**NOVEL β-GLUCOSIDASE *bgl6111* FROM BAGASSE FOUND BY METAGENOMIC LIBRARY**

Fitra Adi Prayogo1, Benjarat Bunterngsook2, Pattanop Kanokratana2, Hermin Pancasakti Kusumaningrum3, Dyah Wulandari4,and Anto Budiharjo3,4\*

**Email:** fitra.adiii@gmail.com, benjarat.bun@biotec.or.th, [pattanopk@biotec.or.th](mailto:pattanopk@biotec.or.th), herminpk@live.undip.ac.id, [wulandaridyah1207@gmail.com](mailto:wulandaridyah1207@gmail.com)

**Corresponding author:** [anto.budiharjo@live.undip.ac.id](mailto:anto.budiharjo@live.undip.ac.id)

1Biomedical Sciences Study Program, Karya Husada University, Jl. R. Kompol Soekanto No.46,

Semarang 50276, Indonesia

2 Enzyme Technology Laboratory, National Centre for Genetic Engineering and Biotechnology

(BIOTEC), National Science and Technology Development Agency, 113 Thailand Science Park,

Phahonyothin Road, Khlong Luang, Pathumthani 12120, Thailand

3Biotechnology Study Program, Diponegoro University, Jl. Prof Soedharto SH – Semarang 50275,

Indonesia

4Molecular and Applied Microbiology Laboratory, Central Laboratory of Research and Service –

Diponegoro University, Jl. Prof. Sudharto SH, Semarang 50275, Indonesia

**Author Contribution**

PK and AB designed, analyzed the research data, and screened the manuscript; FAP and BB conducted and wrote the manuscript. HPK and DW analyzed the data and screened the manuscript. All authors read and approved the manuscript

**NOVEL β-GLUCOSIDASE *bgl6111* FROM BAGASSE FOUND BY METAGENOMIC LIBRARY**

**Abstract**

β-Glucosidase (BGL) is a rate-limiting and critical factor for the hydrolysis of cellulose. These enzymes have several functions in industrial processes. However, study about the exploration of natural BGL enzymes is scarce. Metagenomics is a robust tool for bioprospecting in search of novel enzymes from the entire community genomic DNA present in nature. In this study, we report the cloning, heterologous expression of *bgl6111* genein *Pichia pastoris* KM71 and biochemically characterized the enzyme result. We successfully identified the sequence from bgl6111 with2,520 bp, a GC content of 67%, and encoded 839 amino acids with molecular weight of 89.4 kDa. The amino acid sequence of the *bgl6111* gene showed a similarity (67.61%) to BGL from uncultured bacterium (ABB51613.1). The BGL product had the highest activity on Day 3 = 1.210 U/Ml, categorized as low production. The enzymatic activity could enhance up to 539.8% of 7.742 U/Ml by using the ultrafiltration method. Our findings provide insightful information that *bgl6111* obtained from bagasse metagenome could be an alternative candidate for industrial applications in the future.

**Keywords:** Metagenomic library; Bagasse; β-Glucosidase; Cloning.

# Introduction

β-Glucosidases (BGL) (EC 3.2.1.21) are enzyme that belongs to the cellulase enzyme complex group and are classified among the glycoside hydrolases (GH) (Ahmed et al. 2017; (Rouyi et al. 2014). They hydrolyze glycosidic bonds to release nonreducing terminal glucosyl residues (Rouyi et al. 2014). BGL enzymes are widespread in plants, fungi, bacteria, and animals (Singhania et al. 2017). They are rate-limiting and critical factors for the hydrolysis of cellulose. These enzymes have several functions in industrial processes, including composting (Zang et al. 2018), naringin breakdown in grapefruit juice (Prakash et al. 2002), and ethanol production (Tang et al. 2013). BGL enzymes are also used to produce aromatic compounds for the flour industry, wine production, hydrolysis of anthocyanin products, increase organoleptic levels of fruits and juices (Singh et al. 2016), and used as an additive in animal feed to increase the digestibility of food (Singhania et al. 2017). However, the exploration of natural BGL enzymes is still relatively low. In comparison, nature provides various types and characteristics of different BGL enzymes that can develop in industrial processes.

Cloning and expression are preferred in enzyme exploration because they are simple, cheap, and fast (Zhao et al. 2013). Researchers have explored various *bgl* gene sources encoding BGL enzymes produced by microorganisms. Chamoli et al. (2016) successfully isolated the *bgl* gene from *Bacillus subtilis* and expressed it in *Escherichia coli,* producing an enzyme with 54.04 U/mg activity. Yang et al. (2015) expressed the *bgl* gene from *Thermoanaerobacterium aotearoense* in *E. coli* with an enzyme activity of 740.5 U/mg. In addition to isolating the *bgl* gene from isolated microorganisms, it can be obtained directly from the environmental DNA (Ramírez-escudero et al. 2016; Matsuzawa et al. 2017). Genes in the natural environment extracted directly via a metagenomic approach are potential sources of unexplored enzymes. This potential allows researchers to explore genetic resources that have not been revealed (Prayogo et al. 2020).

The needs for this enzyme are highly demand in the current years due to the multifunctional purpose of β-Glucosidases (BGL) as nutraceuticals and pharmaceuticals because of their recognition ability, signaling processes, also antibiotic properties (Bhatia et al., 2002). The novel BGL from natural sources is abundant in the environment. The potential of the *bgl* gene can be explored based on environmental characteristics. Sugarcane bagasse pile is one of the potential environments for exploring the *bgl* gene. It represents a unique ecological characteristic with a high lignocellulose-rich environment. Microbial communities in this environment provide valuable functional gene resources for discovering lignocellulolytic enzymes (Mhuantong et al. 2015). Exploring the *bgl* gene from the sugarcane bagasse environment may lead to discovering a novel BGL enzyme as an alternative candidate enzyme for industrial purposes.

The aim of this research was attempted to clone the sequence from several lignocellulolytic enzymes which previously constructed and identified through activity-based screening from the bagasse metagenomic library according to (Kanokratana et al. 2015) and Mhuantong et al., 2015 in the pPICZαA plasmid using *E. coli* DH5α as a propagation host and expressed in *Pichia pastoris* KM71. Our result from sequence analysis revealed unique and conserved biomass-degrading enzymes in this metagenomic library denoted as bgl6111, a β-Glucosidase (BGL) novel enzyme similar to the BGL from uncultured bacterium (ABB51613.1). We also provide the detailed structural model of *bgl6111* which the recombinant protein product is closely related to GH3 and analysis of the amino acid showed the molecular weight of this protein of 89.4 kDa. Furthermore, we present the simple ultrafiltration method to enhance the enzyme activity of the *bgl* gene 539.8% higher if compared with non-treatment.

# Materials and Methods

* 1. **Microbial strain, plasmid, and chemicals**

The pPICZαA purchased from Invitrogen was used as the DNA vector. The *bgl6111* gene from the bagasse metagenomic library collected from the bagasse pile at Phu Khieo Bio-Energy Chaiyaphum province, Thailand, was used as a DNA insert (Mhuantong et al. 2015). *E. coli* DH5α was used as the propagation host, and *P. pastoris* KM71 was used as the expression host.

The chemicals used in this study are as follows: GeneJET Gel Extraction Kit (Thermo Scientific) for DNA purification; GeneJET Plasmid Miniprep Kit (Thermo Scientific) for plasmid isolation; GoTaq Green Master Mix (Promega) for colony PCR; Luria Bertani (LB) medium (0.5% w/v yeast extract, 1% w/v peptone, 1% w/v NaCl, and 100 μg/mL ampicillin) for *E. coli* DH5α growth; yeast extract peptone dextrose (YPD) (1% w/v yeast extract, 2% w/v peptone, and 2% w/v dextrose); buffered minimum glycerol (BMGY; 1% v/v glycerol, 2% w/v peptone, 1% w/v yeast extract, 1.34% w/v YNB, 4 × 10 5% w/v biotin, and 100mM potassium phosphate buffer with pH of 6.0) for *P. pastoris* KM71 growth media. Buffered minimum methanol (BMMY; 0.5% v/v methanol, 2% w/v peptone, 1% w/v yeast extract, 1.34% w/v YNB, 4 × 10 −5% w/v biotin, and 100 mM potassium phosphate buffer pH 6.0) was used for the induction of gene expression (Zhang et al. 2019). The activity of the BGL enzyme was assayed by the p-Nitrophenyl-β-D-glucopyranoside (pNPG) method (Gao et al. 2016).

* 1. **Construction of cloned *bgl* expression vector**

A full-length sequence of the *bgl* gene (contig no. 6111) was successfully identified from the metagenomic library constructed from the bagasse pile. The gene was amplified using gene-specific primers BGL6111/EcoRI/F (5′ GAATTCATGGCATGCGTGCTCGCAGCCTTT 3′) and BGL6111/XbaI/R (5′ TCTAGATCATCCCGTGCACGGAAGGGTGCC 3′). A total of 50 µL of 50 ng DNA template, 1× Phusion GC buffer, 200 µM each of dNTPs, 0.5 µM each of the primers, and one unit of Phusion DNA polymerase were used for PCR amplification. The amplification was conducted at 95 °C for 5 min and involved 25 cycles of denaturing at 95 °C for 30 s, annealing at 55 °C for 30 s, elongation at 72 °C for 3 min (extension efficiency was 30 s/1 kb), and a final extension at 72 °C for 10 min.

The GeneJET Gel Extraction Kit was used to purify the PCR product. EcoRI and XbaI digested the blunt-ended purified PCR product to make a sticky end. The digestion reaction involved 2× concentration of Tango buffer, 1 μg of PCR product, and five units of *Eco*RI and *Xba*I incubated at 37 °C for two h. The digested fragment was then ligated to the pPICZαa plasmid. The ligation mixture contained 50 ng of insert DNA, 50 ng of plasmid, 1× concentration of T4 ligase buffer, and five units of T4 ligase enzyme incubated at 22 °C for 16–18 h. The ligation mixture was then transformed into *E. coli* DH5α by using the heat shock method. The transformants harbouring the corrected recombinant plasmid were selected from LB agar containing 25 µg/mL of Zeocin. The transformants were screened by colony PCR using 5′-AOX1forwardand 3′-AOX1 reverse primers. The gene sequence in the plasmid was subsequently confirmed by conventional sequencing (Macrogen).

* 1. **Screening and expression of *P. pastoris* KM71 transformants**

The recombinant plasmid pPICZαA-bgl6111 present in *E. coli* DH5α were segregated using the GeneJET Plasmid Miniprep Kit. Subsequently, the fragments were linearized by *Pme*I to facilitate the integration of the pPICZαA–*bgl6111* and the AOX1 locus of *P. pastoris* KM71. The linearized fragments were transformed into *P. pastoris* KM71 via electroporation and then cultured into YPD containing 100 μg/mL of Zeocin at 30 °C for three days. The positive recombinant cells were confirmed using colony PCR with 5′-AOX1 forward and 3′-AOX1 reverse primers and OUT-AOX\_F and REV-α-965-R. The PCR reaction was carried out with 2× GoTaq Green buffer (Promega) in a total volume of 25 μL. The amplification was conducted at 95 °C for 5 min, followed by 25 cycles at 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 3 min. A final extension was conducted at 72 °C for 10 min.

Positive transformant colonies were randomly selected from the master plate and then cultured on 50 mL of BMGY incubated at 30 °C until the OD600 reached 2–6. The culture was centrifuged at a rate of 3,000 ×g for 5 min. The culture pellets were grown in 10 mL of BMMY media incubated at 30 °C for three days. Induction was maintained for three days through the daily addition of methanol to maintain the methanol concentration at 0.5%. The enzyme activity was measured every day for three days (Day 1 = D1; Day 2 = D2; Day 3 = D3).

The ultrafiltration technique could enhance enzyme concentration [(Zuhair et al. 2018)]. The BMGY culture was scaled up to 200 mL and incubated at 30 °C until the OD600 reached 2–6. The culture was centrifuged at 3,000 ×g for 5 min and cultivated in 40 mL of BMMY at 30 °C for three days. After that, the culture was centrifuged at 3,000 ×g for 10 min. The resulting supernatant was concentrated using ultrafiltration (30 kDa cutoff centrifugal filter, Amicon).

* 1. **Enzyme assay and SDS-PAGE analysis**

The measurement of BGL activity involved using one mM p-nitrophenyl-β-D-glucopyranoside (pNPG) as a substrate. The pNPG was dissolved in 50 mM citrate buffer with a pH of 5.0. The blank and substrate solution containing 80 μL of 1 mM pNPG in each tube was incubated at 50 °C for 5 min. After that, 20 μL of BGL enzyme was added to the substrate solution and then incubated at 50 °C for 10 min. The reaction was stopped by adding 25 μL of 1 M Na2CO3 to the blank and substrate solutions. Exactly 20 μL of BGL enzyme was then added to the blank solution. The solutions were transferred into a 96-well plate to calculate the concentration using spectrophotometry at a wavelength of 405 nm [(Gao et al. 2016)]. The unit of BGL activity was defined as one nmol of p-nitrophenol, which is released per milliliter of enzyme per minute under standard test conditions.

SDS-PAGE was conducted to prove and determine the size of the protein product. Exactly 10 µL of the expression sample was mixed with 1× loading dye in a microtube. The mixture was then heated at 100 °C for 5 min and centrifuged at a speed of 2,000 rpm for 5 s at four °C. Approximately 20 μL of samples and ten μL of markers were inserted into the well. Markers (14.4–120 kDa) were used to calculate the molecular weight of the protein produced. SDS-PAGE was run for 80 min with an electric current of 30 Ω. Coomassie blue staining was used to visualize the resulting protein product band.

* 1. **Bioinformatic analysis**

The BLAST was used for molecular analysis of the *bgl6111* gene [(Zhang et al. 2017)]. ExPASy Translate Tool (https://www.expasy.org/) was applied to translate the *bgl6111* DNA sequence into an amino acid sequence and predict its molecular weight. The prediction of secondary structures was carried out with the Chou & Fasman method [(Kumar 2013)]. The 3D structural model of *bgl6111* protein was generated by Phyre 2 [(Kelley et al. 2015)]. Multiple sequence alignment was analyzed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [(Lu et al. 2013; Chamoli et al. 2016)]. The phylogenetic tree was built using MEGA 7 software for analysis based on the neighbor-joining method under a bootstrap value of 1,000 [(Dodda et al. 2018)]. The gene sequence is available in the NCBI database with accession number MW221260.

1. **Results and Discussion**
   1. **Cloning of *bgl6111* gene on pPICZαA**

Cloning of the *bgl6111* gene was initiated by amplifying it to multiply and add the site of the *Eco*RI restriction enzyme at the start (5'-end) and *Xba*I at the end (3'-end). The amplification process used specific pairs of *bgl6111*/*Eco*RI\_F and *bgl6111*/*Xba*I\_R to produce a DNA fragment length of 2,535 bp. **Figure 1** shows the band proving the presence of the *bgl6111* gene.

**Figure 1.** Gel PCR electrophoresis results of the *bgl6111* gene product show a band size of 2,535 bp. {M = 100bp plus DNA ladder (Thermo Scientific); S = sample *bgl6111* gene}.

Digestion was carried out on the DNA vector (pPICZαA) and DNA insert (*bgl6111* gene) to make them compatible. After digestion, the DNA insert produced a 2,529 bp size, while the vector DNA fragments produced a 3,530 bp size. Both insert fragments and vectors were then ligated to produce the size of 6,059 bp. Recombinant plasmids (6,059 bp) were inserted into *E. coli* DH5α using the heat shock method. The transformation result showed that colonies were growing in the 25 µg/mL Zeocin media. Subsequently, the growing colonies were verified by colony PCR by using the *bgl6111*/*Eco*RI\_F and *bgl6111*/*Xba*I\_R primers. The verification result indicated that the *bgl6111* gene was successfully transformed into *E. coli* DH5α (data not shown). The recombinant plasmid was then sequenced to ascertain the nucleotide sequence of the *bgl6111* gene*.*

* 1. **Expression of recombinant plasmid pPICZαA–*bgl6111***

The GeneJET Plasmid Miniprep Kit (Thermo Scientific) isolated the recombinant plasmid pPICZαA–*bgl6111* from *E. coli* DH5α. PmeI digested purified plasmid recombinants to linearize the fragments. The linear form of recombinant plasmids was then transformed into *P. pastoris* KM71 through the electroporation method. The result of the transformation revealed that colonies were growing in the 100 μg/mL Zeocin media. The 5′-AOX1 forward and 3′-AOX1 reverse were used to confirm the presence of expression cassettes (3,056 bp).

Moreover, OUT-AOX\_F and REV-α-965-R were used to confirm the integration of the AOX1 promoter site to some α-factor secretion signal sites (~1,300 bp). The results showed that the *bgl6111* gene was successfully integrated into *the P. pastoris* genome. Such integration indicates a successful transformation process.

The selected colonies were then expressed through a culture process to produce recombinant proteins by induction in a fed-batch fermentation mode [(Looser et al. 2015)]. Methanol content in BMMY induced the gene expression process via the AOX1 promoter. SDS-PAGE checked the presence of proteins, and the specific activity was measured with the pNPG substrate. The enzyme-specific activity checking was carried out every day for three days, where D1 was the first, D2 was the second, and D3 was the third.

* 1. **Proving the presence of protein products**

The product enzyme was incubated with the pNPG substrate and measured at a wavelength of 405 nm. This enzyme was expressed as a soluble protein with the activity of 0.775 U/mL at D1, 1.173 U/mL at D2, and 1.210 U/mL at D3. The ultrafiltration technique was used to increase enzyme activity. The resulting recombinant protein activity enhanced using ultrafiltration was 7.742 U/mL (**Figure 2**). The yield enhanced using ultrafiltration increased up to 539.8%.

**Figure 2.** The Graph of β-glucosidase recombinant activity with a standard deviation line (D1 = first day of fermentation; D2 = second day of fermentation; D3 = third day of fermentation; H1 = optimized sample).

The SDS-PAGE method was used to prove the presence of recombinant BGL. According to the results of SDS-PAGE (**Figure 3**), the H1 well showed the presence of a protein with a size of 89.4 kDa; this outcome matched the previous prediction. Meanwhile, the control showed no band in the gels. This result proved that the *bgl6111* gene derived from the sugarcane bagasse metagenome sample could be expressed in *P. pastoris* KM71.

**Figure 3.** The SDS-Page results of recombinant product β-glucosidase protein. The protein band shows a size of 89.4 kDa. M = marker 14.4-120.0 kDa, LabAid; H1= optimized sample; C= control, empty plasmid).

* 1. **Sequence analysis of *bgl6111* gene**

The *bgl6111* gene has a fragment length of 2,520 bp, encodes 839 amino acid proteins, and contains 67% GC. The BLAST analysis showed that the *bgl6111* gene had a 67.61% similarity to BGL from uncultured bacterium (ABB51613.1). The recombinant protein product from the *bgl6111* gene was predicted to have a molecular weight of 89.4 kDa. Theoretical isoelectric points and instability indices were 5.41 and 40.44, respectively. An instability index exceeding 40 is classified as an unstable protein (Shrestha et al. 2017).

Phyre2 analysis predicted the tertiary structure of *bgl6111* proteins with a confidence level of 100% and identity of 49% based on the PDB template c3f93D, a glycosyl hydrolase family 3 (GH3) from *Pseudoalteromonas* sp. BB1. **Figure 4A** shows the ligand site predicted by 3DligandSite is highlighted in green; **Figure 4B** shows catalytic sites at Asp 320 as a nucleophilic site and Glu 520 as a proton donor (Wass et al. 2010; Kelley et al. 2015).

**Figure 4.** Structural model of bgl6111 protein generated by Phyre2. **(4A)** Ligand site is highlighted in green, and **(4B)** the catalytic site is highlighted in red.

Based on Pfam’s analysis, three conserved domains were detected in the sequence of the *bgl6111*: the glycosyl hydrolase family 3 (GH3), N-terminal domain; GH3, C-terminal domain; and galactose-binding domain-like. These results indicated that the recombinant protein product is closely related to GH3. The secondary structure prediction analysis using the Chou & Fasman method revealed that the main structure of the recombinant BGL in this study was α-helix, with a percentage of 66.6% [(Kumar 2013)].

The phylogenetic tree analysis revealed that the BGL sequence (*bgl6111*) in this study was closely related to the amino acid sequence of BGL from prokaryotic organisms. Two large clads classified ten amino acid sequences used for the phylogenetic analysis (**Figure 5**). The first clad consisted of amino acids belonging to *Steroidobacter*, including glucan 1,4-β-glucosidase (GFE91065.1), GH3 C-terminal domain-containing protein (WP\_161830347.1) from *Steroidobacter* sp. SA29-B, glucan 1,4-β-glucosidase (GFE78006.1), and GH3 C-terminal domain-containing protein (WP\_161809948.1) from *Steroidobacter* sp. YU21-B.

**Figure 5.** Phylogenetic analysis based on amino acids from β-glucosidase (*bgl6111*) using the neighbor-joining (NJ) method with a bootstrap value of 1,000. *bgl6111*= sample used in this study, ABB51613.1= β-glucosidase from uncultured bacterium, GFE91065.1= glucan 1,4-beta-glucosidase and WP\_161830347.1= GH3 C-terminal domain-containing protein from Steroidobacter sp. SA29-B, GFE78006.1= glucan 1,4-beta-glucosidase and WP\_161809948.1= GH3 C-terminal domain-containing protein from *Steroidobacter* sp. YU21-B, WP\_129641816.1= Exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter agariperforans*, WP\_129781004.1= Exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter* sp. JW-3, WP\_116808789.1= Exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter cummioxidans*, and WP\_129794185.1= Exo 1,3/1,4-beta-D-glucan glucohydrolase from Sphingosinicella sp. CPCC 101087.

The first clad also included exo 1,3/1,4-beta-D-glucan glucohydrolase from other *Steroidobacter*, such as *S. agariperforans* (WP\_129641816.1), *Steroidobacter* sp. JW-3 (WP\_129781004.1), and *S. cummioxidans* (WP\_116808789.1), as well as two types of organisms that have yet to be identified, namely, an uncultured bacterium (ABB51613.1) and metagenomic samples used in this study (*bgl6111*). The second clad comprised Exo 1,3/1,4-beta-D-glucan glucohydrolase from *Sphingosinicella* sp. CPCC 101087 (WP\_129794185.1).

In this study, we used sequence information from the bagasse metagenomic library. Mhuantong et al., (2015) previously reported several genes that contribute to the degradation of lignocellulose. The *bgl6111* gene was selected in this work and performed as a heterologous expression sample. The cloning and expression herein aimed to explore the discovery of new biomass-degrading enzymes from the bagasse pile environment.

Gene cloning was successfully carried out by producing positive transformants, i.e., *E. coli* DH5α. *E. coli* DH5α was used as the host for propagation. This strain has been used in various studies as it can multiply pDNA well (Borja et al. 2012; Trivedi et al. 2014). The identified *bgl6111* gene has a length of 2,520 bp. Several researchers found the gene encoding BGL from the environment with varying sizes. Pozo et al. (2012) found a gene encoding BGL (SRF2g14) with a size of 2,361 bp from bovine rumen microorganisms. Gomes-pepe et al. (2016) also found a gene encoding BGL (Bgl10) with a size of ~2,300 bp from soil microorganisms. The differences in gene sizes indicate the diversity of genes in the environment. Genes come from various microorganisms but encode the same orthology function (Pearson 2013).

The full-length *bgl6111* gene was obtained via PCR using specific primers designed to modify compatible cloning sites with the vectors (pPICZαA) (Hoseini and Sauer 2015). The recombinant enzyme was successfully produced and secreted. Chen et al. (2011) successfully cloned the gene that encodes BGL in the pPICZαA plasmid. Wang et al. (2017) also succeeded in cloning the *bgl2* gene in the pPICZαA plasmid. Both researchers took advantage of the plasmid features called the α-factor. It functions as a signal in the process of secretion in *P. pastoris*.

The recombinant plasmid pPICZαA–*bgl6111* was successfully transformed into *P. pastoris* KM71 in the current work. Genome integration of the insert occurred at the AOX1 sites by homologous recombination (Vogl et al. 2018). The enzyme products were verified by measuring the enzyme activity and conducting an SDS-PAGE analysis. The enzyme activities of D1, D2, D3, and H1 are known below. SDS analysis also revealed low levels of gene expression. Such a result was marked by the thinness of the band generated from SDS-PAGE (Choi and Gelutu 2018). This outcome strengthens the argument of previous research about an existing relationship between low protein activity and thin bands in SDS-PAGE results.

The process of protein expression is complicated. Various complex factors can affect the results of recombinant enzyme expression. The factors may come from genetic or cultural processes. Genetic factors include promoters, gene doses, gene sequences, and post-translational protein modification (Wang et al. 2017; Yu et al. 2017). Yu et al. (2014) stated that the factors of the culture process, such as temperature, induction duration, and culture volume, could significantly influence the yield of recombinant protein production in *P. pastoris*. However, the study did not analyze those factors. It only measured the CAI (Codon Adaptation Index), which was supposed to be one of the causes of low gene expression (Chuck et al. 2009).

According to Behura and Severson (2012), codon bias is when specific codons are used more frequently than other synonym codons. It often occurs in heterogeneous gene expression. Quax et al. (2016) stated that the codon dimension could be seen through the CAI index value. Therefore, CAI analysis was carried out in the current work to ensure the use of codon bias with the help of a web-based application from GenScript (www.genscript.com) (Farshadpour et al. 2015). The results revealed that the CAI value of the *bgl6111* gene sequence was 0.47. CAI values ​​are considered if approved from 0.8 to 1.0. A low CAI value indicates a low level of gene expression. Hence, codon bias may be concluded as one of the causes of the low activity of BGL products in the current work.

An important factor affecting low expression is using a sample from metagenomic samples. Metagenomic samples are eDNA (environment DNA) directly extracted from the environment (Lewin et al. 2017). Various types of microorganisms, including many uncultured microbes in the environment, can be identified through a metagenomic approach. Wooley and Ye (2009) explained that although metagenome samples are helpful in uncovering the diversity of microorganisms, they are difficult to clone and express into a host. Unknown samples and limited information can be obtained from these genes, which are thus difficult to study, especially gene cloning and expression.

The new *bgl6111* gene from the metagenomic library has been identified and heterologously expressed. The gene had a size of 2,520 bp with a GC content of 67%. BLAST analysis showed that the *bgl6111* gene had a 67.61% similarity to BGL from uncultured bacterium (ABB51613.1). The *bgl6111* gene-encoded 839 amino acids, which were predicted to have a molecular weight of 89.4 kDa. Phylogenetic analysis revealed that *bgl6111* was closely related to BGL from prokaryotic organisms.

The *bgl6111* gene was cloned in the pPICZαA plasmid and was successfully expressed in *P. pastoris* KM71. Even though the band was thin, SDS-PAGE results proved the presence of BGL protein products by showing a band of 89.4 kDa. The BGL product had the highest activity at D3 (1.210 U/mL).

These experimental results show that *bgl6111* may be defined as a β-glucosidase with low activity. Although classified as low, the enzyme activity optimized by ultrafiltration could enhance up to 539.8% (7.742 U/mL). This study may provide valuable information for various novel β-glucosidase from nature, which may be used as an alternative enzyme for industrial purposes. However, this result may explain not only the enzyme activity to hydrolyze the substrates but also that we could boost the activity higher from the metagenomic approach.

1. **Conclusion**

The bgl6111 gene is a novel β-glucosidase obtained from a metagenomic library from sugarcane bagasse. The bgl gene was successfully cloned and expressed in Pichia pastoris KM71. The bgl6111 gene was successfully synthesized and produced the β-glucosidase enzyme product. This enzyme may be helpful for its application in industrial production in the future and needs to be re-modified to produce high activity. Moreover, as we did here, heterologously expressed functional enzymes by genetically modifying the host organism can also be practical tools in developing enzymes with desired properties.

**Acknowledgments**

Diponegoro University supports this study for **RPI grant number 474-91/UN7.P4.3/PP/2018.**

**Competing Interests**

All authors declare that there are not any conflicts of interest.

**References**

Ahmed A, Nasim F ul-H, Batool K, Bibi A. 2017. Microbial β-Glucosidase: Sources, Production and Applications. Journal of Applied & Environmental Microbiology. 5(1):31–46. doi:10.12691/JAEM-5-1-4.

Behura SK, Severson DW. 2013. Codon usage bias : causative factors, quantification methods and genome-wide patterns : with emphasis on insect genomes. Biological Reviews Cambridge Philosophical Society. 88:49–61. doi:10.1111/j.1469-185X.2012.00242.x.

Borja GM, Mora EM, Barrón B, Gosset G, Ramírez OT, Lara AR. 2012. Engineering *Escherichia coli* to increase plasmid DNA production in high cell-density cultivations in batch mode. Microbial Cell Factories. 11(132):1–9. https://doi.org/10.1186/1475-2859-11-132.

Chamoli S, Kumar P, Navani NK, Verma AK. 2016. Secretory expression, characterization and docking study of glucose-tolerant β-glucosidase from *B. subtilis*. International Journal of Biological Macromolecules. 85:425–433. http://dx.doi.org/10.1016/j.ijbiomac.2016.01.001.

Chen P, Fu X, Bun T, Ye NX. 2011. Expression of a secretory β-glucosidase from Trichoderma reesei in Pichia pastoris and its characterization. Biotechnology Letters. 33:2475–2479. doi:10.1007/s10529-011-0724-3.

Choi T, Gelutu TT. 2018. High level expression and purification of recombinant flounder growth hormone in *E. coli*. Journal of Genetic Engineering and Biotechnology. 16(2):347–355. https://doi.org/10.1016/j.jgeb.2018.03.006.

Chuck C, Wong C, Chow LM, Fung K, Waye MM, Tsui SK. 2009. Expression of SARS-coronavirus spike glycoprotein in *Pichia pastoris*. Virus Genes. 38:1–9. doi:10.1007/s11262-008-0292-3.

Dodda SR, Aich A, Sarkar N, Jain P, Jain S, Mondal S, Aikat K, Mukhopadhyay SS. 2018. Structural and functional insights of β-glucosidases identified from the genome of *Aspergillus fumigatus*. Journal of Molecular Structure. 1156:105–114. https://doi.org/10.1016/j.molstruc.2017.11.078.

Farshadpour F, Makvandi M, Taherkhani R. 2015. Design, Construction, and Cloning of Truncated ORF2 and tPAsp-PADRE-Truncated ORF2 Gene Cassette from Hepatitis E Virus in the pVAX1 Expression Vector. Jundishapur Journal of Microbiology. 8(12):1–11. doi:10.5812/jjm.26035.

Gao G, Wang A, Gong B, Li Q, Liu Y, He Z, Li G. 2016. A novel metagenome-derived gene cluster from termite hindgut: Encoding phosphotransferase system components and high glucosetolerant glucosidase. Enzyme and Microbial Technology. 84:24–31. http://dx.doi.org/10.1016/j.enzmictec.2015.12.005.

Gomes-pepe S, Guillermo E, Sierra M, Pereira R, Cristina T, Castellane L, De EG, Lemos M. 2016. Bg10 : A Novel Metagenomics Alcohol-Tolerant Suitable for Lactose-Free Milk Preparation. PLoS One. 11(12):1–25. doi:10.1371/journal.pone.0167932.

Hoseini SS, Sauer MG. 2015. Molecular cloning using polymerase chain reaction , an educational guide for cellular engineering. Journal of Biological Engineering. 9(2):1–12. doi:10.1186/1754-1611-9-2.

Kanokratana P, Eurwilaichitr L, Pootanakit K, Champreda V. 2015. Identification of glycosyl hydrolases from a metagenomic library of microflora in sugarcane bagasse collection site and their cooperative action on cellulose degradation. Journal of Bioscience and Bioengineering. 119(4):384–391. http://dx.doi.org/10.1016/j.jbiosc.2014.09.010.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 10(6):845–858. doi:10.1038/nprot.2015-053.

Kumar TA. 2013. CFSSP : Chou and Fasman Secondary Structure Prediction server. Wide Spectrum. 1(9):15–19. doi: 10.5281/zenodo.50733

Lewin A, Lale R, Wentzel A. 2017. Expression Platforms for Functional Metagenomics : Emerging Technology Options Beyond *Escherichia coli*. In: Functional Metagenomics: Tools and Applications. Cham: Springer International Publishing. p. 13–44.

Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, Fotiadis D, Kovar K. 2015. Cultivation strategies to enhance productivity of *Pichia pastoris* : A review. Biotechnology Advances. 33(6):1177–1193. http://dx.doi.org/10.1016/j.biotechadv.2015.05.008.

Lu J, Du L, Wei Y, Hu Y, Huang R. 2013. Expression and characterization of a novel highly glucose-tolerant β-glucosidase from a soil metagenome. Acta Biochimica et Biophysica Sinica. 45(8):664–673. doi:https://doi.org/10.1093/abbs/gmt061.

Matsuzawa T, Watanabe M, Yaoi K. 2017. Improved thermostability of a metagenomic glucose-tolerant β-glycosidase based on its X-ray crystal structure. Applied Microbiology and Biotechnology. 101:8353–8363. doi:10.1007/s00253-017-8525-9.

Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsornruang S. 2015. Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. Biotechnology for Biofuels. 8(16). doi:10.1186/s13068-015-0200-8.

Pearson WR. 2013. An Introduction to Sequence Similarity (“Homology”) Searching. Current Protocols in Bioinformatic.:1–9. doi:10.1002/0471250953.bi0301s42.An.

Pozo MV Del, Fernández-arrojo L, Gil-martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, et al. 2012. Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnology for Biofuels. 5(73):1–13. doi:10.1186/1754-6834-5-73.

Prakash S, Singhal RS, Kulkarni PR. 2002. Enzymic debittering of Indian grapefruit (Citrus paradisi) juice. Journal of the Science Food and Agriculture. 82:394–397. doi:10.1002/jsfa.1059.

Prayogo FA, Budiharjo A, Kusumaningrum HP, Wijanarka W, Suprihadi A, Nurhayat N. 2020. Metagenomic applications in exploration and development of novel enzymes from nature : a review. Journal of Genetic Engineering and Biotechnology. 18:1–10. doi:https://doi.org/10.1186/s43141-020-00043-9.

Quax TEF, Claassens NJ, Oost J Van Der, Haven N. 2016. Codon Bias as a Means to Fine-Tune Gene Expression. Molecular Cell. 59(2):149–161. doi:10.1016/j.molcel.2015.05.035.

Ramírez-escudero M, Pozo M V, Marín-navarro J, González B, Golyshin PN, Polaina J, Ferrer M, Sanz-aparicio J. 2016. Structural and Functional Characterization of a Ruminal β-Glycosidase Defines a Novel Subfamily of Glycoside Hydrolase Family 3 with Permuted Domain Topology \*Edited by Gerald Hart. Journal of Biological Chemistry. 291(46):24200–24214. doi:10.1074/jbc.M116.747527.

Rouyi C, Baiya S, Lee S, Mahong B, Jeon J, Ketudat- JR. 2014. Recombinant Expression and Characterization of the Cytoplasmic Rice β-Glucosidase Os1BGlu4. Plos one. 9(5):1–13. doi:10.1371/journal.pone.0096712.

Shrestha A, Palmieri N, Abd-elfattah A, Ruttkowski B, Pagès M, Joachim A. 2017. Cloning, expression and molecular characterization of a Cystoisospora suis specific uncharacterized merozoite protein. Parasit Vectors. 10(68):1–13. doi:10.1186/s13071-017-2003-1.

Singh G, Verma AK, Kumar V. 2016. Catalytic properties, functional attributes and industrial applications of β-glucosidases. 3 Biotech. 6(1):1–14. doi:10.1007/s13205-015-0328-z.

Singhania RR, Patel AK, Saini R, Pandey A. 2017. Industrial Enzymes: β-Glucosidases. In: Current Developments in Biotechnology and Bioengineering: Production, Isolation and Purification of Industrial Products. Amsterdam: Elsevier B.V. p. 103–125.

Tang H, Hou J, Shen Y, Xu L, Yang H, Fang X, Bao X. 2013. High β-Glucosidase Secretion in *Saccharomyces cerevisiae* Improves the Efficiency of Cellulase Hydrolysis and Ethanol Production in Simultaneous Saccharification and Fermentation. Journal of Microbiology and Biotechnology. 23(11):1577–1585. http://dx.doi.org/10.4014/jmb.1305.05011.

Trivedi RN, Akhtar P, Meade J, Bartlow P, Ataai MM, Khan SA, Domach MM. 2014. High-Level Production of Plasmid DNA by Introducing inc Mutations. Applied and Environmental Microbiology. 80(23):7154–7160. doi:10.1128/AEM.02445-14.

Vogl T, Gebbie L, Palfreyman RW. 2018. Effect of Plasmid Design and Type of Integration Event on Recombinant Protein Expression in *Pichia pastoris*. Applied and Environmental Microbiology. 84(6):e02712-17. https://doi .org/10.1128/AEM.02712-17.

Wang J, Lu L, Feng F. 2017. Combined strategies for improving production of a thermo-alkali stable laccase in *Pichia pastoris*. Electronic Journal of Biotechnology. 28:7–13. http://dx.doi.org/10.1016/j.ejbt.2017.04.002

Wass MN, Kelley LA, Sternberg MJE. 2010. 3DLigandSite : predicting ligand-binding sites using similar structures. Nucleic Acids Research. 38:469–473. doi:10.1093/nar/gkq406.

Wooley JC, Ye Y. 2009. Metagenomics: Facts and Artifacts, and Computational Challenges. Journal of Computer Science and Technology. 25(1):71–81. doi:10.1007/s11390-010-9306-4.

Yang F, Yang X, Li Z, Du C, Wang J, Li S. 2015. Overexpression and characterization of a glucose-tolerant β-glucosidase from *T. aotearoense* with high specific activity for cellobiose. Applied Microbiology and Biotechnology. 99(21):8903–8915. doi:10.1007/s00253-015-6619-9.

Yu X, Sun W, Wang Y, Xu Y. 2017. Identification of novel factors enhancing recombinant protein production in multi-copy *Komagataella phaffii* based on transcriptomic analysis of overexpression effects. Scientific Reports. 7(16249):1–12. http://dx.doi.org/10.1038/s41598-017-16577-x.

Yu Y, Zhou X, Wu S, Wei T, Yu L. 2014. High-yield production of the human lysozyme by *Pichia pastoris* SMD1168 using response surface methodology and high-cell-density fermentation. Electronic Journal of Biotechnology. 17(6):311–316. http://dx.doi.org/10.1016/j.ejbt.2014.09.006.

Zang X, Liu M, Fan Y, Xu J, Xu X, Li H. 2018. The structural and functional contributions of β‑glucosidase‑producing microbial communities to cellulose degradation in composting. Biotechnology for Biofuels. 11(51):1–13. https://doi.org/10.1186/s13068-018-1045-8.

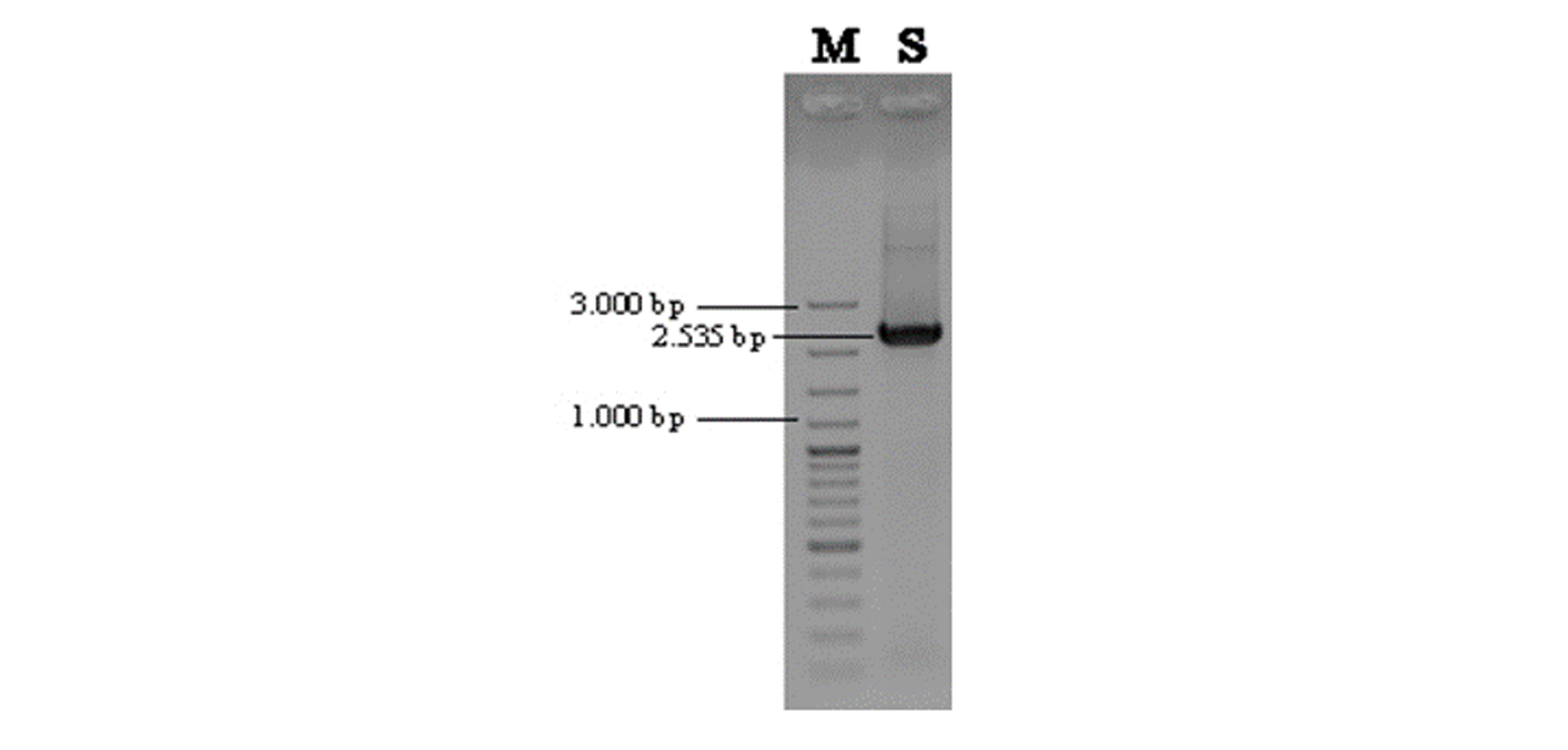
Zhang L, Fu Q, Li W, Wang B, Yin X, Liu S, Xu Z, Niu Q. 2017. Identification and characterization of a novel β-glucosidase via metagenomic analysis of *Bursaphelenchus xylophilus* and its microbial flora. Scientific Reports. 7:1–11. http://dx.doi.org/10.1038/s41598-017-14073-w.

Zhang XF, Ai YH, Xu Y, Yu XW. 2019. High-level expression of *Aspergillus niger* lipase in *Pichia pastoris*: Characterization and gastric digestion in vitro. Food Chemistry. 274(2019):305–313. https://doi.org/10.1016/j.foodchem.2018.09.020.

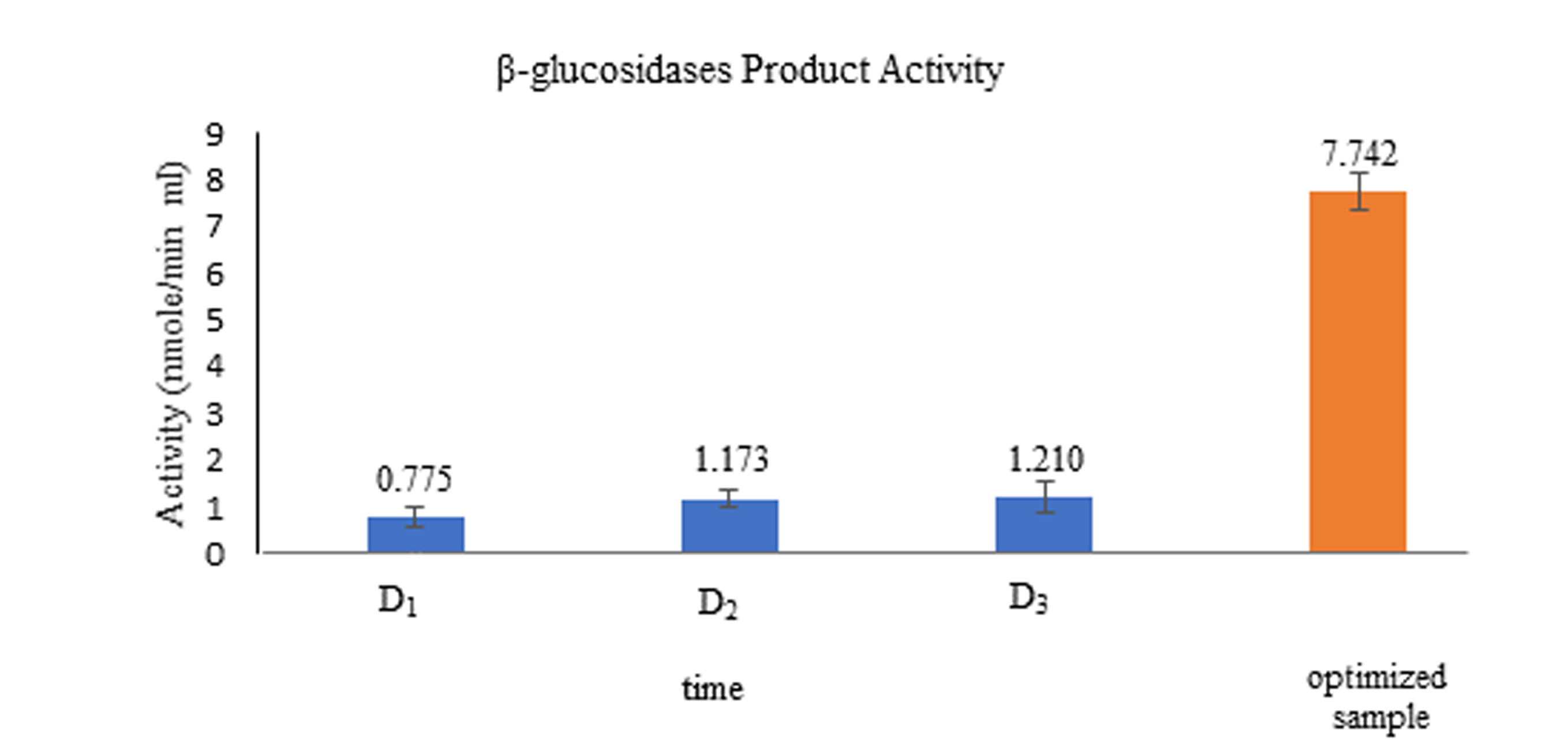
Zhao L, Xie J, Zhang X, Cao F, Pei J. 2013. Overexpression and characterization of a glucose-tolerant β-glucosidase from *Thermotoga thermarum* DSM 5069T with high catalytic efficiency of ginsenoside Rb1 to Rd. Journal of Molecular Catalysis B: Enzymatic. 95:62–69. http://dx.doi.org/10.1016/j.molcatb.2013.05.027.

Zuhair M, Nor M, Vasiljevic T. 2018. Performance of a Two-Stage Membrane System for Bromelain Separation from Pineapple Waste Mixture as Impacted by Enzymatic Pretreatment and Diafiltration. Food Technology and Biotechnology. 56(2):218–227. doi:10.17113/ftb.56.02.18.5478.

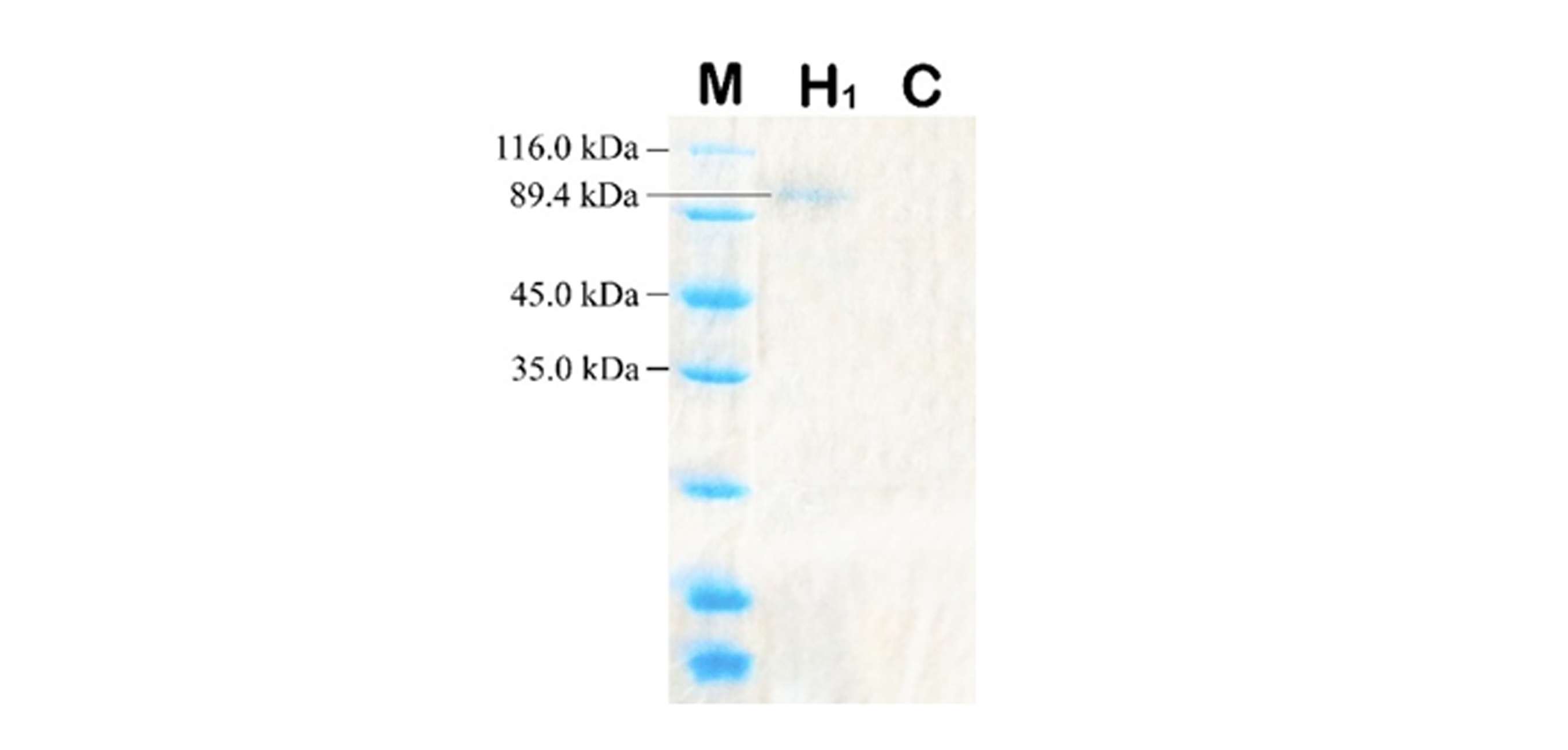
**Figures**

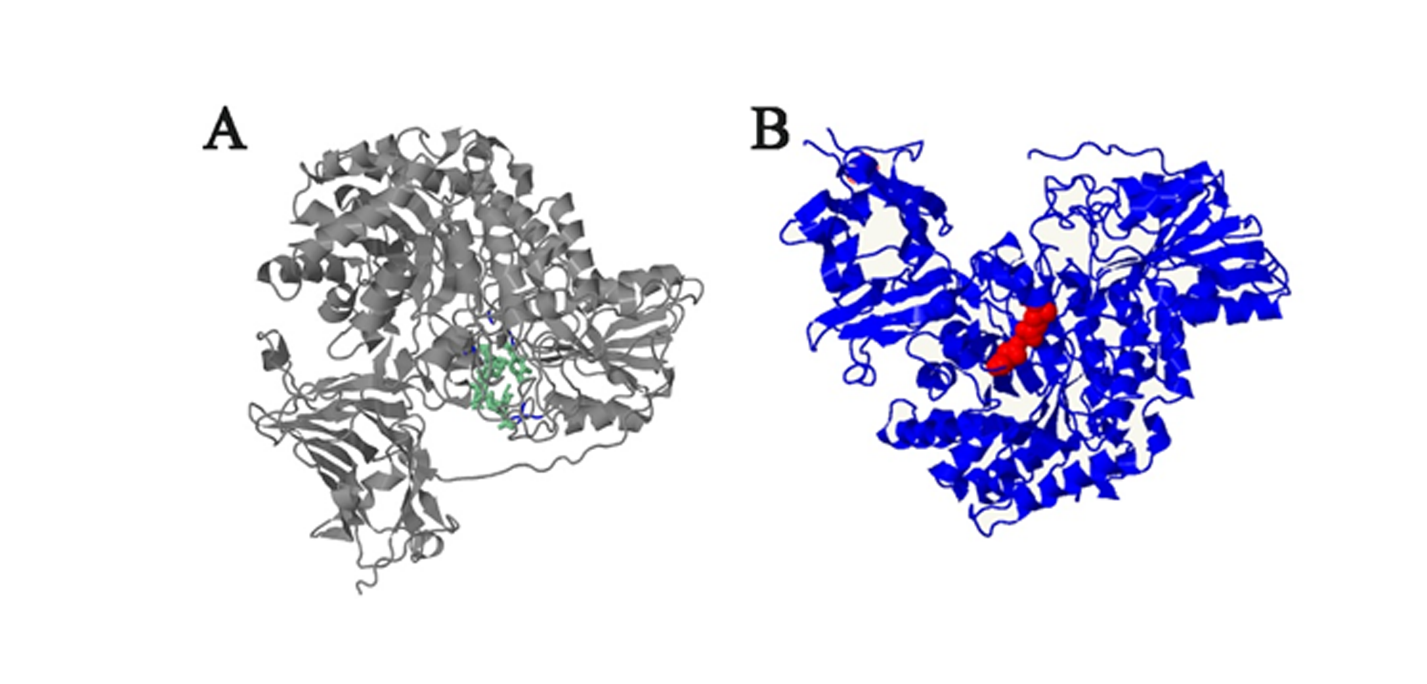


**Figure 1.** Gel PCR electrophoresis results of the *bgl6111* gene product show a band size of 2,535 bp. {M = 100bp plus DNA ladder (Thermo Scientific); S = sample *bgl6111* gene}.

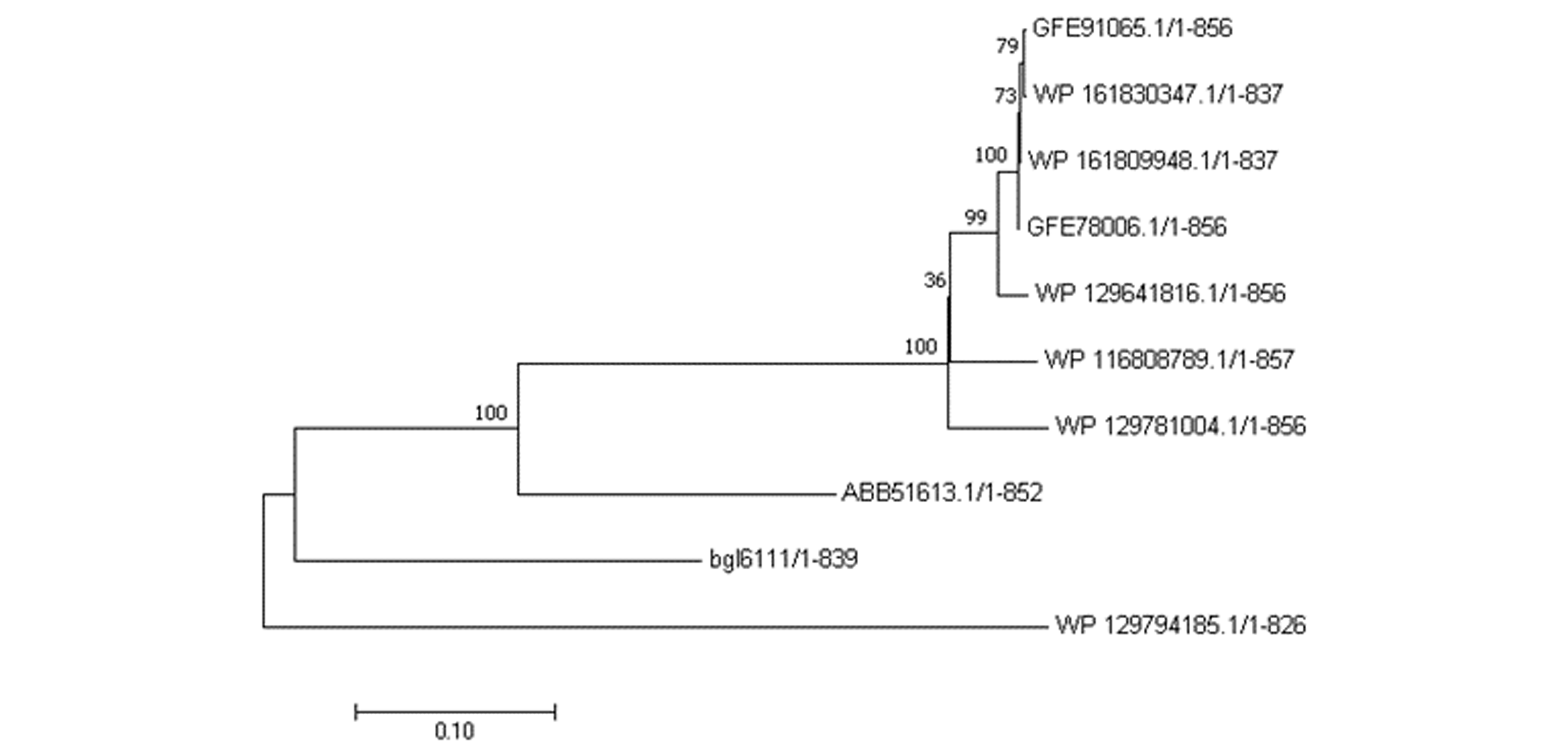


**Figure 2.** The Graph of β-glucosidase recombinant activity with a standard deviation line (D1 = first day of fermentation; D2 = second day of fermentation; D3 = third day of fermentation; H1 = optimized sample).



**Figure 3.** The SDS-Page results of recombinant product β-glucosidase protein. The protein band shows a size of 89.4 kDa. M = marker 14.4-120.0 kDa, LabAid; H1= optimized sample; C= control, empty plasmid).

**Figure 4.** Structural model of bgl6111 protein generated by Phyre2. **(4A)** Ligand site is highlighted in green, and **(4B)** the catalytic site is highlighted in red.



**Figure 5.** Phylogenetic analysis based on amino acids from β-glucosidase (*bgl6111*) using the neighbour-joining (NJ) method with a bootstrap value of 1,000. *bgl6111*= sample used in this study, ABB51613.1= β-glucosidase from uncultured bacterium, GFE91065.1= glucan 1,4-beta-glucosidase and WP\_161830347.1= GH3 C-terminal domain-containing protein from Steroidobacter sp. SA29-B, GFE78006.1= glucan 1,4-beta-glucosidase and WP\_161809948.1= GH3 C-terminal domain-containing protein from *Steroidobacter* sp. YU21-B, WP\_129641816.1= exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter agariperforans*, WP\_129781004.1= exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter* sp. JW-3, WP\_116808789.1= exo 1,3/1,4-beta-D-glucan glucohydrolase from *Steroidobacter cummioxidans*, and WP\_129794185.1= exo 1,3/1,4-beta-D-glucan glucohydrolase from Sphingosinicella sp. CPCC 101087.