*DNA Research*, 2017, 24(6), 649–656
doi: 10.1093/dnares/dsx032
Advance Access Publication Date: 7 August 2017
Full Paper

OXFORD

Full Paper

# Isolation of highly thermostable β-xylosidases from a hot spring soil microbial community using a metagenomic approach

**Masaru Sato[1], Migiwa Suda[1], Jiro Okuma[2], Tomohiko Kato[1], Yoshitsugu Hirose[2], Asuka Nishimura[2,†], Yasuhiko Kondo[2], and Daisuke Shibata[1,*]**

[1]Department of Biotechnology Research and Development, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan, and [2]Honda Research Institute Japan Co., Ltd., Wako, Saitama 351-0188, Japan

*To whom correspondence should be addressed. Tel. 81 438 52 3947. Fax: +81 438 52 3948. Email: shibata@kazusa.or.jp

[†]Present address: Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

Edited by Dr. Katsumi Isono

Received 17 February 2017; Editorial decision 12 July 2017; Accepted 31 July 2017

## Abstract

The DNA extracted from a high-temperature environment in which micro-organisms are living will be a good source for the isolation of thermostable enzymes. Using a metagenomic approach, we aimed to isolate thermostable β-xylosidases that will be exploited for biofuel production from lignocellulosic biomass. DNA samples obtained from the soil near a spout of a hot spring (70°C, pH7.2) were subjected to sequencing, which generated a total of 84.2 Gbp with 967,925 contigs of >500 bp in length. Similarity search for β-xylosidase in the contigs revealed the presence of 168 candidate sequences, each of which may have arisen from more than one gene. Individual genes were amplified by PCR using sequence-specific primers. The resultant DNA fragments were cloned and introduced into *Escherichia coli* BL21 Star(DE3). Consequently, 269 proteins were successfully expressed in the *E. coli* cells and then examined for β-xylosidase activity. A total of 82 proteins exhibited β-xylosidase activity at 50°C, six of which retained the activity even at 90°C. Out of the six, three proteins were originated from a single candidate sequence, AR19M-311. An amino acid sequence comparison suggested the amino acid residues that appeared to be crucial for thermal stability of the enzymes.

**Key words:** thermostable enzyme, β-xylosidase, metagenome, xylan-degradation

## 1. Introduction

β-Xylosidases, which cleave the β-1,4-linkage of xylan from the non-reducing terminus to release D-xylose, have received attention in the past decade with respect to biofuel production from lignocellulosic biomass due to their potential to degrade hemicellulose, the second most abundant lignocellulose biomass, coordinately with endoxylanases and some glycosidases, such as arabionosidases and gluclonidases.[1] Although the sequences of over 800 β-xylosidases have been registered in the enzyme database BRENDA,[2] there is interest in the identification of thermostable β-xylosidases, because enzymatic processing at high temperatures prevents contamination with saprophytic bacteria and lowers the viscosity of the reaction

© The Author 2017. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

mixture, thus increasing the high reaction efficiency and reducing energy consumption during agitation of the mixture.[3] Thermostable β-xylosidases have been isolated from thermophilic micro-organisms from hot compost material and hot spring soils. The thermophilic bacteria *Thermotoga maritima* and *T. thermarum* produce β-xylosidases active at temperatures above 90 °C.[4,5] Thermostable β-xylosidases from *Caldicellulosiruptor bescii* and *Thermoanaerobacter ethanolicus* have also been reported.[6,7] A thermostable β-xylosidase exhibiting an optimum reaction temperature of 85 °C was isolated from the archea *Sulfolobus solfataricus*.[8] A β-xylosidase isolated from the fungus *Aureobasidium pullulans* was reported to be active at 80 °C.[9] No amino acid sequence motifs clearly associated with the thermal stability of these thermostable enzymes have been identified, however. The characterization of additional thermostable β-xylosidases will hopefully enhance our understanding of the molecular features associated with thermal stability, which in turn could facilitate genetic engineering of hydrolytic enzymes suitable for industrial use.

Recent advancements in next-generation sequencing (NGS) technologies now enable researchers to more easily access the vast stores of DNA sequence data for mixed micro-organisms, known as the metagenome. As such, the use of metagenomic approaches to identify enzymes of interest has increased in recent years in many areas of biological research.[10] Although β-xylosidases have been isolated from sources of mixed DNA (such as compost or yak rumen microbes, for example) using traditional screening of genomic libraries,[11,12] no attempts have been made, to our knowledge, to isolate thermostable β-xylosidases using an NGS-based metagenomic approach.

In this study, we employed NGS to obtain an 84.2-Gbp sequence from a microbial community inhabiting a 70 °C hot spring. The sequence contained 168 open reading frames (ORFs) encoding β-xylosidase/β-glucosidase enzymes, most of which were active when expressed in *Escherichia coli*. Among the isolated enzymes, six β-xylosidases, including ones originated from clones with single nucleotide polymorphisms (SNPs), were highly thermostable, exhibiting activity at 90 °C. Interestingly, one protein, AR19M-311-27, lost activity at 90 °C, while three other homologous proteins were active at 90 °C. Based on these data, amino acid residues important for maintaining thermal stability were tentatively identified.

## 2. Materials and methods

### 2.1. Metagenomic DNA extraction and sequencing

The soil sample was collected by scratching the surface of a spout of a hot spring (70 °C and pH 7.2 when sampled) by a spatula in the Onikobe-onsen-kyo geothermal area at lat. 38°48' and at long. 140°40' (Miyagi, Japan) in June, 2009. Metagenomic DNA was extracted from microbes in 10 g of the soil using an ISOIL Large for Beads, ver. 2, DNA extraction kit (Nippon Gene, Tokyo, Japan). Five micrograms of the extracted DNA were sequenced using a GS FLX Titanium DNA sequencer (Roche Diagnostics, Basel, Switzerland) according to the supplier's protocol. In addition, the genomic DNA was amplified using a GenomiPhi V2 DNA amplification kit (GE Healthcare, Little Chalfont, UK) and then sequenced using a Hiseq 2000 DNA sequencer (Illumina, San Diego, CA, USA) according to the supplier's protocol. Sequences with a total length of 1.11 Gbp and average length of 396 bp for each read were obtained from the former sequence, and sequences with a total length of 83.1 Gbp and average length of 92.7 bp for each read were obtained from the latter sequence. Uncertain sequences were trimmed from the raw sequences based on the quality score defined, and the resulting sequences were used for *de novo* assembly using CLC Genomics Workbench software, ver. 4.0.

### 2.2. Prediction of the genes encoding β-xylosidases

Genes and their ORFs in the assembled contigs were extracted using MetaGeneAnnotator software,[13] with the multiple species option for metagenome analysis, as recommended by the software supplier. The homology to β-xylosidase (E.C. 3.2.1.37) and β-glucosidase (E.C. 3.2.1.21) amino acid sequences of the extracted ORFs were searched using BLAST (Supplementary Table S1) with sequences downloaded from the UniProt database, release date November 2011 (http://www.uniprot.org). The list contained 2,140 amino acid sequences, although the latest release (2017_06) contain 13,585 sequences for the enzymes, which could be useful for future research. ORFs exhibiting homology to the amino acid sequences of β-glucosidases/β-xylosidases with an expected value $<1 \times 10^{-20}$ were categorized as glycosyl hydrolase (GH) family members, with the hydrolase motifs defined in the Pfam database, version 23.0 (http://pfam.sanger.ac.uk), using HMMER software, version 2.3 (hmmer.org).

### 2.3. Cloning of the predicted genes

Predicted gene sequences were amplified from the metagenomic DNA using KOD-FX DNA polymerase (Toyobo, Osaka, Japan) with PCR primers shown in Supplementary Table S2. The PCR-amplified fragments with four additional bases (CACC) at the 5'-terminus for directional TOPO® cloning were then inserted into the pET101/D-TOPO vector (Thermo Fisher Scientific, Waltham, MA, USA) and introduced into *E. coli* TOP10, which was a host for the vector recommended by the supplier for efficient cloning. The recombinant plasmids were extracted from the *E. coli* culture using a Wizard plus SV Minipreps DNA Purification System (Promega, Fitchburg, WI, USA). The inserted sequences of the plasmids were analysed on an ABI 3730 DNA analyser (Thermo Fisher Scientific).

### 2.4. Preparation of cell-free extracts

To produce the proteins, the plasmids isolated from *E. coli* TOP10 were introduced into *E. coli* BL21 Star (DE3) (Thermo Fisher Scientific), which is designed to efficiently express foreign DNA. The *E. coli* colonies were inoculated into 10 ml of Luria-Bertani medium containing 100 μg/ml of ampicillin and cultivated until the optical density at 600 nm reached approximately 0.2 to 0.8. Isopropyl-β-D(−)-thiogalactopyranoside was then added to the culture to induce expression of the foreign DNA encoded by the vector, and cultivation was continued for 16 h. The cells were harvested from the cultures by centrifugation and suspended in 1 ml of 50 mM Tris-HCl buffer (pH 8). The cell suspensions were homogenized using an Astrason 3000 homogenizer (Misonix, Farmingdale, NY, USA) and filtered using a 0.45-μm pore-size filter to prepare cell-free extracts used for the initial assay shown in Supplementary Fig. S2.

To assess the thermal stability of the all β-xylosidases, recombinant *E. coli* cells were inoculated into 1 ml of Overnight Express™ instant LB medium (Merck, Darmstadt, Germany) containing 50 μg/ml of carbenicillin and cultivated for 20 h at 30 °C with rotary shaking at 1,000 rpm. The cells were harvested by centrifugation at 3,000 ×g for 10 min, washed twice with 50 mM sodium acetate buffer (pH 5.0), and lysed by freeze-thaw cycling in 200 μl of the buffer. The resulting lysates were centrifuged at 3,000 ×*g* for 10 min to remove cell debris and used for the assay.

In the assay for determining the optimal reaction temperature and pH and thermal stability of the proteins originated from AR19M-311, recombinant *E. coli* cells were inoculated into 100 ml of Overnight Express$^{TM}$ instant LB medium containing 50 µg/ml of carbenicillin and cultivated for 20 h at 30 °C with rotary shaking at 180 rpm. The cells were then washed twice with 50 mM sodium acetate buffer (pH 5.0), homogenized using a Sonifier 250 homogenizer (Branson Ultrasonics, St. Louis, MO, USA), and heat-treated at 70 °C for 10 min. After centrifugation at 20,000 ×*g* for 30 min, the supernatants were subjected to the β-xylosidase activity assay.

### 2.5. Enzyme purification by column chromatography

The cell-free extract was fractionated by anion exchange chromatography using an ÄKTA FPLC (GE Healthcare) system equipped a HiTrap Q HP column (GE Healthcare) equilibrated with 50 mM Tris-HCl buffer. The active fractions were collected and applied onto a HiTrap Phenyl HP column (GE Healthcare) for hydrophobic interaction chromatography. The enzyme was finally desalted by gel permeation chromatography using a PD-10 column (GE Healthcare).

### 2.6. Enzymatic activity assay

Hydrolysis of *p*-nitrophenylxylopyranoside (*p*NPX) by cell-free extracts containing the recombinant proteins was evaluated by determining *p*-nitrophenol release by measuring the absorbance at 400 nm. The portion (50 µl) of cell-free extracts were mixed with 1.7 mM of *p*NPX in 200 µl of 50 mM acetate or sodium phosphate buffer (pH 5.5 or 6.8) and incubated for 20 min. After the reaction, the same volume of 200 mM $Na_2CO_3$ was added to the reaction mixture, and the absorbance at 400 nm was determined.

Thermal stability was evaluated at 50°C and 90 °C in 50 mM sodium acetate buffer (pH 5.0). *p*NPX was used as the substrate to assay β-xylosidase activity in crude cell-free extract. Before the assay, the cell lysates were incubated at 50°C or 90°C for 15 min. The assay was performed by adding 50 µl of the cell lysates to 200 µl of buffer containing 1.0 mM *p*NPX at 50°C for 15 min. Following addition of the same volume of 200 mM $Na_2CO_3$, the release of *p*-nitrophenol was measured by determining the absorbance at 400 nm. Instead of the cell lysate, water was added to the reaction mixture for the blank assay. One unit of the enzyme activity was defined as the amount of the enzyme releasing 1 µmol of *p*-nitrophenol per minute.

The optimal reaction temperature (50° C to 90° C) and pH (2 to 10) of the proteins originated from AR19M-311 were assayed in 50 mM GTA buffer, which consisted of 50 mM 3,3-dimethylglutaric acid, 50 mM 2-amino-2-methyl-1,3-propanediol, and 50 mM tris(hydroxymethyl)aminomethane. After the reaction, the amount of *p*-nitrophenol released was monitored using the same method described above. For assays of the thermal stability of the proteins originated from AR19M-311, the enzymes were incubated at temperatures ranging from 50°C to 90° C for 0 to 60 min in 50 mM sodium acetate buffer (pH 5.0). After incubation, the residual activity was measured at 90°C for 15 min using *p*NPX as the substrate.

Concentrations of proteins were determined by Coomassie Plus (Bradford) Assay Kit (Thermo Fischer Scientific) with bovine serum albumin as a standard.

### 2.7. Phylogenetic analysis

The amino acid sequences of known β-xylosidases were obtained from either UniProt database entries assigned as β-xylosidases or the BRENDA database containing thermostable β-xylosidases, and the sequences were then reviewed. These sequences and that of AR19M-311 were aligned using ClustalX ver. 2.0[14] with default parameters; subsequently, the phylogenetic tree was generated. The phylogenetic tree was visualized using FigTree ver. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

## 3. Results

### 3.1. Sequencing the metagenomic DNA of a hot spring microbial community

Metagenomic DNA was extracted from a microbial community isolated from the soil of a hot spring (70 °C, pH 7.2) in Miyagi, Japan, and then subjected to NGS using a GS FLX Titanium DNA sequencer. The average length of the reads produced from each DNA fragment was 396 bp. The assembled sequences, totaling 1.1 Gbp, were searched against a list of amino acid sequences of β-xylosidases/β-glucosidases downloaded from the UniProt database (Materials and methods), resulting in the identification of 54 ORF sequences homologous to β-xylosidases/β-glucosidases. To find more genes, we then sequenced a portion of the DNA amplified from the metagenomic DNA using a Hiseq 2000 DNA sequencer, which has a higher sequencing output than the former sequencer. The resulting sequences (83.1 Gbp) were assembled with the previously generated 1.1 Gbp sequence. A total of 967,925 contigs with a length >500 bp were finally obtained, of which the longest contig was 287,641 bp.

### 3.2. Isolation of a gene encoding a thermostable β-xylosidase

From the assembled sequences, 168 genes potentially belonging to GH families 1, 3, 31, and 43 were obtained from ORF prediction using MetaGeneAnnotator, followed by homology searching of the ORFs using the programs BLAST and HAMMER. A total of 105 ORFs appeared to be full-length, and the remainder lacked 5'- and/or 3'-terminal sequences. The ORFs exhibited a wide distribution of amino acid sequence similarity with known sequences of β-xylosidases and β-glucosidases in the UniProt database, as shown in Fig. 1. Most of the ORFs exhibited <80% homology with known sequences, indicating that metagenomic approaches are useful for identifying novel sequences.

Using PCR, we amplified DNAs corresponding to ORFs, and cloned the resulting products into the pET101/D-TOPO vector using *E. coli* TOP10 as a cloning host. The CACC sequence, necessary for TOPO® cloning, was added on the 5'-terminal of the forward primer. Each of the DNA primer set was designed to amplify the region from initial codon to terminal codon of each full-length ORF. However, when amplifying the ORFs lacking 5'- and/or 3'-terminal sequences, the DNA primer sets were designed to amplify the region of the longest in-frame sequences of the ORFs, and added lacking initial and/or terminal codons at the terminals. One to six independent clones for each ORF were isolated to confirm DNA insertion by partial sequencing. In total, 269 genes originating from 120 ORFs were successfully cloned. These 269 genes were consisted of the ones that were identical to the original ORF sequence and the others that possessed SNPs. Then we distinguished all of the genes by adding branch numbers. The host of the plasmids was then changed from *E. coli* TOP10 to *E. coli* BL21 Star (DE3), which was more suitable for protein production. Cell-free lysates of the new host were used for evaluating β-xylosidase activity using *p*NPX as the substrate (Supplementary Table S2). When assayed at 50 °C, 82 crude cell-free extracts showed activity against the substrate (Fig. 2A). However, only six extracts retained the activity when assayed at 90 °C,

indicating that most of the enzymes were not thermostable at higher temperatures (Fig. 2B). The highly thermostable enzymes included three proteins originated from AR19M-311 (AR19M-311-2, AR19M-311-11, and AR19M-311-21), each single protein originated from AR19M-319 (AR19M-319-21), AR19M-329 (AR19M-329-1), and AR19M-333 (AR19M-333-4). AR19M-311-27 originated from AR19M-311 was lost its activity at 90°C, while three other proteins were highly thermostable. Then, we further investigated the enzyme characteristics of proteins originated from AR19M-311 to reveal the relationship between thermostability and SNPs.

### 3.3. Effects of temperature and pH on the hydrolysis of *p*NPX by the β-xylosidases

We characterized the proteins originated from AR19M-311 further after removal of host-derived proteins by heat treatment at 70 °C, which caused most proteins in the lysate, except the heat-stable enzymes, to precipitate. A protein, AR19M-311-27, lost activity during incubation at 70 °C. AR19M-311-2, AR19M-311-11, and AR19M-311-21 showed similar optimum temperatures (Fig. 3A). These enzymes showed an optimal pH of 5.0, but no activity was detected at pH 4.5 when assayed at 90 °C, whereas AR19M-311-27 showed negligible activity at all pH values examined (Fig. 3B). The drastic difference in activity over such a narrow pH range (5.0 to 4.5) suggests that there are structural differences between the pH range, and this will be an interesting target of research when the three-dimensional structures of the enzymes become available. Similar pH dependency was reported for a β-xylosidase from *Enterobacter* sp.[15]

As the three enzymes AR19M-311-2, AR19M-311-11 and AR19M-311-21 showed almost the same optimum pH and temperature for the enzymatic reaction (Fig. 3), only AR19M-311-2 was further purified by column chromatography using HiTrap Q HP and HiTrap Phenyl HP columns to determine the kinetic parameters under the optimal reaction conditions using *p*NPX as the substrate. The $k_{cat}$ of AR19M-311-2 was $1.30 \times 10^2\ s^{-1}$, and the *K*m for *p*NPX was 0.872 mM. The lineweaver-Burk plot for the determination of the kinetic parameters was shown in Supplementary Fig. S1.

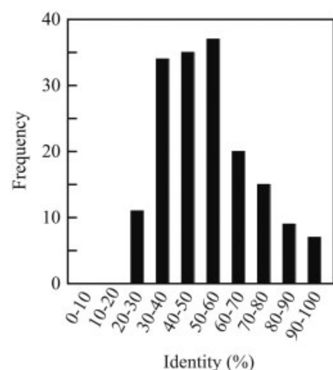


**Figure 1.** Frequency distribution of the 168 ORFs identified by identity searches against known β-xylosidases/β-glucosidases contained in the UniProt database.

### 3.4. Thermostability of the β-xylosidases

We analysed the thermal stability of AR19M-311-2, AR19M-311-11, and AR19M-311-21 by incubating the partially purified enzymes without substrate at 50°C, 60°C, 70°C, 80°C, and 90 °C and pH 5.0 for 0 to 60 min (Fig. 4). The enzymes retained above 80% of their initial activities at 50°C, 60°C, and 70°C for more than 60 min, except AR19M-311-2 whose activity was decreased to 50% after incubation at 70°C for 60 min. The three enzymes lost 50% of their activities within 10 min at 80°C. While the activity was immediately lost when the substrate *p*NPX was absent at 90°C, the hydrolysis proceeded at 90°C at least 15 min as shown in Fig. 3 when the substrate existed. Hence, it suggests that the substrate stabilize these enzymes. When the crude cell-free extracts were used, the activities were retained after incubation at 90°C without the substrate as shown in Fig. 2, indicating the enzymes were stabilized in the extract.

### 3.5. Classification of AR19M-311 and other β-xylosidases based on sequence homology

We searched for proteins similar to the ones originated from AR19M-311, which consisted of 752 amino acid residues, in the National Center for Biotechnology Information amino acid sequence database. The highest identity (71%) was exhibited by a predicted xylosidase/arabinosidase encoded by the genome of the bacterium *Dictyoglomus thermophilum*, although there is no experimental evidence confirming that this protein is a hydrolase. Among the known thermostable β-xylosidases, the enzyme shared the highest homology (56%) with a β-xylosidase from the archaeon *S. solfataricus*[8] and the

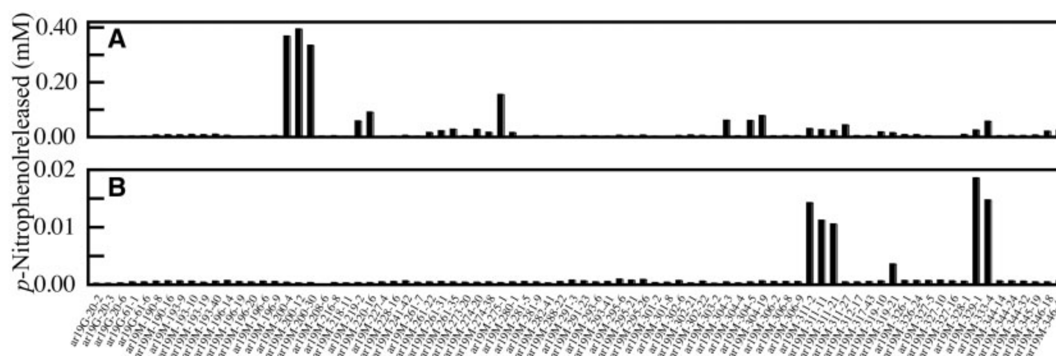


**Figure 2.** The amount of *p*-nitrophenol released from the hydrolysis of *p*NPX by the crude cell-free extracts with the 82 β−xylosidases isolated from the metagenomic DNA. The enzymes were incubated at 50 °C (A) or 90 °C (B) for 15 min, then subjected to the β-xylosidase assay using *p*NPX as the substrate. After the reaction, the amount of *p*-nitrophenol was measured to detect the residual activity.

lowest homology (29%) with a β-xylosidase from the fungus *A. pullulans*.[9]

The phylogenetic relationship between AR19M-311 and other β-xylosidases was determined using the ClustalX program (Fig. 5). AR19M-311 was classified in a thermostable β-xylosidase cluster that included like enzymes from *S. solfataricus* and the bacteria *T. thermarum*, *T. maritima*, *Thermoanaerobacter ethanolicus*, and *Caldicellulosiruptor bescii*.

### 3.6. Amino acid substitutions of proteins originated from AR19M-311

The consensus amino acid sequence of AR19M-311 and those of the four proteins were compared, as shown in Fig. 6. Amino acid substitutions were observed for 11 residues ($Thr^{57}$, $Arg^{187}$, $Val^{194}$, $Ala^{206}$, $Thr^{311}$, $Val^{355}$, $Ala^{451}$, $Gly^{519}$, $Ile^{625}$, $Val^{635}$, and $Lys^{739}$) of the consensus sequence. Comparing the thermostable proteins, AR19M-311-2, AR19M-311-11, and AR19M-311-21, and the non-thermostable protein AR19M-311-27, there were five residues that were substituted only with non-thermostable protein ($Thr^{57}$ to $Ile^{57}$, $Ile^{311}$ to $Thr^{311}$, $Ala^{451}$ to $Thr^{451}$, $Gly^{519}$ to $Ser^{519}$, and $Lys^{739}$ to $Asn^{739}$). AR19M-311-2 exhibited unique substitutions of $Val^{355}$ with Ala and $Val^{635}$ with Ile, but these substitutions might not be crucial for maintaining high activity because they were not found in the other thermostable proteins. Interestingly, $Arg^{187}$, $Val^{194}$, and $Ala^{206}$ were only seen in the consensus sequence and not in the four proteins.

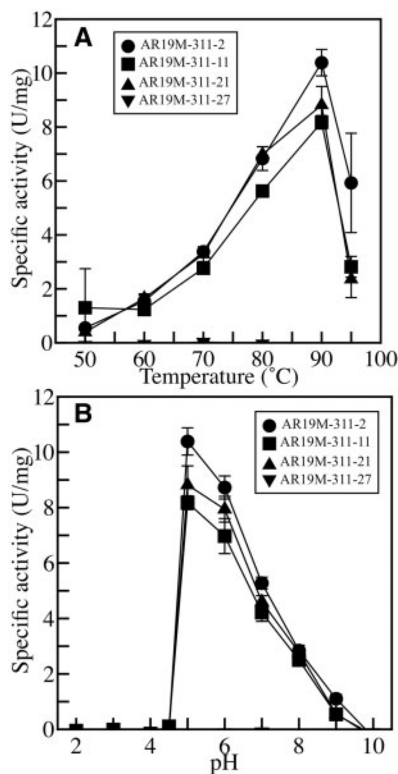


**Figure 3.** Optimal β-xylosidase activity reaction temperature (A) and pH (B) for the proteins originated from AR19M-311. β-Xylosidase activity was assayed at several temperatures in GTA buffer (pH 5.0) in (A) and in buffers adjusted to different pH values at 90 °C in (B). Activity was assessed by measuring the release of *p*-nitrophenol at 400 nm.

## 4. Discussion

Here, we carried out a comprehensive survey of thermostable β-xylosidases using the metagenomic DNA of a microbial community isolated from the soil of a hot spring, which was 70 °C and pH 7.2 when sampled. Using two types of NGS analysers, a total of 84.2 Gbp of sequences were obtained from the metagenomic DNA. From 967,925 contigs calculated from the sequences, 168 ORFs were identified, most of which shared homology of less than 80% with known β-xylosidases (Fig. 1). Of these 168 ORFs, 120 were successfully cloned from the metagenomic DNA. Including the SNPs that were identified by PCR-based cloning of the 120 ORFs, a total of 269 individual clones were obtained, and 82 proteins produced from these clones showed β-xylosidase activity at 50 °C, demonstrating the

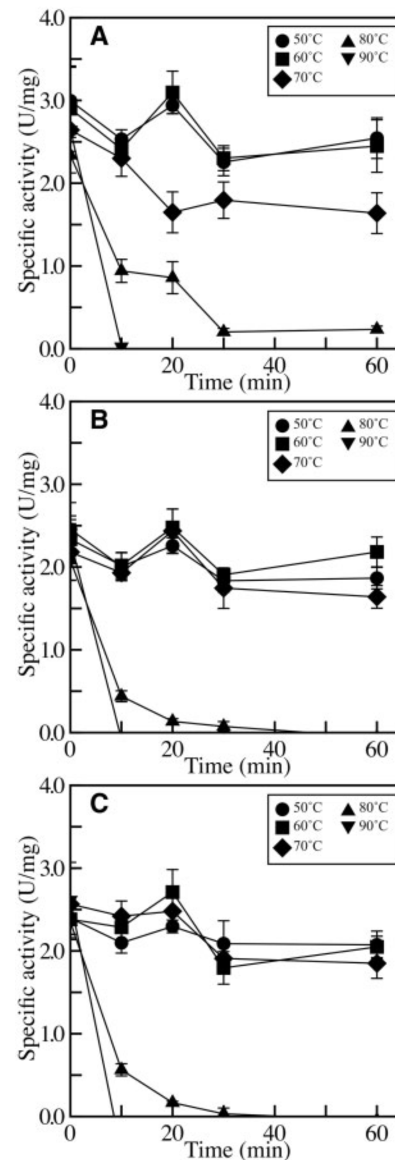


**Figure 4.** Thermal stability of AR19M-311-2 (A), AR19M-311-11 (B), and AR19M-311-21 (C). The enzymes were preincubated at 70°C for 10 min to precipitate host-derived proteins, subsequently, the enzymes were incubated at 50, 60, 70, 80, or 90 °C for 0 to 60 min and then subjected to the β-xylosidase assay using *p*NPX as the substrate. Residual activity was assessed by measuring the release of *p*-nitrophenol at 400 nm.

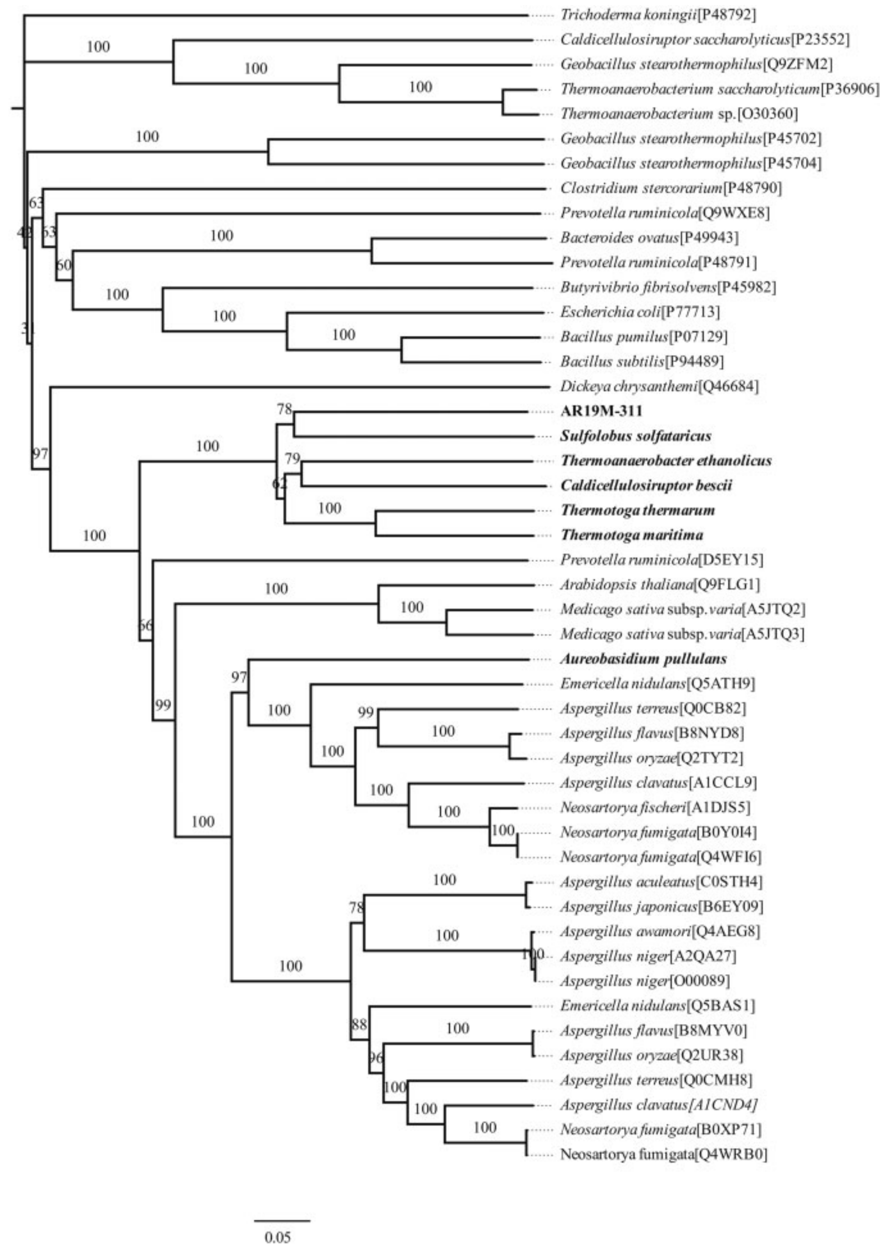


**Figure 5.** Phylogenetic tree of the proteins originated from AR19M-311 and representatives of known β-xylosidases.

usefulness of metagenomic approaches for identifying novel β-xylosidases. Six genes, including three genes originated from AR19M-311, and genes originated from AR19M-319, AR19M-329, and AR19M-334, encoded enzymes showing thermal stability up to 90 °C (Fig. 2). This is a large-scaled survey for enzymatic activity of thermostable cellulases isolated from metagenomic DNA resources, which is, to our knowledge, the first report of thermostable β-xylosidases isolated by a metagenomic approach.

This study provided novel sequences of thermostable β-xylosidases for further consideration of structure-activity relationships. AR19M-311-2, AR19M-311-11, and AR19M311-21 showed the highest activity at 90 °C and were stable at temperatures up to 70 °C for at least 1 h. As the β-xylosidase from *T. ethanolicus* showed the highest optimal reaction temperature (93 °C),[7] it was concluded that the three proteins are highly thermostable β-xylosidases. Despite similarities in their thermal properties, the amino acid sequences shared 54% identity with *T. ethanolicus* β-xylosidase. The β-xylosidases from *S. solfataricus* and *T. maritima*, which have an optimal reaction temperature of 90 °C,[4,8] are also considered highly thermostable β-xylosidases. However, the sequence identity between these β-xylosidases and the three proteins we describe was <60%. The sequence differences between these highly thermostable β-xylosidases helped us identify a number of consensus and essential amino acids in the sequences of the highly thermostable β-xylosidases, as discussed below.

Analysis of the phylogenetic tree indicated that the prokaryotic thermostable β-xylosidase cluster is more closely related to eukaryotic β-xylosidases (which include those of plants such as *Arabidopsis thaliana* and *Medicago sativa* and those of the fungi *A. pullulans*[9] and *Aspergillus* sp.) than to the prokaryotic β-xylosidases, which

```
                                                                         57
AR19M-311       MEERWLQRRVEELLSKMTLEEKIAQLGSIPSGKLVENGKFSREKAKELLKNGIGQITRVAGYAEREPEESIELINEIQRF   80
AR19M-311-2     MEERWLQRRVEELLSKMTLEEKIAQLGSIPSGKLVENGKFSREKAKELLKNGIGQITRVAGYAEREPEESIELINEIQRF   80
AR19M-311-11    MEERWLQRRVEELLSKMTLEEKIAQLGSIPSGKLVENGKFSREKAKELLKNGIGQITRVAGYAEREPEESIELINEIQRF   80
AR19M-311-21    MEERWLQRRVEELLSKMTLEEKIAQLGSIPSGKLVENGKFSREKAKELLKNGIGQITRVAGYAEREPEESIELINEIQRF   80
AR19M-311-27    MEERWLQRRVEELLSKMTLEEKIAQLGSIPSGKLVENGKFSREKAKELLKNGIGQIIRVAGYAEREPEESIELINEIQRF   80

AR19M-311       LKEETRLGIPAIIHEECLSGVMTKGATTFPQAIGMASTFEPDDIQRMTSIIRKEMKAFGVHQGLSPVLDIPRDPRWGRTE  160
AR19M-311-2     LKEETRLGIPAIIHEECLSGVMTKGATTFPQAIGMASTFEPDDIQRMTSIIRKEMKAFGVHQGLSPVLDIPRDPRWGRTE  160
AR19M-311-11    LKEETRLGIPAIIHEECLSGVMTKGATTFPQAIGMASTFEPDDIQRMTSIIRKEMKAFGVHQGLSPVLDIPRDPRWGRTE  160
AR19M-311-21    LKEETRLGIPAIIHEECLSGVMTKGATTFPQAIGMASTFEPDDIQRMTSIIRKEMKAFGVHQGLSPVLDIPRDPRWGRTE  160
AR19M-311-27    LKEETRLGIPAIIHEECLSGVMTKGATTFPQAIGMASTFEPDDIQRMTSIIRKEMKAFGVHQGLSPVLDIPRDPRWGRTE  160

                                          187    194         206
AR19M-311       ETFGEDPYLVSKMAESYIKGLQGEDWREGIIATVKHFTAYGISEGARNLGPARVSERELREVFLFPFEVAIRKANAGSVM  240
AR19M-311-2     ETFGEDPYLVSKMAESYIKGLQGEDWKEGIIATAKHFTAYGISEGGRNLGPARVSERELREVFLFPFEVAIRKANAGSVM  240
AR19M-311-11    ETFGEDPYLVSKMAESYIKGLQGEDWKEGIIATAKHFTAYGISEGGRNLGPARVSERELREVFLFPFEVAIRKANAGSVM  240
AR19M-311-21    ETFGEDPYLVSKMAESYIKGLQGEDWKEGIIATAKHFTAYGISEGGRNLGPARVSERELREVFLFPFEVAIRKANAGSVM  240
AR19M-311-27    ETFGEDPYLVSKMAESYIKGLQGEDWKEGIIATAKHFTAYGISEGGRNLGPARVSERELREVFLFPFEVAIRKANAGSVM  240

                                                                                 311
AR19M-311       NAYHEIDGVPCASSKFLLTKILREEWGFKGFVVSDYSAIEMLHTFHKVAKDLKTAAIKALEAGIEIELPETKCYGEPLLS  320
AR19M-311-2     NAYHEIDGVPCASSKFLLTKILREEWGFKGFVVSDYSAIEMLHTFHKVAKDLKTAAIKALEAGIEIELPEIKCYGEPLLS  320
AR19M-311-11    NAYHEIDGVPCASSKFLLTKILREEWGFKGFVVSDYSAIEMLHTFHKVAKDLKTAAIKALEAGIEIELPEIKCYGEPLLS  320
AR19M-311-21    NAYHEIDGVPCASSKFLLTKILREEWGFKGFVVSDYSAIEMLHTFHKVAKDLKTAAIKALEAGIEIELPEIKCYGEPLLS  320
AR19M-311-27    NAYHEIDGVPCASSKFLLTKILREEWGFKGFVVSDYSAIEMLHTFHKVAKDLKTAAIKALEAGIEIELPETKCYGEPLLS  320

                                                  355
AR19M-311       AVKEGKVSVSVIDTAVARVLRAKILLGLLDDIIYVDPSKIRAVLDNPEHRAFARELARKSIVLLKNDGILPISKGVKTIA  400
AR19M-311-2     AVKEGKVSVSVIDTAVARVLRAKILLGLLDDIIYADPSKIRAVLDNPEHRAFARELARKSIVLLKNDGILPISKGVKTIA  400
AR19M-311-11    AVKEGKVSVSVIDTAVARVLRAKILLGLLDDIIYVDPSKIRAVLDNPEHRAFARELARKSIVLLKNDGILPISKGVKTIA  400
AR19M-311-21    AVKEGKVSVSVIDTAVARVLRAKILLGLLDDIIYVDPSKIRAVLDNPEHRAFARELARKSIVLLKNDGILPISKGVKTIA  400
AR19M-311-27    AVKEGKVSVSVIDTAVARVLRAKILLGLLDDIIYVDPSKIRAVLDNPEHRAFARELARKSIVLLKNDGILPISKGVKTIA  400

                                                                  451
AR19M-311       VIGPSADSTKNLHGDYSYTSHIAGVADGVRTVTVLEGIKNKVSSGTTVLYAKGCELSDESREGFKEALDIASRSDVIIAV  480
AR19M-311-2     VIGPSADSTKNLHGDYSYTSHIAGVADGVRTVTVLEGIKNKVSSGTTVLYAKGCELSDESREGFKEALDIASRSDVIIAV  480
AR19M-311-11    VIGPSADSTKNLHGDYSYTSHIAGVADGVRTVTVLEGIKNKVSSGTTVLYAKGCELSDESREGFKEALDIASRSDVIIAV  480
AR19M-311-21    VIGPSADSTKNLHGDYSYTSHIAGVADGVRTVTVLEGIKNKVSSGTTVLYAKGCELSDESREGFKEALDIASRSDVIIAV  480
AR19M-311-27    VIGPSADSTKNLHGDYSYTSHIAGVADGVRTVTVLEGIKNKVSSGTTVLYTKGCELSDESREGFKEALDIASRSDVIIAV  480

                                                      519
AR19M-311       MGENSGLFKRGISGEGNDRIDLKLPGVQEELLKALKEVGKPIVLVLVNGRPLSIKWEKENIPAILEVWYPGEEGGNAIAD  560
AR19M-311-2     MGENSGLFKRGISGEGNDRIDLKLPGVQEELLKALKEVGKPIVLVLVNGRPLSIKWEKENIPAILEVWYPGEEGGNAIAD  560
AR19M-311-11    MGENSGLFKRGISGEGNDRIDLKLPGVQEELLKALKEVGKPIVLVLVNGRPLSIKWEKENIPAILEVWYPGEEGGNAIAD  560
AR19M-311-21    MGENSGLFKRGISGEGNDRIDLKLPGVQEELLKALKEVGKPIVLVLVNGRPLSIKWEKENIPAILEVWYPGEEGGNAIAD  560
AR19M-311-27    MGENSGLFKRGISGEGNDRIDLKLPGVQEELLKALKEVSKPIVLVLVNGRPLSIKWEKENIPAILEVWYPGEEGGNAIAD  560

                                                                                625       635
AR19M-311       VIFGDYNPGGKLPISFPKDVGQIPVYYNRKPSAFSEYLTTDTKPLFPFGHGLSYTTFEYSELKIIPENVMPGGYVDISFK  640
AR19M-311-2     VIFGDYNPGGKLPISFPKDVGQIPVYYNRKPSAFSEYLTTDTKPLFPFGHGLSYTTFEYSELKIIPENVMPGGYIDISFK  640
AR19M-311-11    VIFGDYNPGGKLPISFPKDVGQIPVYYNRKPSAFSEYLTTDTKPLFPFGHGLSYTTFEYSELKIIPENVMPGGYVDISFK  640
AR19M-311-21    VIFGDYNPGGKLPISFPKDVGQIPVYYNRKPSAFSEYLTTDTKPLFPFGHGLSYTTFEYSELKITPENVMPGGYVDISFK  640
AR19M-311-27    VIFGDYNPGGKLPISFPKDVGQIPVYYNRKPSAFSEYLTTDTKPLFPFGHGLSYTTFEYSELKITPENVMPGGYVDISFK  640

AR19M-311       VRNTGNIDGDEVVQLYIHDEWASVERPIKELKGFKRIHLKAREEKKVTFRLFTDQLAFYDEVMRFVVEAGTFEVMVGSSS  720
AR19M-311-2     VRNTGNIDGDEVVQLYIHDEWASVERPIKELKGFKRIHLKAREEKKVTFRLFTDQLAFYDEVMRFVVEAGTFEVMVGSSS  720
AR19M-311-11    VRNTGNIDGDEVVQLYIHDEWASVERPIKELKGFKRIHLKAREEKKVTFRLFTDQLAFYDEVMRFVVEAGTFEVMVGSSS  720
AR19M-311-21    VRNTGNIDGDEVVQLYIHDEWASVERPIKELKGFKRIHLKAREEKKVTFRLFTDQLAFYDEVMRFVVEAGTFEVMVGSSS  720
AR19M-311-27    VRNTGNIDGDEVVQLYIHDEWASVERPIKELKGFKRIHLKAREEKKVTFRLFTDQLAFYDEVMRFVVEAGTFEVMVGSSS  720

                                  739
AR19M-311       EDIRLTGKFEVLETKVITKDRKFASEVIIEP   751
AR19M-311-2     EDIRLTGKFEVLETKVITKDRKFASEVIIEP   751
AR19M-311-11    EDIRLTGKFEVLETKVITKDRKFASEVIIEP   751
AR19M-311-21    EDIRLTGKFEVLETKVITKDRKFASEVIIEP   751
AR19M-311-27    EDIRLTGKFEVLETKVITNDRKFASEVIIEP   751
```

**Figure 6.** Sequence alignment of AR19M-311 and its clones. The alignment was created by ClustalX version 2.0 and the residues substituted in any of the clone were highlighted by black background color and the residue numbers were indicated at the top of the residues. AR19M311 is the consensus sequence obtained from the ORF search. AR19M-311-2, AR19M-311-11, and AR19M-311-21 are the thermostable proteins, and AR19M-311-27 is the non-thermostable protein.

have not been reported as thermostable (Fig. 5). Therefore, comparative sequence analysis of these β-xylosidases could also provide information that could enhance our understanding of the mechanisms of high-temperature stability.

From an integrated consideration of both the complete comparative sequence analysis and the local effects of amino acid substitutions, we identified some of the amino acid residues essential for the thermal stability of β-xylosidases. Comparing the three thermostable proteins and the non-thermostable protein AR19M-311-27, the five residues were found as unique substitutions in AR19M-311-27 ($Thr^{57}$ to $Ile^{57}$, $Ile^{311}$ to $Thr^{311}$, $Ala^{451}$ to $Thr^{451}$, $Gly^{519}$ to $Ser^{519}$, and $Lys^{739}$ to $Asn^{739}$). In alignment of the sequences of the thermostable prokaryotic β-xylosidases and eukaryotic β-xylosidases (Supplementary Fig. S2), the $Thr^{57}$ and the $Ala^{451}$ was conserved in all of the thermostable enzymes, and the $Gly^{519}$ was conserved in five of the six thermostable enzymes. Hence, the three substitutions of $Thr^{57}$ to $Ile^{57}$, $Ala^{451}$ to $Thr^{451}$ and $Gly^{519}$ to $Ser^{519}$ were considered to affect the low thermostability of AR19M-311-27 at 90°C. The alignment also showed that insertion of two amino acids, including $Thr^{311}$ or the corresponding residues of other β-xylosidases, occurs only in the thermostable prokaryotic β-xylosidases and the moderately thermostable β-xylosidase from *Trichoderma koningii*.[16] The Thr residue is conserved in the four prokaryotic β-xylosidases and substituted with Ile in the others. Therefore, we suggest that the insertion of the two amino acids contributes to the thermal stability of β-xylosidases, which should be confirmed by future biochemical works. The contribution of the substitution of $Lys^{739}$ to $Asn^{739}$ on the thermostability is considered to be insignificant, because $Lys^{739}$ is not conserved in thermostable enzymes. AR19M-311-2 also possessed the unique substitutions of $Val^{355}$ with Ala and $Val^{635}$ with Ile, substitutions not found in AR19M-311-11 and AR19M-311-21. These substitutions apparently have only a slight influence on thermostability, as Val, Ala, and Ile are all hydrophobic residues with similar properties.

As shown by this study, PCR-based cloning of genes from metagenomic DNA is a useful method because it can lead to the identification of SNPs in consensus ORF sequences, which is advantageous for elucidating structure-activity relationships. There are two possible approaches to discover functional enzymes from metagenomic data. One is the homology search method demonstrated in this study, while the other is the activity-driven one. One of the advantages of the former method is that it could find highly homologous sequences

with distinct activities from vast sequences, especially encoding active and non-active enzymes, which may lead speculation of function of each amino acid residue on the enzyme activity, as shown in our results of active forms, AR19M-311-2, AR19M-311-11, and AR19M-311-21 and a non-active form AR19M-311-27, which could not be obtained by the latter method alone. Recently, isolation of novel β-galactosidases not predictable from gene sequences was reported,[17] in which the enzymes were identified from a cosmid library with ∼33 kbp metagenomic DNA in surrogate hosts by the activity-driven method. They emphasize the importance of developing of multi-host systems for functional screening to substantially extend the breadth of gene discovery. Therefore, either method may be used according to the purpose of research.

## Data availability

The sequences of AR19M metagenome were deposited in DDBJ Sequence Read Archive maintained by the DNA Data Bank of Japan under accession numbers of DRA005406 and DRA005407.

## Acknowledgements

We thank Fumiko Fujii for her excellent technical support in DNA sequencing using the GS FLX Titanium sequencer and Kumiko Suzuki, Miyuki Inde, Masumi Nagaki, Akiko Kisen, Takahiro Matsuyama, and Satomi Saito for their technical support in DNA cloning and protein expression analyses.

## Supplementary data

Supplementary data are available at *DNARES* online.

## Funding

This research was supported by Honda Research Institute Japan and the Kazusa DNA Research Foundation.

## Conflict of interest

None declared.

## References

1. van Dyk, J.S. and Pletschke, B.I. 2011, A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy, *Biotechnol. Adv.*, **30**, 1458–80.
2. Schomburg, I., Chang, A., Placzek, S., et al. 2013, BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA, *Nucleic Acids Res.*, **41**, D764–72.
3. Haki, G. 2003, Developments in industrially important thermostable enzymes: a review, *Bioresour. Technol.*, **89**, 17–34.
4. Xue, Y. and Shao, W. 2004, Expression and characterization of a thermostable beta-xylosidase from the hyperthermophile, *Thermotoga maritima*, *Biotechnol. Lett.*, **26**, 1511–5.
5. Shi, H., Li, X., Gu, H., et al. 2013, Biochemical properties of a novel thermostable and highly xylose-tolerant β-xylosidase/α-arabinosidase from *Thermotoga thermarum*, *Biotechnol. Biofuels*, **6**, 27.
6. Su, X., Han, Y., Dodd, D., et al. 2013, Reconstitution of a thermostable xylan-degrading enzyme mixture from the bacterium *Caldicellulosiruptor bescii*, *Appl. Environ. Microbiol.*, **79**, 1481–90.
7. Shao, W. and Wiegel, J. 1992, Purification and characterization of a thermostable β-xylosidase from *Thermoanaerobacter ethanolicus*, *J. Bacteriol.*, **174**, 5848–53.
8. Morana, A., Paris, O., Maurelli, L., Rossi, M., and Cannio, R. 2007, Gene cloning and expression in *Escherichia coli* of a bi-functional beta-D-xylosidase/alpha-L-arabinosidase from *Sulfolobus solfataricus* involved in xylan degradation, *Extremophiles*, **11**, 123–32.
9. Dobberstein, J. and Emeis, C.C. 1991, Purification and characterization of β-xylosidase from *Aureobasidium pullulans*, *Appl. Microbiol. Biotechnol.*, **35**, 210–5.
10. Sebastian, R., Kim, J.Y., Kim, T.H., and Lee, K.T. 2013, Metagenomics: a promissing approach to assess enzymes biocatalyst for biofuel production, *Asian J. Biotechnol.*, **5**, 33–50.
11. Allgaier, M., Reddy, A., Park, J.I., et al. 2010, Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community, *PLoS One*, **5**, e8812.
12. Bao, L., Huang, Q., Chang, L., Sun, Q., Zhou, J., and Lu, H. 2012, Cloning and characterization of two β-glucosidase/xylosidase enzymes from yak rumen metagenome, *Appl. Biochem. Biotechnol.*, **166**, 72–86.
13. Noguchi, H., Taniguchi, T., and Itoh, T. 2008, MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes, *DNA Res.*, **15**, 387–96.
14. Larkin, M.A., Blackshields, G., Brown, N.P., et al. 2007, Clustal W and Clustal X version 2.0, *Bioinformatics*, **23**, 2947–8.
15. Campos, E., Alvarez, M.J.N., di Lorenzo, G.S., et al. 2014, Purification and characterization of a GH43 β-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. *Microbiol. Res.*, **169**, 213–20.
16. Li, Y.K., Yao, H.J., and Cho, Y.T. 2000, Effective induction, purification and characterization of *Trichoderma koningii* G-39 beta-xylosidase with high transferase activity. *Biotechnol. Appl. Biochem.*, **31**, 119–25.
17. Cheng, J., Romantsov, T., Engel, K., et al. 2017, Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS One*, **12**, e0172545.