerobic and acidic environment of the porcupine hind-gut [29,30]. Indeed, it is possible that our putative β-xylanase may have been inhibited by aerobic assay conditions; previous studies have shown that reactive oxygen species can alter enzyme folding and function as described previously for nitrogenase enzymes from anaerobic bacteria *Azotobacter* and *Cyanothece* [31,32]. pH also has been shown to affect protein folding, stability and function [33,34]. Future studies will thoroughly test different environmental conditions, particularly those that match the environment of the porcupine hind-gut.

References

1. Dunn KA, Moore-Connors J, MacIntyre B, Stadnyk A, Thomas NA, Noble A, et al. The gut microbiome of pediatric Crohn's disease patients differs from healthy controls in genes that can influence the balance between a healthy and dysregulated immune response. Inflamm Bowel Dis. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27760077. pmid:27760077
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 May 1;135:513-22.
3. Gruninger RJ, McAllister TA, Forester RJ. Bacterial and archaeal diversity in the gastrointestinal tract of the North American beaver (Castor canadensis). PLoS One. 2016;11(5): e0156457.
4. Hongzhang C. Biotechnology of Lignocellulose: Theory and Practice. [Internet]. 1st ed. Dordrecht: Springer Netherlands; c2014. Available from: http://www.springer.com/gp/book/9789400768970. DOI: 10.1007/978-94-007-6898-7
5. Graham D. Porcupine. South Dakota Department of Game, Fish and Parks, Division of Wildlife. Available from: http://www3.northern.edu/natsource/MAMMALS/Porcup1.htm.
6. Finlayson-Trick ECL, Getz LJ, Slaine PD, Thornbury M, Lamoureux E, Cook J, Langille MG, Murray LE, McCormick C, Rohde JR, Cheng Z. Taxonomic differences of gut microbiomes drive cellulolytic enzymatic potential within hind-gut fermenting mammals. PloS one. 2017 Dec 27;12(12):e0189404.
7. Wi SG, Cho EJ, Lee DS, Lee SJ, Lee YJ, Bae HJ. Lignocellulose conversion for biofuel: a new pretreatment greatly improves downstream biocatalytic hydrolysis of various lignocellulosic materials. Biotechnol Biofuels. 2015 Dec;8(1):228.
8. Dodd D, Cann IK. Enzymatic deconstruction of xylan for biofuel production. GCB Bioenergy. 2009 Feb 1;1(1):2-17.
9. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. May 2015;31(10):1674-6.
10. Hyatt D, Chen WL, LoCascio PF, Miriam L, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11:119.
11. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999 Jul 8; 234(2): 187-208
12. Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, Endy D et al. Measurements of translation initiation from all 64 codons in E. coli. Nucleic Acids Res. 2017 Apr 20;45(7):3615-3626
13. Johnson SL, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010 Aug;11:431.
14. Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. J Nucleic Acids. 2011 Jul;39:W29–W37
15. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank Nucleic Acids Research. 2000, 28: 235-242. www.rcsb.org
16. Sockolosky JT, Szoka FC. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. Protein Expr Purif. 2013 Feb;87(2):129-135
17. Froger A, Hall JE. Transformation of plasmid DNA into E. coli using the heat shock method. J Vis Exp. 2007 Aug 1;(6):253. doi:10.3791/253.

18. Chen HM, Armstrong Z, Hallam SJ, Withers SG. Synthesis and evaluation of a series of 6-chloro-4-methylumbelliferyl glycosides as fluorogenic reagents for screening metagenomic libraries for glycosidase activity. Carbohydr Res. 2016 Feb 8;421:33-39
19. Srebotnik E, Messner K. A simple method that uses differential staining and light microscopy to assess the selectivity of wood delignification by white rot fungi. Appl Environ Microbial. 1994 Apr; 60(4): 1383-1386.
20. Barros RRO, Oliveira RA, Gottschalk LMF, Bon EPS. Production of cellulolytic enzymes by fungi Acrophialophora nainiana and Ceratocystis paradoxa using different carbon sources. Appl Biochem Biotech. 2010 Feb 20; 161:448.
21. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. 2011 Jan 28; 331(6016):463-467
22. Ferbiyanto A, Rusmana I, Raffiudin R. Characterization and identification of cellulolytic bacteria from gut of worker Mactotermes gilvus. HAYATI J Biosci. 2015 Oct; 22(4):197-200.
23. North American Porcupine. [cited 17 March 2018]. In: NatureWorks Website [Internet]. Available from: http://www.nhptv.org/natureworks/porcupine.htm
24. U.S. DOE. 2006. Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda. Report from the December 2005 Workshop, DOE/SC-0095. U.S. Department of Energy Office of Science (www.genomicscience.energy.gov/biofuels/).
25. Karasov WH, Douglas AE. Comparative digestive physiology. Compr Physiol. 2013 Apr;3(2): 741-783
26. Zhang X, Sherman DM, Sherman LA. The uptake hydrogenase in the unicellular diazotrophic cyanobacterium Cyanothece sp. strain PCC 7822 protects nitrogenase from oxygen toxicity. J Bacteriol. 2014 Feb;196(4):840-849.
27. Dilworth MJ, Kennedy IR. Oxygen inhibition in Azotobacter vinelandii Some enzymes concerned in acetate metabolism. Biochim Biophys Acta. 1963;67: 240-253.
28. Alexov E. Numerical calculations of the pH of maximal protein stability. FEBS J. 2004 Jan;271(1):173-185
29. Pace CN, Grimsley GR, Sholtz JM. Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility. J Biol Chem. 2009 May; 284:13285-13289.
30. Vispo C, Hume ID. The digestive tract and digestive function in the North American porcupine and beaver. Can J Zool. 1995; 73(5): 967-974
31. Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufield JD, Charles TC. Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences.PLoS ONE. 2017;12(3): e0172545.
32. Nelson N. A photometric adaptation of the Somogyi method for the determination of glucose. J biol Chem. 1944 May 1;153(2):375-80.
33. Meddeb-Mouelhi F, Moisan JK, Beauregard M. A comparison of plate assay methods for detecting extracellular cellulase and xylanase activity. Enzyme Microb Technol. 2014 Nov 1;66:16-9.
34. Honda S. High-performance liquid chromatography of mono-and oligosaccharides. Anal Biochem. 1984 Jul 1;140(1):1-47.