Indonesian Journal of Biotechnology

VOLUME 28(4), 2023, 200-208 | RESEARCH ARTICLE

Cloning and characterization of *bgl6111* gene encoding β-glucosidase from bagasse metagenome

Fitra Adi Prayogo[1], Benjarat Bunterngsook[2], Pattanop Kanokratana[2], Hermin Pancasakti Kusumaningrum[3], Dyah Wulandari[4], Anto Budiharjo[3,4,*]

[1]Biomedical 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
[3]Biotechnology Study Program, Diponegoro University, Jl. Prof Soedharto SH, Semarang 50275, Indonesia
[4]Molecular and Applied Microbiology Laboratory, Central Laboratory of Research and Service, Diponegoro University, Jl. Prof. Sudharto SH, Semarang 50275, Indonesia
*Corresponding author: anto.budiharjo@live.undip.ac.id

SUBMITTED 20 January 2023 **REVISED** 5 May 2023 **ACCEPTED** 22 August 2023

ABSTRACT β-Glucosidase (BGL) is an essential enzyme for the hydrolysis of cellulose in industrial processes, but natural BGL enzymes are poorly understood. Metagenomics is a robust tool for bioprospecting in the search for novel enzymes from the entire community's genomic DNA present in nature. The metagenomics approach simplifies the process of searching for new BGL enzymes by extracting DNA and retrieving its gene information through a series of bioinformatic analyses. In this study, we report the gene cloning, heterologous expression of the *bgl6111* gene (accession number MW221260) in *Pichia pastoris* KM71, and the biochemical characterization of the recombinant enzyme. We successfully identified the *bgl6111* sequence of 2,520 bp and 839 amino acids with a molecular size of 89.4 kDa. The amino acid sequence of the *bgl6111* gene showed 67.61% similarity to BGL from an uncultured bacterium (ABB51613.1). The BGL product has the highest activity on the third day at 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 β-Glucosidase; Bagasse; Cloning; Metagenomic library

1. Introduction

The β-glucosidases (BGL) (EC 3.2.1.21) are enzymes that belong to the cellulase enzyme complex group and classified among the glycoside hydrolases (GH) (Ahmed et al. 2017; Rouyi et al. 2014). These enzymes play a crucial role in hydrolyzing glycosidic bonds to release nonreducing terminal glucosyl residues (Rouyi et al. 2014). BGL enzymes are vital contributors to the hydrolysis of cellulose in various biotechnology industrial processes, including composting (Zang et al. 2018), breakdown of naringin in grapefruit juice (Prakash et al. 2002), and ethanol production (Tang et al. 2013). Additionally, BGL enzymes find applications in producing aromatic compounds for the flour industry, wine production, hydrolysis of anthocyanin products, enhancing organoleptic qualities of fruits and juices (Singh et al. 2016), and serving as additives in animal feed to improve food digestibility (Singhania et al. 2016). Despite their wide-ranging applications, the exploration of natural BGL enzymes remains relatively limited. Nature offers a diverse array of BGL enzymes with different types and characteristics that hold potential for development in industrial processes.

Cloning and expression of recombinant protein methods are necessary for enzyme exploration (Zhao et al. 2013). Researchers have explored various *bgl* encoding BGL enzymes produced by microorganisms. It was successfully isolated the *bgl* gene from *Bacillus subtilis* and expressed in *Escherichia coli*, to produce the recombinant enzyme with 54.04 U/mg specific activity Chamoli et al. (2016). Yang et al. (2015) expressed the *bgl* gene from *Thermoanaerobacterium aotearoense* in *E. coli* with a specific activity of 740.5 U/mg. In addition, to isolate the *bgl* gene from microorganisms, it can be obtained directly from the environmental DNA (Mercedes 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 the 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, and 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. The composition contained in the environment can characterize the types of genes present in the environment. For example, 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 in the sugarcane bagasse environment may lead to the discovery of a novel BGL enzyme as an alternative candidate enzyme for industrial purposes.

The aim of this research was to attempt to clone the sequence from several lignocellulolytic enzymes which was previously constructed and identified through activity-based screening from the bagasse metagenomic library (Kanokratana et al. 2015; Mhuantong et al. 2015) . This recombinant enzyme was constructed 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 predicted 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, which is 539.8% higher if compared with non-treatment.

2. Materials and Methods

2.1. Microbial strain, plasmid, and culture

The purchased vector pPICZαA from Invitrogen was used as the DNA vector for gene construction. 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). The resulting recombinant DNAs were introduced into *E. coli* DH5α by electroporation (Biorad) with 15 kv/cm, 100 Ω, and 25 µF. *E. coli* DH5α was used as the propagation host, and *P. pastoris* KM71 was used as the expression host (Haniyya et al. 2021).

The medium used in this study are as follows: 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 100 mM 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 pept% 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).

2.2. 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/*Eco*RI/F (5′ GAATTCATGGCATGCGTGCTCGCAGCCTTT 3′) and BGL6111/*Xba*I/R (5′TCTAGATCATCCCGTGCACG-GAAGGGTGCC 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 (Thermo Scientific) was used to purify the PCR product. *Eco*RI and *Xba*I 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 pPICZαplasmid. 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 harboring the corrected recombinant plasmid were selected from LB agar containing 25 µg/mL of Zeocin. The transformants were screened by colony PCR using 5′-AOX1 forward and 3′-AOX1 reverse primers. The gene sequence in the plasmid was subsequently confirmed by conventional sequencing (Macrogen).

2.3. 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. A total of 300 colonies of the transformants were collected by dots on the YPD containing 100 µg/mL of Zeocin. The positive clones 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 OD_{600} 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 and 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 (Nor et al. 2018). The BMGY culture was scaled up to 200 mL and incubated at 30 °C until the OD_{600} 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).

2.4. Enzyme assay and SDS-PAGE analysis

The measurement of BGL activity involved using 1 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 Na_2CO_3 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.

Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was conducted to confirm 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 4 °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.

2.5. 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 (Ashok 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 has been submitted by the authors in the NCBI database with accession number MW221260.

3. Results and Discussion

3.1. 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.

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

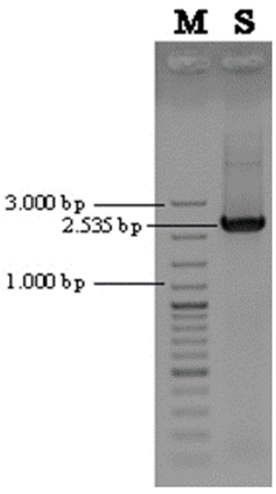

FIGURE 1 Gel PCR electrophoresis of the bgl6111 gene showed a band size of 2,535 bp. M = 100 bp plus DNA ladder (Thermo Scientific); S = sample *bgl6111* gene.

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 using Sanger sequencing method.

3.2. Expression of recombinant plasmid pPICZαA–bgl6111

The GeneJET Plasmid Miniprep Kit (Thermo Scientific) isolated the recombinant plasmid pPICZαA–*bgl6111* from *E. coli* DH5α. *Pme*I 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.

3.3. Confirmation of the presence of BGL

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%.

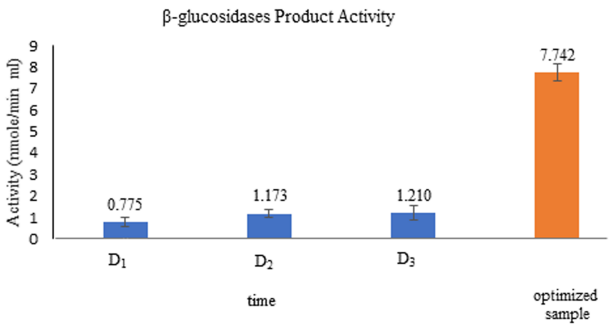

FIGURE 2 The graph of β-glucosidase recombinant activity with a standard deviation line (D1 = first day of fermentation; D2 = second day of fermentation; D3 = third day of fermentation; H1 = optimized sample).

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.

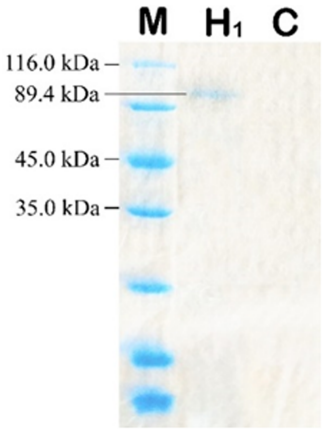

FIGURE 3 The SDS-Page results of recombinant product β-glucosidase protein. The protein band shows a size of 89.4 kDa. M = marker 14.4-120.0 kDa, LabAid; H1= optimized sample; C= control, empty plasmid).

3.4. 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).

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,

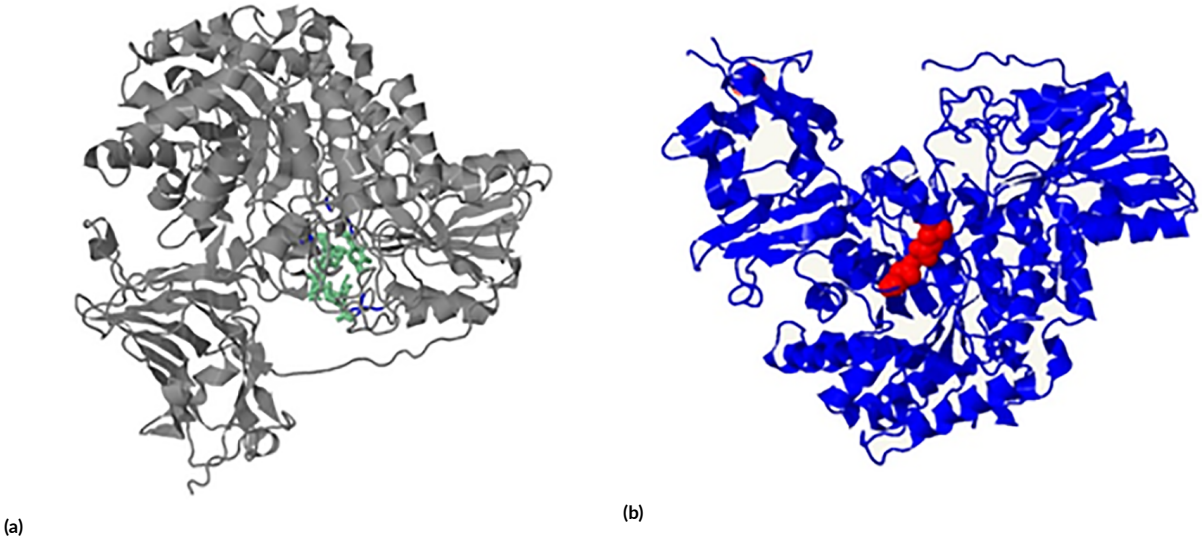

FIGURE 4 Structural model of *bgl6111* protein generated by Phyre2. (a) Ligand site is highlighted in green, and (b) the catalytic site is highlighted in red.

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% (Ashok 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.

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). In the previous report, several genes 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. Del 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 Geletu 2018). This outcome strengthens the argument of previous research about an existing relationship between low

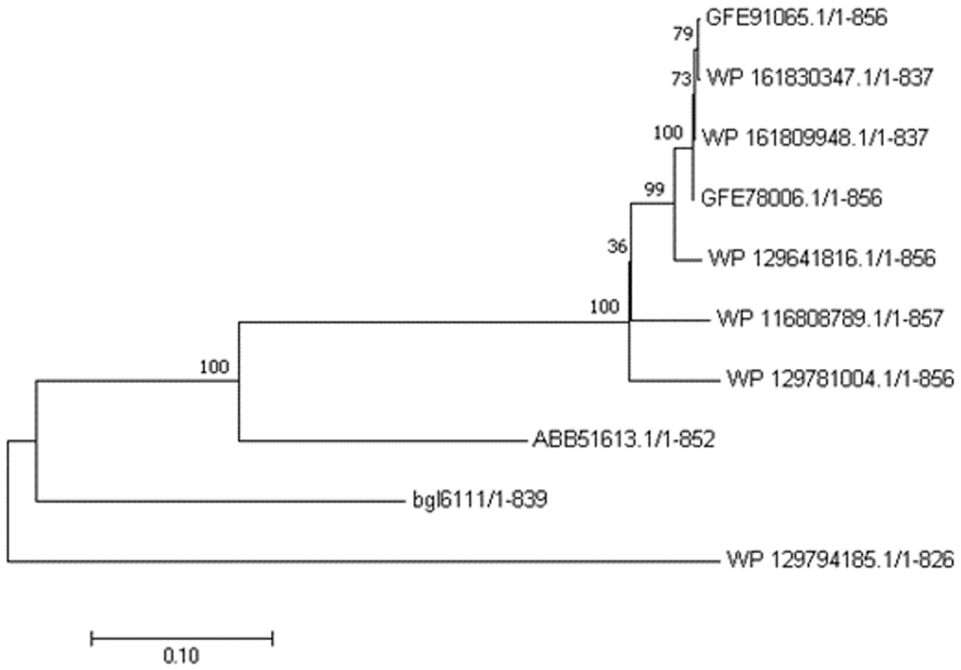

FIGURE 5 Phylogenetic analysis based on amino acids from β-glucosidase (*bgl6111*) using the neighbor-joining (NJ) method with a bootstrap value of 1,000. *bgl6111*= sample used in this study, ABB51613.1= β-glucosidase from uncultured bacterium, GFE91065.1= glucan 1,4-beta-glucosidase and WP_161830347.1= 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.

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 (2013), codon bias is when specific codons are used more frequently than other synonym codons. It often occurs in heterogeneous gene expression. Quax et al. (2015) 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 to be 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 environmental DNA (eDNA) that 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, particularly in 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 showed that *bgl6111* might 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.

4. Conclusions

The *bgl6111* gene is a novel β-glucosidase obtained from a metagenomic library from sugarcane bagasse. The *bgl6111* 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 remodified 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

AB would like to thanks Diponegoro University for supporting this project under RPI grant number 185-80/UN7.6.1/PP/2021 and WCU grant for staff exchange 2022. DW would like to thank Diponegoro University for the Postdoc Grant SK Rektor: 216/UN7.A/HK/XI/2023.

Authors' contributions

AB and PK concept the research and manuscript. HPK and BB carried out literature search. DW and FAP worked to shape and edit the manuscript. All authors contributed to the final manuscript.

Competing interests

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

References

Ahmed A, Nasim FUH, Batool K, Bibi A. 2017. Microbial β-glucosidase: Sources, production, and applications. J. Appl. Environ. Microbiol. 5(1):31–46. doi:10.12691/jaem-5-1-4.

Ashok Kumar T. 2013. CFSSP: Chou and Fasman secondary structure prediction server. Wide Spectr. 1(9):15–19. doi:10.5281/zenodo.50733.

Behura SK, Severson DW. 2013. Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. Biol. Rev. 88(1):49–61. doi:10.1111/j.1469-185X.2012.00242.x.

Bhatia Y, Mishra S, Bisaria VS. 2002. Microbial β-glucosidases: Cloning, properties, and applications. Crit. Rev. Biotechnol. 22(4):375–407. doi:10.1080/07388550290789568.

Borja GM, Meza Mora E, 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. Microb. Cell Fact. 11:132. doi: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*. Int. J. Biol. Macromol. 85:425–33. doi:10.1016/j.ijbiomac.2016.01.001.

Chen P, Fu X, Ng TB, Ye XY. 2011. Expression of a secretory β-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. Biotechnol. Lett. 33(12):2475–9. doi:10.1007/s10529-011-0724-3.

Choi TJ, Geletu TT. 2018. High level expression and purification of recombinant flounder growth hormone in *E. coli*. J. Genet. Eng. Biotechnol. 16(2):347–355. doi:10.1016/j.jgeb.2018.03.006.

Chuck CP, Wong CH, Chow LMC, Fung KP, Waye MMY, Tsui SKW. 2009. Expression of SARS-coronavirus spike glycoprotein in *Pichia pastoris*. Virus Genes 38(1):1–9. doi:10.1007/s11262-008-0292-3.

Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, Golyshina OV, Newbold CJ, Polaina J, Ferrer M, Golyshin PN. 2012. Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnol. Biofuels 5(1):73. doi:10.1186/1754-6834-5-73.

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*. J. Mol. Struct. 1156:105–114. doi: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 J. Microbiol. 8(12):e26035. doi:10.5812/jjm.26035.

Gao G, Wang A, Gong BL, Li QQ, Liu YH, He ZM, Li G. 2016. A novel metagenome-derived gene cluster from termite hindgut: Encoding phosphotransferase system components and high glucose tolerant glucosidase. Enzyme Microb. Technol. 84:24–31. doi:10.1016/j.enzmictec.2015.12.005.

Gomes-Pepe ES, Sierra EGM, Pereira MR, Castellane TCL, De Lemos EGM. 2016. Bg10: A novel metagenomics alcohol-tolerant and glucose-

stimulated gh1 β-glucosidase suitable for lactose-free milk preparation. PLoS One 11(12):e0167932. doi:10.1371/journal.pone.0167932.

Haniyya, Mulyawati L, Helianti I, Pinmanee P, Kocharin K, Cantasingh D, Nimchua T. 2021. Characterization of recombinant *Bacillus halodurans* CM1 xylanase produced by *Pichia pastoris* KM71 and its potential application in bleaching process of bagasse pulp. Indones. J. Biotechnol. 26(1):15–24. doi:10.22146/IJBIOTECH.57701.

Hoseini SS, Sauer MG. 2015. Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. J. Biol. Eng. 9:2. 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. J. Biosci. Bioeng. 119(4):384–91. doi:10.1016/j.jbiosc.2014.09.010.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10(6):845–58. doi:10.1038/nprot.2015.053.

Lewin A, Lale R, Wentzel A. 2017. Expression platforms for functional metagenomics: Emerging technology options beyond *Escherichia coli*. Funct. Metagenomics Tools Appl. p. 13–44. doi:10.1007/978-3-319-61510-3_2.

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. Biotechnol. Adv. 33(6):1177–93. doi: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 Biochim. Biophys. Sin. (Shanghai). 45(8):664–73. doi: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. Appl. Microbiol. Biotechnol. 101(23-24):8353–8363. doi:10.1007/s00253-017-8525-9.

Mercedes RE, Julia MN, Del Pozo MV, González B, Golyshin PN, Polaina J, Ferrer M, Julia SA. 2016. Structural and functional characterization of a ruminal β-glycosidase defines a novel subfamily of glycoside hydrolase family 3 with permuted domain topology. J. Biol. Chem. 291(46):24200–24214. doi:10.1074/jbc.M116.747527.

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

Nor MZM, Ramchandran L, Duke 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 Technol. Biotechnol. 56(2):218–227. doi:10.17113/ftb.56.02.18.5478.

Pearson WR. 2013. An introduction to sequence similarity ("homology") searching. Curr. Protoc. Bioinforma. Chapter 3:3.1.1–3.1.8. doi:10.1002/0471250953.bi0301s42.

Prakash S, Singhal RS, Kulkarni PR. 2002. Enzymic debittering of Indian grapefruit (*Citrus paradisi*) juice. J. Sci. Food Agric. 82(4):394–397. doi:10.1002/jsfa.1059.

Prayogo FA, Budiharjo A, Kusumaningrum HP, Wijanarka W, Suprihadi A, Nurhayati N. 2020. Metagenomic applications in exploration and development of novel enzymes from nature: A review. J. Genet. Eng. Biotechnol. 18:39. doi:10.1186/s43141-020-00043-9.

Quax TE, Claassens NJ, Söll D, van der Oost J. 2015. Codon bias as a means to fine-tune gene expression. Mol. Cell 59(2):149–161. doi:10.1016/j.molcel.2015.05.035.

Rouyi C, Baiya S, Lee SK, Mahong B, Jeon JS, Ketudat-Cairns JR, Ketudat-Cairns M. 2014. Recombinant expression and characterization of the cytoplasmic rice β-glucosidase Os1BGlu4. PLoS One 9(5):e96712. 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. Parasites and Vectors 10:68. 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):3. doi:10.1007/s13205-015-0328-z.

Singhania RR, Patel AK, Saini R, Pandey A. 2016. Industrial enzymes: b-Glucosidases. Amsterdam: Elsevier B.V. p. 103–125. doi:10.1016/j.jclepro.2017.05.040.

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. J. Microbiol. Biotechnol. 23(11):1577–85. doi: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 *Escherichia coli* DH5$\alpha\Omega$sacB by introducing inc mutations. Appl. Environ. Microbiol. 80(23):7154–60. doi:10.1128/AEM.02445-14.

Vogl T, Gebbie L, Palfreyman RW, Speight R. 2018. Effect of plasmid design and type of integration event on recombinant protein expression in *Pichia pastoris*. Appl. Environ. Microbiol. 84(6):e02712–17. doi:10.1128/AEM.02712-17.

Wang J, Lu L, Feng F. 2017. Combined strategies for im-

proving production of a thermo-alkali stable laccase in *Pichia pastoris*. Electron. J. Biotechnol. 28:7–13. doi:10.1016/j.ejbt.2017.04.002.

Wass MN, Kelley LA, Sternberg MJ. 2010. 3DLigandSite: Predicting ligand-binding sites using similar structures. Nucleic Acids Res. 38:W469–73. doi:10.1093/nar/gkq406.

Wooley JC, Ye Y. 2009. Metagenomics: Facts and artifacts, and computational challenges. J. Comput. Sci. Technol. 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. Appl. Microbiol. Biotechnol. 99(21):8903–15. doi:10.1007/s00253-015-6619-9.

Yu XW, Sun WH, Wang YZ, Xu Y. 2017. Identification of novel factors enhancing recombinant protein production in multi-copy *Komagataella phaffii* based on transcriptomic analysis of overexpression effects. Sci. Rep. 7(1):16249. doi: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. Electron. J. Biotechnol. 17(6):311–316. doi: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. Biotechnol. Biofuels 11:51. doi: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. Sci. Rep. 7:14850. doi: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 Chem. 274:305–313. doi: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. J. Mol. Catal. B Enzym. 95:62–69. doi:10.1016/j.molcatb.2013.05.027.