Journal of Applied Biological Sciences
E-ISSN: 2146-0108 16(1): 46-61, 2022
DOI: https://doi.org/10.71336/jabs.918
Research Article

A PUTATIVE β-GLUCOSIDASE AND AN ENDO-1,4-β-GLUCANASE FROM POME METAGENOMIC DNA

Farah Fadwa Benbelgacem[1], Adibah Parman[1], Oualid Abdelkader Bellag[1], Nabila Akhyar[1], Md Zahangir Alam[1], Hamzah Mohd. Salleh[2*]

[1]*International Islamic University Malaysia, Kulliyyah of Engineering, Department of Biotechnology Engineering, Kuala Lumpur, Malaysia*
[2]*International Islamic University Malaysia, International Institute for Halal Research and Training (INHART), Kuala Lumpur, Malaysia*

**Corresponding Author:*
E-mail: hamzah@iium.edu.my

(Received 23rd March 2021; accepted 26th June 2021)

ABSTRACT. Functional metagenomic approach with high-throughput screening can be used to identify tapped and untapped biocatalysts. Metagenomic DNA libraries of 4.49 Gbase were constructed from microbes in Malaysian palm oil mill effluent (POME). After culture experiment based on natural selection metagenomic DNA was extracted and cloned to pCC1FOS vector and transformed into EPI300T1^{R}. Cellulose-degrading enzyme activity was screened with microtiter assay using methylumbelliferyl-β-D-glucopyranoside (MUGlc) and methylumbelliferyl-β-D-cellobioside (MUC) as fluorogenic substrates. Reads were normalized using robust ***z***-score and 100 highly rated clones were selected. Fosmids of these clones were isolated and sequenced with Hiseq strategy. Using Solexa, Velvet, SSPACE, Prodigal and Blastp, genes IDs of 96 putative cellulose-degrading enzymes were identified. Two putative metagenomic cellulose-degrading enzymes, MCDE1 with β-glucosidase activity and MCDE3 with endo-1,4-β-glucanase activity were produced, purified, and partially biochemically characterized.

Keywords: *Metagenomics, palm oil mill effluent (POME), high-throughput screening, next-generation sequencing, glucosidase, glucanase.*

INTRODUCTION

Amongst the highly demanded industrial enzymes are cellulose-degrading enzymes where they are used in various industries including detergents, fermentation, food, textile, pulp and paper as well as biomass saccharification for bioethanol production which seems to be the most popular application. Using these enzymes in enzymatic hydrolysis of biomass for bioethanol production may reduce the process cost greatly [1,2,3]. Cellulose-degrading enzymes are commonly produced by fungi, bacteria and actinomycetes during their growth on cellulosic and hemi-cellulosic materials; they may be thermophilic, mesophilic, aerobic or anaerobic.

Biocatalysts can be obtained through cultivation and subsequent screening of pure strains of microorganisms. It was later found out that only 1-15% of microbial genomes are cultivable under laboratory conditions while more than 85% have never been studied before [4]. A new method was then introduced namely metagenomics, to overcome the inability of identifying diversity present in the original environment. The new method

involves direct DNA isolation from an environmental sample followed by direct cloning for subsequent screening and product expression to allow unbiased genomic representation of the microbes. The success of the cloning may be proven by positive presence of novel proteins through function-based screening where the clones' biological activity is monitored [5].

Screening may be performed through two alternatives: sequence-based and activity or function-based screening. It is recommended to combine both methods to obtain the complete spectrum of the community. Sequence-based screening allows the identification of the sequence of interest while functional screening provides identification of unknown and novel genes that might not be recognizable by only sequence-based screening [6].

Adopting metagenomic approach has proven its efficiency to bioprospecting various glycosyl hydrolases, [7] were able to find one cellulase in 40,000 plasmid library while [8] have found 2 β-galactosidases in 2,843 BAC library, and [9] have found 5 cellulases in 11,520 fosmid library which reflect the importance of screening large metagenomic library to cover the microbiota in a special environment. More recently, [10] have reported finding a GH5 endo-β-1-4-glucanase from a fosmid library of 45,000 clones while [11] have identified 223 clones carrying a polyketide synthase (PKS) and or a non-ribosomal peptide synthase (NRPS) from a soil metagenomic library of 19,200 clones.

In the present work, fosmid library of palm oil mill effluent (POME) metagenomic DNA was manually constructed and screened by high-throughput screening microtiter test. Fluorogenic substrates were used as they are more sensitive than chromogenic substrates. The other essential point taken into consideration is the high-throughput screening data normalization and analysis which is an extremely important step in fluorescence-based assay comparing to absorbance measurements. Novel robust statistical method for drug discovery [12] was adopted and adapted, where *de facto* negative control samples instead of using negative and positive controls, to be efficient for fluorescence essays.

MATERIALS AND METHODS

POME sample was collected from fresh, cooled, and anaerobic ponds of FELDA Mempaga Mill, Bentong, Pahang, Malaysia. Metagenomic DNA cloning was conducted using CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector and Phage T-1 Resistant EPI300™-T1^{R} *E. coli* plating strain (Cat. No. CCFOS110, Epicentre). The fosmid DNA was extracted using FosmidMAX™ DNA Purification Kit (Epicentre, USA). Fosmids were sequenced with Illumina Hiseq 2000 at Malaysia Genome Institute (MGI), Bangi, Selangor, Malaysia.

Enrichment Strategy

For enrichment and metagenomic DNA library construction purpose, enrichment media was prepared and sterilised prior to sampling. On the day of sampling, 2 ml samples were added into 150 ml minimal medium consisting of 15.00 ml/l raw POME sludge, 0.5 g/L KH_2PO_4, 0.55 g/L $CaCl_2$, 0.13 g/L $MgSO_4.7H_2O$, 0.46 g/L yeast extract, 0.0056 g/L $ZnSO_4$, 0.009 g/L $CuSO_4$, 0.011 g/L $MnSO_4$ and 8.0 g/L carboxymethyl cellulose (CMC) or cellobiose respectively as enrichment substrate in separate flasks. All the media used were conditioned to have similar nature to the original habitat of the microorganisms, for

which the pH is adjusted to 4.2 as it is within the pH range of POME. This strategy is adopted and adapted from [13] and [14]. On the other hand, since sample was collected from outdoor fresh, cooled and anaerobic treatment pond, room temperature was used throughout the experiment. Initially, the growth medium contained minimum amount of each nutrients.

Metagenomic DNA Libraries Construction

DNA extraction protocol of [15] was used on both enriched and non-enriched samples, and the amount of extracted DNA was then measured using a spectrophotometer (VivaSpec LS, Sartorius Stedim, Germany). Thereafter, a library was created using CopyControl Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre, USA). Accordingly, the insert DNA was end-repaired followed by size selection. The DNA of interest was ~40 kb in size and was determined using low melting point gel electrophoresis (20 cm 1% agarose gel) for overnight run. The gel was not exposed to UV light or ethidium bromide to prevent decrease in cloning efficiency later. Therefore, only a fraction of the gel was stained and exposed to locate the position of the gel to be fractionated. DNA in gel was then recovered using GELase enzyme (Epicentre, USA). The recovered DNA was further ligated into pCC1FOS vector depending on the concentration of insert DNA quantified after end-repair process. Following that, ligated DNA was packaged and inserted into Phage T-1 Resistant EPI300™-T1R *E. coli* plating strain (Epicentre, USA). Finally, the culture was plated (LB agar + 12.5 µg/ml chloramphenicol) for library construction. Using sterile toothpicks, the library was constructed by carefully picking single colonies of EPI300-T1R culture. Each colony was inoculated to transparent 384-well plates filled with 80 µl of LB broth with sterile 10% glycerol and was incubated overnight in 37 °C humid incubator. The libraries were then kept at -80 °C freezer for further screening. Metagenomic DNA libraries construction is illustrated in Fig. 1A.

High Throughput Screening and Data Normalization

Colonies from each plate were inoculated into black 384-well plates filled with LB broth + 12.5 µg/ml chloramphenicol + 100 µg/ml L-arabinose (respective to their positions in the library plates) (Fig. 1B). Screening buffer (50 mM potassium acetate pH 5.5 and lysis mix (10% Triton x-100, 100 mM Tris pH 7.4 and 10 mM EDTA) were added to break open the cells. The fluorogenic substrates used were methylumbelliferyl-ß-D-glucopyranoside (MUGlc), methylumbelliferyl-ß-D-cellobioside (MUC) and chlorocoumarin-xylobioside (CCX). The fluorogenic substrates stock solutions were dissolved in DMSO and added to the assay mix of 50 µM of both MUGlc and MUC and 40 µM for CCX. After an overnight incubation at 37 °C without shaking, the clone's fluorescence was measured using Infinite F200PRO microplate reader (Tecan, Switzerland) for fluorescence that suggests the presence of the enzymes. The relative fluorescence units (RFU) given by the microplate reader were converted to robust ***z***-score to select the high rated hits. Robust ***z***-score was calculated for each microplate independently with the following equation (Eqn. 1) [12,16]:

$$Rz = \frac{s_i - med(s_{all})}{MAD(s_{all})}$$
$$MAD = 1.4826 \times med|(s_i - meds_{all})|$$

Eqn. 1

R_z= robust ***z***-score. med= median. MAD= median absolute deviation. s_i= sample value. S_{all}= all samples values.

Fosmid DNA Isolation and Sequencing

In Fig. 1C, the selected clones were auto-induced in 50 ml flasks; each contains 2 ml of SOC media, 12.5 µg/µl chloramphenicol and 100 µl of 500x auto-induction solution for each 50 ml of the media. The colonies were inoculated from the transparent microplates to the auto-induction preparation. The cultures grew overnight (19 h) at 37 °C with 250 rpm agitation. The fosmid isolation was performed the day after with the FosmidMaxTM DNA purification kit (Cat. No. FMAX046 from Epicentre, USA) following the manufacturer instructions. The integrity and size of the fosmids were checked by 0.8 % agarose gel electrophoresis (35 min / 120 V). Moreover, the concentration and purity ratios were given by MultiSkan Go spectrophotometer (Thermo-Scientific, USA). The fosmids which migrated in agarose gel appeared in different isotypes (more than one band). The fosmids were sequenced with Illumina Hiseq strategy at the Malaysia Genome Institute. The procedure followed to analyse the NGS-data is presented in [17]. In Figure 1D, the raw data of the Hiseq sequencer was the input of the Velvet software to arrange the data into contigs [18]. After filtering the fosmids sequences and checking the quality with SOLEXA. Besides, for the Phred, values less than 30 data were trimmed, filtered and checked again until the desired quality was obtained. SSPACE scaffolding software was used to order and orient the contigs and make them scaffolds [19]. Genes were further predicted with prodigal program [20] and genes IDs were assigned with blastp against NP.

Cloning of MCDE1 and MCDE3 Genes

The selected genes were recovered from the metagenomic DNA library by PCR amplification using FR-MCDE1 primer (5'-atgaaatggctttgtgtggtgag-3'), and rv-MCDE1 primer (5'-cttcagcgtcaagctatgcgcc-3'), FR-MCDE3 primer (5'-atgctgcaactcctgtcgct-3'), rv-MCDE3 (5'-ttttgaagttgtgcattggccg-3'). Amplified DNA was purified from 1% agarose gel (150 V / 50 min) and ligated to the pBAD TOPO vector with insert:vector molar ratio of 1:1. Recombinant DNA was transformed to the One Shot® TOP10 competent *E. coli* by heat-shock and plated in 100 µg/ml ampicillin LB agar plates.

Production of Recombinant MCDE1 and MCDE3

Transformed *E. coli* was inoculated, grown in 2 ml of LB containing 100 µg/ml ampicillin, and incubated overnight at 37 °C with shaking (225-250 rpm) to reach OD_{600} = 1 to 2. Two milliliters of the overnight culture was transferred to 1 liter LB media and L-arabinose (0.2% of w/v) was added to the culture. The pilot expression was then incubated at 20 °C (200 rpm for 16 h).

Purification of Recombinant MCDE1 and MCDE3

Cells were harvested by centrifugation and broken by ultrasonication method after adding binding buffer (20 mM sodium phosphate pH 7.4 with 0.5 M NaCl) to the cell pellet. Ten cycles of 10 seconds with 30 seconds intervals of ultrasonication were performed for cell lysis. Cell debris was eliminated by centrifugation (12,000 rpm for 30 min) and the supernatant was filtered with 0.45 µm membrane. HisTrap HP column for his6x tagged proteins (1 ml capacity) was employed for protein purification aided by a fast protein liquid chromatography (FPLC) system (GE Healthcare Life Science, USA). The column was equilibrated with 5 ml binding buffer (20 mM sodium phosphate pH 7.4 with 0.5 M NaCl) operated at 1 ml/min flow rate. After sample application (2 ml), the column was eluted with 20 ml binding buffer and bound proteins were eluted with elution buffer (binding buffer containing 0.5 M imidazole) and 0.5 ml fractions were collected for protein and enzyme assays.

Enzyme Assay

Recombinant *E. coli* clones were used in activity assays on LB plates supplemented with L-arabinose (0.2% w/v final concentration), ampicillin (100 µg/ml final concentration) and 1% final concentration of CMC. Activity from recombinant clones was detected by staining plates, after incubation at 37 °C for 24 h, with 1% (w/v) Congo red dye and de-stained with 1 M NaCl. Negative control TOP10 *E. coli* with non-recombinant pBAD TOPO vector was used following the same procedure. On the other hand, the FPLC eluted fractions which have indicated the presence of protein were tested with fluorogenic substrates MUG, MUGlc and CCX. Enzyme assay was carried out by mixing 30 µl potassium acetate buffer (50 mM, pH 5.5), eluted fraction (19 µl) and 1 µl substrate (50 µM final concentration) of were used for the enzymes assay. The negative control of the assay was 30 µl potassium buffer, 19 µl elution buffer and 1 µl of substrates. The expression lysate of each protein was also tested to check the presence of protein in the expressed cells. All assay tubes were incubated at room temperature and monitored hourly and qualitatively by handheld UV light starting from 'time of enzyme addition'.

Protein Determination, Molecular Weight Estimation and in silico Validation

Aliquots of eluted fractions were analyzed by SDS-PAGE [21] on 12% polyacrylamide gel (150 V / 55 min) and stained with Coomassie Brilliant Blue [22]. The molecular weight of each protein was estimated from the gene sequence length. For protein structure and function prediction based on available information of proteins presents in Protein Data Bank (PDB), amino acid sequences without signal peptide of MCDE1 and MCDE3 were submitted to the automated comparative protein modelling iTASSER server [23,24,25]. ScanProsite tool of the on-line bioinformatics resource operated by the Swiss Institute of Bioinformatics (https://prosite.expasy.org/scanprosite/) was used to detect the protein families and domains [26].

Biochemical Characterization

A standard curve of *p*-nitrophenol (*p*NP) was prepared at different concentrations (0 to 20 mM with 2.5 mM intervals). For chromogenic measurement of the catalytic activity of putative ß-glucosidase, 4-nitrophenyl-β-D-glucopyranoside (*p*NPG) was used and for putative glucanase, 4-nitrophenyl β-D-cellobioside (*p*NPC) was used. A reaction mixture

of 2 ml contained 50 mM citrate buffer, pH 4 with 10 mM *p*NPG/*p*NPC and 20 µg of purified enzyme. The mixtures were prepared in duplicate and incubated at 40 °C for 30 minutes. One hundred microliters of 1 M Na_2CO_3 was added to stop the reaction and *p*-nitrophenol product was measured at 405 nm (VivaSpec LS, Sartorius Stedim, Germany). The optimum pH for enzyme MCDE1 and MCDE3 activities were determined in the presence of 10 mM of *p*-nitrophenyl cellobioside (*p*NPC) over the pH range of 3.0 to 10.0 (50 mM sodium citrate buffer for pH 3.0 to 5.0), (50 mM citrate phosphate buffer for pH 6.0), (50 mM phosphate buffer for pH 7.0), and (50 mM Tris-HCl buffer for pH 8.0 to 10.0). Optimum temperature was determined by incubating the reaction mixture at 30-70 °C for 30 min in the presence of 10 mM *p*NPC and 50 mM citrate buffer (pH 4.0).

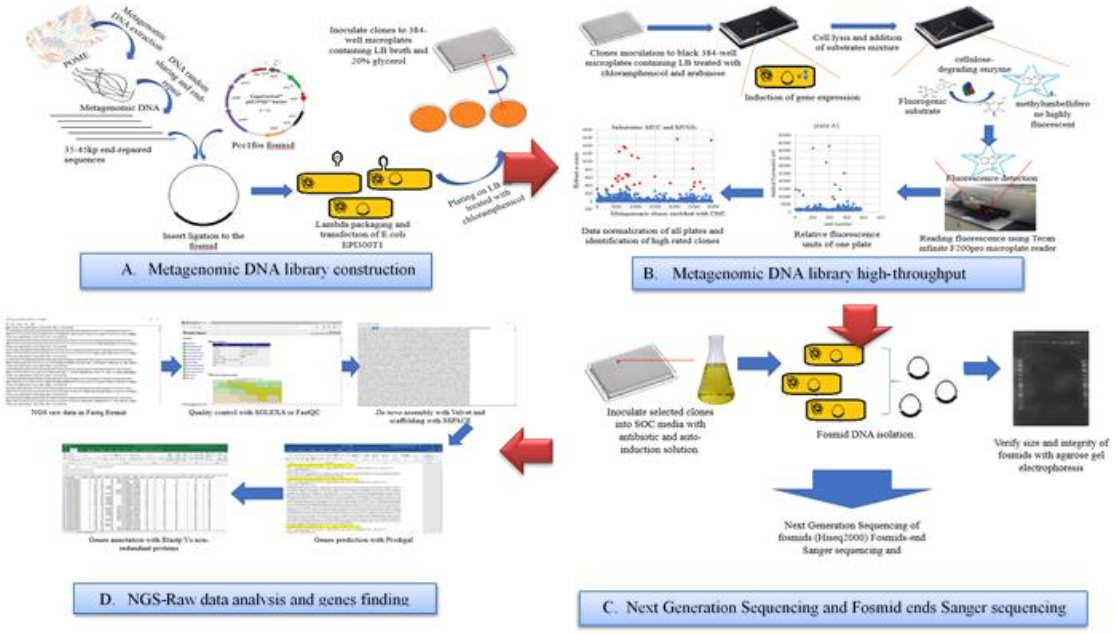

Fig. 1. *Overall scheme of bioprospecting cellulose-degrading enzymes from POME metagenomic DNA library.*

RESULTS AND DISCUSSION

Metagenomic DNA Libraries Construction and High-Throughput Screening of Metagenomic DNA Libraries

In the present study, an enrichment strategy was developed to increase the existence level of desired genes. The results of these steps have been reported elsewhere [27]. Construction of a large-sequence metagenomic DNA library is necessary to avoid problems such as the non-expression of enzymes due to the unrecognition of the promoter, ineffectiveness of translation, and imperfect post-translational modification [28,29,30]. There are many technical challenges in such approach; the first challenge when using phage transformation system is the length of required DNA to be packaged and this condition may not be assured while extracting metagenomic DNA even while using gentle lysis method [31,32,33]. The DNA size needs to be checked before ligation with pulse field gel electrophoresis (PFGE) or a large agarose gel electrophoresis method. The other challenge of large-insert metagenomic DNA library is the formation of

restricted and dephosphorylated vector as well as the packaging conditions have been reported [34,35,36,37].

It is also a challenge to develop effective and sensitive functional screening. Therefore, high-throughput screening (HTS) assay is introduced where screens are conducted in 384-well plate format to increase the efficiency and comparability between samples. Cell lysis was also imposed in this method to overcome intracellular accumulation of enzyme activity. Fluorogenic substrates, MUGlc and MUC (as well as CCX for xylanase, not reported) (Fig. 2) may interact with the enzyme released from the cells and allow the measurement of enzyme activity using fluorescence-based microplate readers [38]. In the screening, substrate mixture was used to check for more than one enzyme in one assay. A microplate reader (Tecan F200 Pro) with the suitable filter can detect fluorescence with high sensitivity and in a high-throughput manner when the fluorogenic substrates are cleaved with the desired enzyme [7]. The DNA sequencing step at a later stage will identify which enzyme is exactly present.

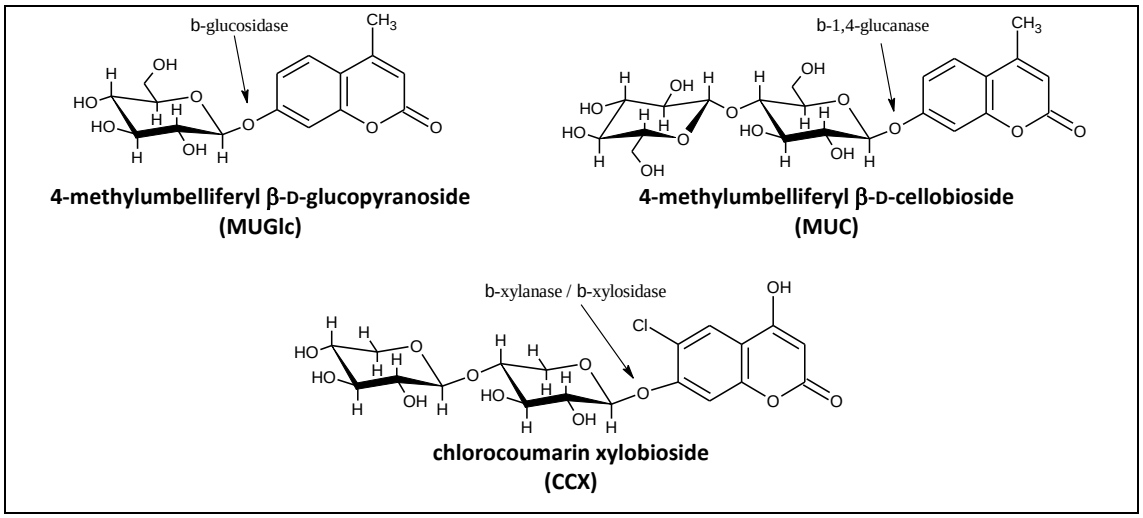

Fig. 2. *Fluorogenic substrates (MUGlc, MUC and CCX) used in HTS for detection of specific enzymes.*

Normalization of High-Throughput Screening Data

New statistical tool was adopted to confidently identify high rated clones. Robust statistical ways are able to find credible hits on sensitive and specific screening. HTS-data of 3072 clones from eight 384-well microplates (Fig. 3A), is represented in relative fluorescence unit (RFU), where the comparison between microplates is uncorrected due to the difference on population values distribution from microplate to another. Data normalization of HTS results was carried out by converting relative fluorescence unit to robust ***z***-score to eliminate errors and to combine all microplates' results in a proper way (Fig. 3B). Normalization of reads was performed separately to permit a correct positive clones' selection. In the current work, instead of using control-based screening, more accurate type of screening was followed by using the samples as *de facto* negative controls [12]. The HTS-normalized data of the four libraries is presented and in red are the 25 high rated clones from each library chosen as positive clones and their fosmids

were sequenced and analyzed to decode genes encoding putative cellulose-degrading enzymes (Fig. 4).

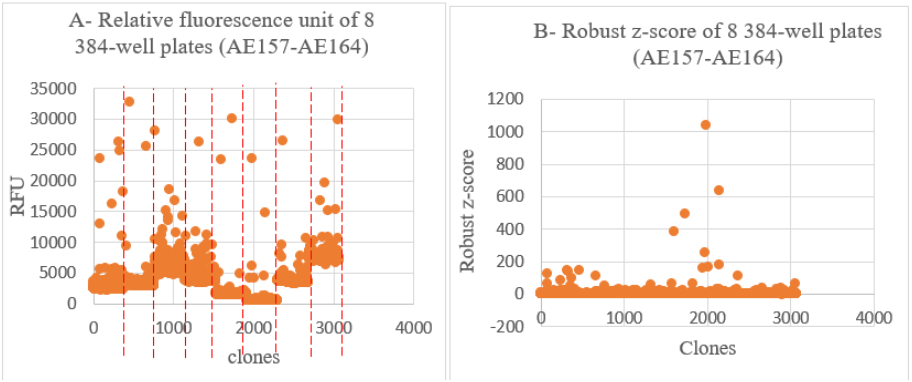

Fig. 3. (A) Scatter plot of high-throughput screened libraries represented in relative fluorescence unit. (B) Scatter plot of high-throughput screened libraries after data normalization represented in robust z-score.

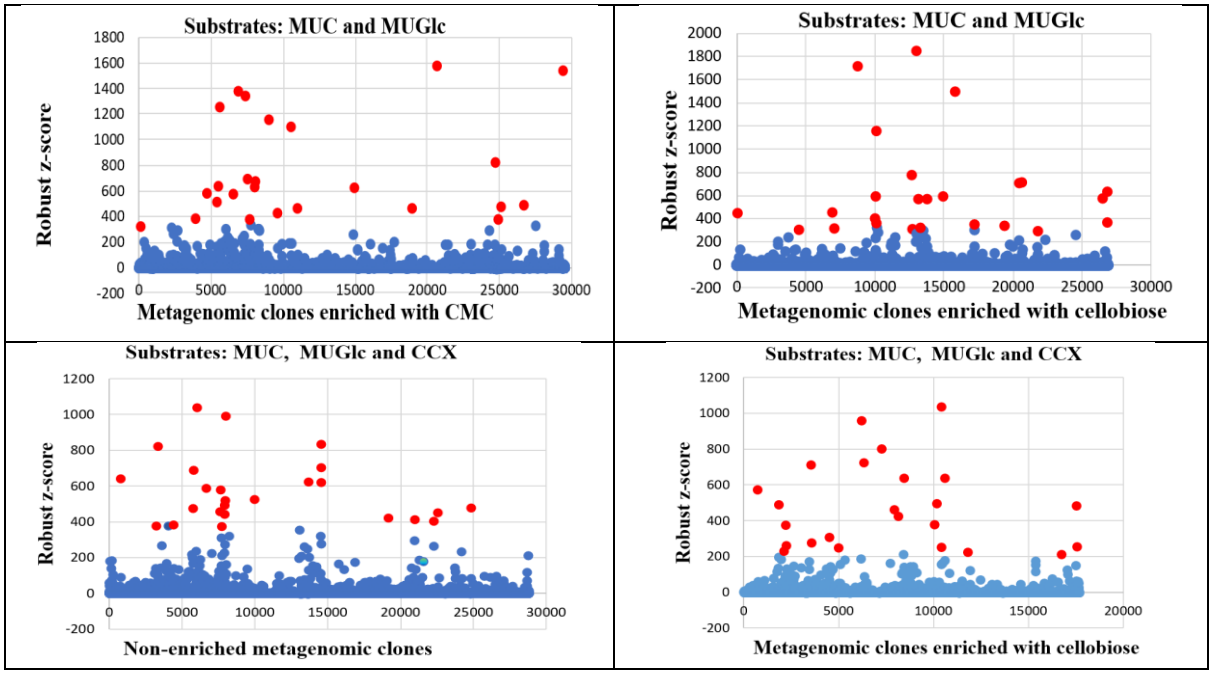

Fig. 4. *Scatter of robust **z**-score of high-throughput screened libraries. Red spots present the high rated clones chosen for sequencing.*

Next-Generation Sequencing and Genetic Analysis for Gene Finding

The highest one hundred rated clone were selected for genetic analysis whereby their fosmids DNA were isolated and pooled together to be sequenced using Hiseq strategy.

DNA sequences raw data were analyzed to identify gene encoding putative cellulose-degrading enzymes. Four thousand nine hundred and one contigs were obtained at the end of *de novo* assembly where the maximum contig's size was 2,036,785 bp. The number of scaffolds was 3.540 where the maximum scaffold size was 2,838,309 bp. At the end, 42,247 codon DNA sequences were obtained.

A total of 96 putative cellulose-degrading enzymes found in the metagenomic libraries are represented (Fig. 5). Protein sequences were compared to those available in the Protein Data Bank (PDB) and divided into 5 categories based of similarity percentages (Fig. 5a). The protein sequences of 15 putative cellulose-degrading enzymes are 100% similar to protein sequences available in protein databases, while 40 putative enzymes show similarities between 80-99%, 24 putative enzymes show similarities between 60-79%, 14 putative enzymes between 40-59% and 3 putative enzymes show less than 40% similarity. In Fig. 5b, the distribution of types of enzymes is represented; 19.20% are putative glucanases, 31.32% are putative beta glucosidases and 46.48% are putative glucoside hydrolases with cellulose-degrading enzyme conserved domains. In Fig. 5c, 26 microbial genus origins of all the enzymes which have sequence similarities to the bioprospected ones; 21 putative enzymes have similarities to enzymes in *Ochrobactrum* bacteria, 11 from *Phyllobacterum*, 10 from *Serratia*, 7 from *Pseudomonas*, 6 from *Candida*, 5 from *Scheffersomyces* and *Staphylococcus*, 4 from *Brucella*, 3 from *Ralstonia*, *Hyphopichia*, and *Aspergillus*, 2 from *Clavispora*, *Spathaspora*, *Burkholderia*, 1 from *Pichia*, *Bacilli*, *Cupriavidus*, *Ensifer*, *Wickerhamomyces*, *Cyberlindnera*, *Rhizobium*, *Kazachstania*, *Komagataella*, *Coccidioides*, *Streptococcus* and *Xanthomonas*.

Protein sequence information of the two putative cellulose-degrading enzymes MCDE1 and MCDE3 obtained from proteins database are represented in Table 1. MCDE1 enzyme was 99% similar to a ß-D-glucoside glucohydrolase from *Serratia marcescens* and MCDE3 was 99% identical to an endo-1,4-D-glucanase from *Serratia* sp. The crucial reason of choosing these two enzymes for further partial characterization is that the results available in the database are based on protein sequence homology only and there were no protein characterization or validation which may refer to a new enzyme. The enzyme MCDE1 top 6 threading proteins obtained from i-TASSER are: 3U48 and 3U48_B|PDBID|CH referred to a novel dimeric Family 3 ß-glucosidase isolated from compost using metagenomic analysis, 5z87A and 5z87_B|PDBID|CH referred to novel beta-glucosidase EmGH1 from *Erythrobacter marinus*, 4zo6A which is a mutant (D270A) ß-glucosidase from *Listeria innocua* in complex with cellobiose, and 5z9s which is a ß-glucosidase involved in saponin metabolism from intestinal bacteria. The alignment of MCDE1 to the top 6 threading proteins templates of i-TASSER by ScanProsite tool by Expasy has successfully detected the glycosyl hydrolase Family 3 active site from amino acid 273 to amino acid 290 (Fig. 6). The active site of glycosyl hydrolase was not detected in MCDE3 when the same procedure was followed.

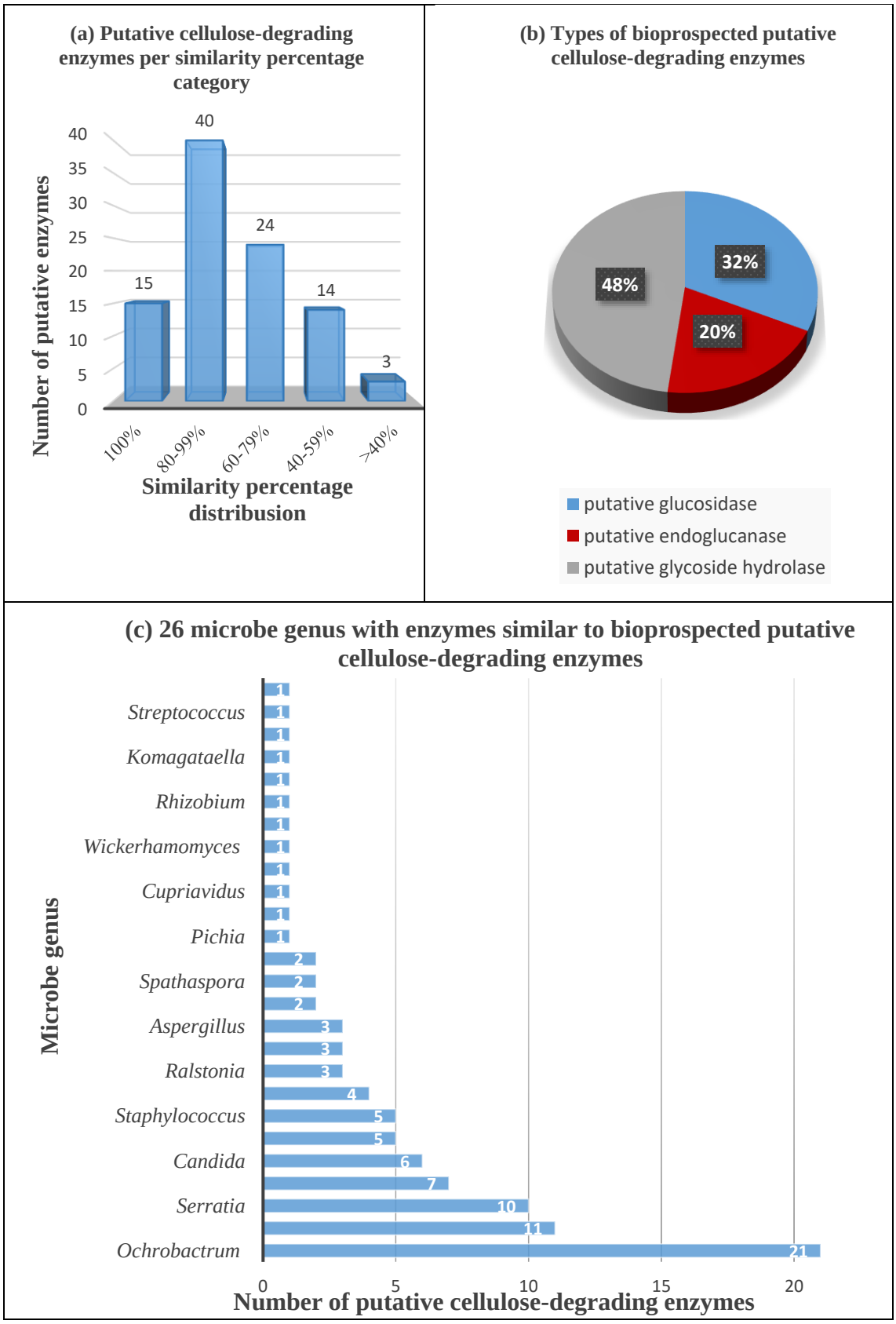

Fig. 5. *Putative cellulose-degrading enzymes found in the metagenomic libraries.*

Table 1. *Protein sequence information of the two putative cellulose-degrading enzymes MCDE1 and MCDE3 obtained from proteins database.*

	Putative protein detected in the protein database	Protein information in the protein database	Expected microorganism	Query/ Identity percentage	Expected protein size (kDa)
MCDE1	ß-D-glucoside glucohydrolase; (provisional)	Uncharacterized protein	*Serratia marcescens*	99%/99%	84.26
MCDE3	Putative endo-1,4-D-glucanase	Derived by automated computational analysis using gene prediction method: Protein Homology	*Serratia* sp.	94%/99%	40.48

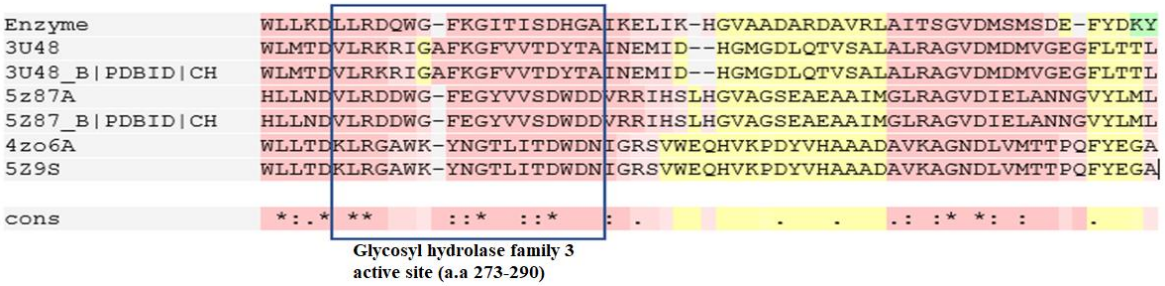

Fig. 6. *MCDE1 alignment to the top 6 threading proteins templates used by i-TASSER. The glycosyl hydrolase active site was detected by ScanProsite tool by Expasy.*

Partial Protein Purification and Characterization

MCDE1 and MCDE3 were successfully partially purified with immobilized metal affinity chromatography (IMAC) since the cloning vector should result into his6x tagged and they showed immediate response to fluorogenic substrates. MCDE1 shown intense release of fluorescence within the first hour while MCDE3 shown the highest intensity after four hours of assay. As shown in Fig. 7 and 8, the cell lysates and the three fractions of each enzyme exhibit fluorescence in the presence of 4-methylumbelliferyl-ß-D-glucopyranoside. It was expected for MCDE3 to hydrolyze the 4-methylumbelliferyl-ß-D-cellobioside as it is a putative endoglucanase, and this substrate is more specific to glucanase activity.

Under the assay conditions used in the study, the optimum enzyme activity of MCDE1 occurred at 60 °C and pH 7 and it displayed almost no activity at 80 °C and pH 9, whereas the enzyme activity of MCDE3 occurred at 50 °C and pH 4 and it displayed almost no activity at 70 °C and pH 9 (Fig. 7). Interestingly, an endoglucanase isolated from *Serratia proteamaculans* CDBB-1961 which has a protein size of 41.2 kDa close to the size of enzyme 3 (40.8 kDa) and show similarities to the amino acids of MCDE3, has different optimum temperature and pH, 40 °C and pH 7, respectively. This *Serratia proteamaculans* endoglucanase is the only characterized endoglucanase of *Serratia* sp. [39] and we have detected this divergence in optimum temperature and pH which may

suggest to the novelty of MCDE3. For further characterization, the expression level and specific activity of the enzyme should be increased; the method of [40] may be adopted where the sub-clones need to be synthesized in the methyltrophic yeast *Pichia pastoris* which can strongly over-expresses foreign proteins. In [41] work, the enzyme activity obtained from shake flasks was 0.029 U/ml and to increase the production of GH1254, fermentation was reperformed in 7.5 l Labfors bioreactor with 3 liter working volume and the obtained enzyme activity then was 0.48 U/ml (16x higher).

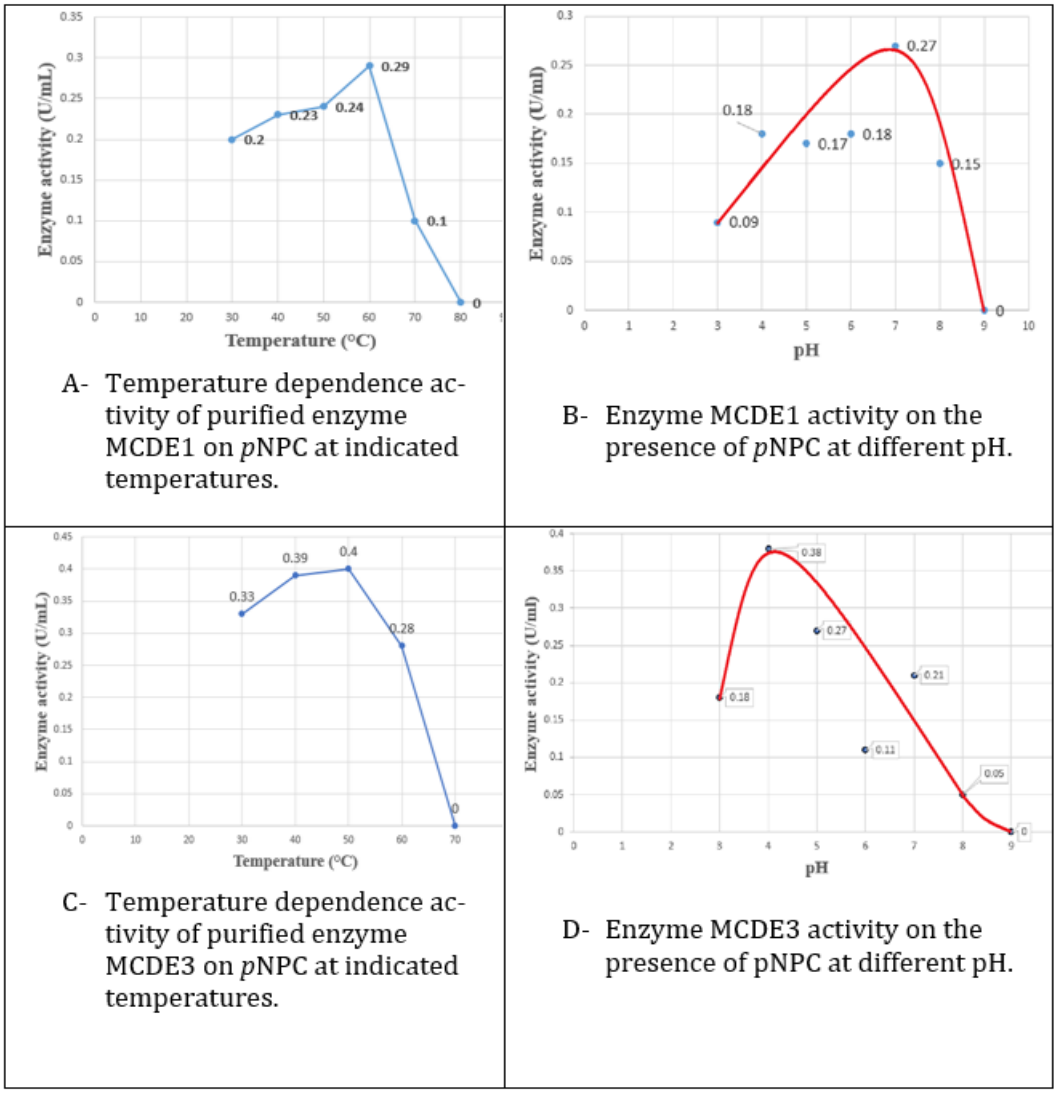

Fig. 7. *MCDE1 and MCDE3 activity at different pH and temperature.*

CONCLUSION

We have demonstrated a number of techniques/approaches that may help to improve the outcome of novel enzyme bioprospection. Metagenomics have proven to be efficient in elucidating the uncovered biodata in natural samples by screening the genome of

majority of microorganisms that are unculturable by genomic methods, thus widening the spectrum of novel enzymes and enzyme diversity. In an approach of functional metagenomics, where targeted enzymes are looked for in a constructed metagenomic library, an enrichment step, pre-culture may considerably increase the chances of finding the enzymes followed by a screening appropriate to the library size. In this context, NGS is as for now the only method capable of handling such big variant data for sequencing. POME has shown to contain cellulose degrading enzymes as hypothesized. We have screened a total of 110,000 variant clones and sequenced the 100 top hits. However, a considerable number of putative cellulose degrading enzymes have been found after the bioinformatic analysis of the raw NGS data using a pipeline of programs and method to clean, trim and assemble all the contigs obtained from the Hiseq sequencer. The constructed POME libraries hide a lot of enzymes to be discovered as we have only targeted cellulose degrading enzymes following the function-based metagenomics. Other rounds of NGS sequencing may also reveal many enzymes and analysis of the NGS raw data may as well reveal other enzymes than cellulose degrading enzymes. Further protein production, purification, and characterization of the metagenomic DNA libraries permit the bioprospecting of known and unknown biocatalysts.

Acknowledgement. The authors are thankful for a Fundamental Research Grant Scheme, Ministry of Higher Education, Malaysia (grant no. FRGS/2/2013/SG05/UIAM/01/3) that supports the research financially, and the Department of Biotechnology Engineering, IIUM for research facilities. We are indebted to Professor SG Withers (UBC, Canada) for CCX substrate and his valuable advice for the research project. We are also grateful to Professor Ibrahim A. Noorbatcha for his valuable discussion time and we acknowledge the excellent technical assistance from Maya Dehimi related to the research.

Conflict of Interest. The authors declared that there is no conflict of interest.

Authorship Contributions. Concept: H.M., Design: F.B., H.M., M.A., Data Collection: F.B., A.P., O.B., N.A., Analysis/Interpretation: F.B., A.P., H.M., Literature Search: F.B., O.B., A.P., Writing: F.B., O.B., Editing: M.A., H.M.

Financial Disclosure. Already stated in the Acknowledgement.

REFERENCES

[1] Kuhad, R. C., Gupta, R., & Singh, A. (2011): Microbial Cellulases and Their Industrial Applications. Enzyme Research, pp. 1–10.

[2] Ghimire, A., Kumar, G., Sivagurunathan, P., Shobana, S., Saratale, G. D., Kim, H. W., … Munoz, R. (2017): Bio-hythane production from microalgae biomass: Key challenges and potential opportunities for algal bio-refineries. Bioresource Technology. 241: 525–536.

[3] Armstrong, Z., Liu, F., Kheirandish, S., Chen, H.-M., Mewis, K., Duo, T., Morgan-Lang, C., Hallam, S.J., & Withers, S. G. (2019): High-Throughput Recovery and Characterization of Metagenome-Derived Glycoside Hydrolase-Containing Clones as a Resource for Biocatalyst Development. MSystems. 4(4). https://doi.org/10.1128/mSystems.00082-19.

[4] Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995): Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. Microbiological Reviews. 59(1): 143–169.

[5] Kumar, A. K. (2015): UV mutagenesis treatment for improved production of endoglucanase and β-glucosidase from newly isolated thermotolerant actinomycetes,

Streptomyces griseoaurantiacus. Bioresources and Bioprocessing. 2(1): 22. https://doi.org/10.1186/s40643-015-0052-x.

[6] Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004): Metagenomics: genomic analysis of microbial communities. Annual Review of Genetics. 38: 525–552.

[7] Martin, M., Biver, S., Steels, S., Barbeyron, T., Jam, M., Portetelle, D., van Michel, G., & Vandenbol, M. (2014): Identification and Characterization of a Halotolerant, Cold-Active Marine Endo-β-1,4-Glucanase by Using Functional Metagenomics of Seaweed-Associated Microbiota. Applied and Environmental Microbiology. 80(16): 4958–4967. https://doi.org/10.1128/AEM.01194-14.

[8] Vester, J., Glaring, M., & Stougaard, P. (2014): Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microbial Cell Factories. 13(1): 72. doi: 10.1186/1475-2859-13-72.

[9] Lewin, A., Zhou, J., Pham, V. T. T., Haugen, T., Zeiny, M. El, Aarstad, O., Leibl, W., Wenzel, A., & Liles, M. R. (2017): Novel archaeal thermostable cellulases from an oil reservoir metagenome. AMB Express. 7(1): 183. doi: 10.1186/s13568-017-0485-z.

[10] Wierzbicka-Woś, A., Henneberger, R., Batista-García, R. A., Martínez-Ávila, L., Jackson, S. A., Kennedy, J., & Dobson, A. D. W. (2019): Biochemical Characterization of a Novel Monospecific Endo-β-1,4-Glucanase Belonging to GH Family 5 From a Rhizosphere Metagenomic Library. Frontiers in Microbiology. 10. https://doi.org/10.3389/fmicb.2019.01342.

[11] Santana-Pereira, A. L. R., Sandoval-Powers, M., Monsma, S., Zhou, J., Santos, S. R., Mead, D. A., & Liles, M. R. (2020): Discovery of Novel Biosynthetic Gene Cluster Diversity from a Soil Metagenomic Library. Frontiers in Microbiology. 11. https://doi.org/10.3389/fmicb.2020.585398.

[12] Birmingham, A., Selfors, L. M., Forster, T., Wrobel, D., Kennedy, C. J., Shanks, E., Santoyo-Lopez, J., Dunican, D.J., Long, A., Kelleher, D., Beijersbergen, R.L., Ghazal, P., & Shamu, C. E. (2009). Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods, 6(8): 569–575.

[13] Rees, H. C., Grant, S., Jones, B., Grant, W. D., & Heaphy, S. (2003): Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. Extremophiles. 7(5): 415–421.

[14] Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K. E., & Streit, W. R. (2003): Prospecting for Novel Biocatalysts in a Soil Metagenome. Applied and Environmental Microbiology. 69(10): 6235–6242.

[15] Lee, S., & Hallam, S. J. (2009): Extraction of high molecular weight genomic DNA from soils and sediments. Journal of Visualized Experiments: JoVE. (33): 2–5.

[16] Asli, N., Sergio, C., & Chen, T. (2013): Data Analysis Approaches in High Throughput Screening. In Drug Discovery. InTech. doi: 10.5772/52508.

[17] Benbelgacem, F. F., Mat Isa, Mohd Noor Abdelrahim, M. A. M., Tumian, Afidalina, Abdelkader Bellag, O., Parman, Adibah Noorbatcha, I. A., & Mohd Salleh, H. (2018): Next Generation Sequencing-Data Analysis for Cellulose- and Xylan-Degrading Enzymes from POME Metagenome. Sains Malaysiana, 47(12): 2951–2960.

[18] http://dx.doi.org/10.17576/jsm-2018-4712-03 Zerbino, D. R., & Birney, E. (2008): Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research. 18(5): 821–829.

[19] Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011): Scaffolding pre-assembled contigs using SSPACE. Bioinformatics, 27(4): 578-579. doi: 10.1093/bioinformatics/btq683.

[20] Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010): Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11(1): 119. https://doi.org/10.1186/1471-2105-11-119.

[21] Laemmli, U. K. (1970): Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 227(5259): 680–685.
[22] Walker, J. M. (2002): Protein Protocols Handbook, The. New Jersey: Humana Press.
[23] Yang, J., & Zhang, Y. (2015): Protein Structure and Function Prediction Using I-TASSER. Current Protocols in Bioinformatics. 52(1): 5.8.1-5.8.15. doi: 10.1002/0471250953.bi0508s52.
[24] Roy, A., Kucukural, A., & Zhang, Y. (2010): I-TASSER: A unified platform for automated protein structure and function prediction. Nature Protocols. 5(4): 725–738.
[25] Y Zhang. (2008): I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9: 40. https://doi.org/10.1186/1471-2105-9-40.
[26] de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006): ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research. 34(Web Server), W362–W365. doi: 10.1093/nar/gkl124.
[27] Benbelgacem, F.F., Bellag, O.A., Ahmad, A. A., Noorbatcha, I.A., & Mohd. Salleh, H. (2017): Palm Oil Mill Effluent Metagenome for Cellulose-Degrading Enzymes. Journal of Basic and Applied Research. 3(3): 101–106.
[28] Uchiyama, T., & Kentaro Miyazaki. (2009): Functional metagenomics for enzyme discovery: challenges to efficient screening. Current Opinion in Biotechnology. 20: 616–622.
[29] Madhavan, A., Sindhu, R., Parameswaran, B., Sukumaran, R. K., & Pandey, A. (2017): Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. Applied Biochemistry and Biotechnology. 183(2): 636–651.
[30] Parks, R., & Graham, F. (1997): A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. Journal of Virology. 71: 3293–3298.
[31] Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996): DNA recovery from soils of diverse composition. Applied and Environmental Microbiology. 62(2): 316–322.
[32] Wexler, M., Bond, P. L., Richardson, D. J., & Johnston, A. W. B. (2005): A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environmental Microbiology. 7(12): 1917–1926.
[33] Kakirde, K. S., Parsley, L. C., & Liles, M. R. (2010): Size does matter: Application-driven approaches for soil metagenomics. Soil Biology and Biochemistry, 42(11): 1911-1923. doi: 10.1016/j.soilbio.2010.07.021
[34] Craig, J. W., Chang, F.-Y., Kim, J. H., Obiajulu, S. C., & Brady, S. F. (2010): Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria. Applied and Environmental Microbiology. 76(5): 1633–1641.
[35] Troeschel, S. C., Drepper, T., Leggewie, C., Streit, W. R., & Jaeger, K.-E. (2010): Novel Tools for the Functional Expression of Metagenomic DANN, pp. 117–139.
[36] Cheng, J., Pinnell, L., Engel, K., Neufeld, J. D., & Charles, T. C. (2014): Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. Journal of Microbiological Methods. 99(1): 27–34.
[37] Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., & Charles, T. C. (2015): Current and future resources for functional metagenomics. Frontiers in Microbiology, 6. https://doi.org/10.3389/fmicb.2015.01196.
[38] Taupp, M., Mewis, K., & Hallam, S. J. (2011): The art and design of functional metagenomic screens. Current Opinion in Biotechnology. 22(3) 465–472.
[39] Cano-Ramírez, C., Santiago-Hernández, A., Rivera-Orduña, F. N., García-Huante, Y., Zúñiga, G., & Hidalgo-Lara, M. E. (2016): Expression, purification and characterization of an endoglucanase from Serratia proteamaculans CDBB-1961, isolated from the gut of *Dendroctonus adjunctus* (Coleoptera: Scolytinae). AMB Express, 6(1): 63. doi: 10.1186/s13568-016-0233-9

[40] Cregg, J. M., Vedvick, T. S., & Raschke, W. C. (1993): Recent advances in the expression of foreign genes in Pichia pastoris. Nature Biotechnology. 11(8): 905–910.
[41] Woon, J. S., Jie, P., King, H., Mohamed, M., Mahadi, N. M., Mohd, W., & Wan, K. (2017): Cloning, Production and Characterization of a Glycoside Hydrolase Family 7 Enzyme from the Gut Microbiota of the Termite Coptotermes curvignathus. Molecular Biotechnology. 59(7): 271–283.