Vietnam Journal of Biotechnology **24**(1): 121-131, 2026. DOI: 10.15625/vjbt-22443

Research Article

USING BIOINFORMATICS TOOLS TO MINE GENES ENCODING 6-PHOSPHO-β-GLUCOSIDASE FROM METAGENOMIC DATA OF BACTERIA SURROUNDING WHITE-ROT FUNGI

Thi Binh Nguyen[1], Thi Thu Hong Le[2] and Nam Hai Truong[2,✉]

[1]*Hanoi Metropolitan University, No. 98 Duong Quang Ham, Nghia Do, Hanoi, Vietnam*

[2]*Institute of Biology, Vietnam Academy of Science and Technology, No. 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

✉To whom correspondence should be addressed. Email: tnhai@ibt.ac.vn

Received: 22 February 2025; Accepted: 30 May 2025; Published online: 10 March 2026

ABSTRACT

6-Phospho-β-glucosidase (EC 3.2.1.86) is an enzyme involved in the hydrolysis of lignocellulose. It cleaves phosphorylated cellobioses to release glucose and glucose-6-phosphate. Therefore, identifying enzyme candidates with novel properties is crucial for enhancing the efficiency of lignocellulose hydrolysis. In this study, we mined 1718 open reading frames (ORFs) encoding 6-phospho-β-glucosidase from the DNA metagenome of bacteria in soil surrounding white-rot fungi, using the KEGG database. Of these, 1699 ORFs were classified as bacteria, 1633 ORFs (96.12%) were classified into three major bacterial phyla: Proteobacteria (1494 ORFs, 87.93%), Firmicutes (126 ORFs, 7.42%), and Actinobacteria (13 ORFs, 0.77%). Interestingly, the most abundant class was *Gammaproteobacteria*, which accounted for 86.64% of the Proteobacteria phylum, and *Enterobacterales* was the most dominant order (81.40%). Among the 1718 mined ORFs, 213 were complete, and 210 encoded proteins with functional regions. Of these, 152 ORFs contained the GH1 domain (84.87% from Proteobacteria), and 58 ORFs contained the GH4 domain (94.83% from Proteobacteria). The representative proteins had 28-39% alpha helices, 12-14% beta strands, and 2-4% disordered regions. The spatial structure models of these proteins were based on a published 6-phospho-β-glucosidase template, with 88-95% coverage and 100% confidence. The expression levels of the ORFs in *E. coli* were predicted using Periscope software. Two ORFs with the GH1 domain and three ORFs with the GH4 domain, from the Proteobacteria and Firmicutes phyla, showed high expression levels ranging from 904 mg/l to 4714 mg/l. All five selected enzymes were alkaline, with optimal enzyme activity at relatively high temperatures. These findings suggest that these enzymes hold significant potential for experimental designs aimed at industrial applications.

Keywords: 6-phospho-β-glucosidase, bioinformatics, DNA metagenome, expression level, taxonomy.

INTRODUCTION

Cellulose is a key biomass resource for industry today. A combination of enzyme

groups, including β-1,4-endoglucanase (EC 3.2.1.4), β-1,4-exoglucanase (EC 3.2.1.91), β-glucosidase (EC 3.2.1.21), and 6-phospho-β-glucosidase (EC 3.2.1.86), is required to effectively degrade biomass. Among these, the enzymes β-1,4-endoglucanase and β-1,4-exoglucanase work together to break down natural cellulose into the disaccharide cellobiose, which can, in turn, inhibit the activity of both enzymes (Zhang *et al.*, 2017). The enzymes β-glucosidase and 6-phospho-β-glucosidase are responsible for cleaving phosphorylated cellobiose and cellobiose into glucose and glucose-6-phosphate molecules. When these two enzymes are reversibly inhibited by the degradation end products, the activities of endoglucanase and exoglucanase are also inhibited (Monteiro *et al.*, 2020). Notably, in bacteria, monosaccharides and disaccharides must be phosphorylated before being absorbed via phosphate-dependent transport systems. Therefore, the bacterial enzyme 6-phospho-β-glucosidase plays a vital role in releasing phosphorylated monosaccharides and disaccharides (Deutscher *et al.*, 2006). Many studies have explored this enzyme in microorganisms to identify genes encoding 6-phospho-β-glucosidase with high industrial potential, such as the *BglA-2* gene from *Streptococcus pneumoniae* TIGR4 (Yu *et al.*, 2013) and the *BlBglH* gene from *Bacillus licheniformis* (Veldman *et al.*, 2020).

Soil is a rich and diverse ecosystem that harbors a wide range of microorganisms (Mhete *et al.*, 2020), making it a promising source for discovering new enzymes with high efficiency in cellulose degradation. This is particularly true for humus samples collected from areas surrounding white-rot fungi, which are capable of effectively metabolizing all components of wood (Pinar *et al.,* 2024). The hydrolysis of wood by white-rot fungi is often linked to enzymes produced by bacteria in the same ecosystem. The interactions between fungi and bacteria in these environments can be either supportive or competitive (Janusz *et al.*, 2017). During wood decomposition, fungi release secondary metabolites into the environment, producing toxins that acidify the surroundings quickly due to oxidation reactions (Pusztahelyi *et al.*, 2015). As a result, environmental conditions become highly selective for bacteria, which must possess traits suited to these harsh conditions in order to survive.

To efficiently explore and identify genes from microorganisms in various ecosystems, metagenomics is employed to discover new genes encoding enzymes from unculturable microorganisms. The combination of metagenomics and next-generation sequencing techniques has generated vast amounts of metagenomic data. To maximize the potential of this data, bioinformatics tools are used to identify and characterize candidate genes encoding proteins of interest before experimental validation. In this study, we demonstrate the use of bioinformatics tools to mine genes encoding 6-phospho-β-glucosidase from the microbial metagenome of humus surrounding white-rot fungi in Cuc Phuong National Park.

MATERIALS AND METHODS

The 51.8 Gb metagenomic DNA data from the humus sample surrounding wood-hydrolyzed white-rot fungi in Cuc Phuong National Park were obtained using the HiSeq Illumina sequencing system (Illumina, San Diego, USA) at BGI, China. A total of 4,104,872 open reading frames (ORFs) were identified using the MGA software

(http://metagene.nig.ac.jp/metagene/metagene.html). These ORFs were then searched against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database for the identification of ORFs encoding 6-phospho-β-glucosidase with an e-value threshold less than 1e-5 (Le *et al.*, 2022).

Taxonomic prediction of genes encoding 6-phospho-β-glucosidase

Gene classification was performed using BLASTp with protein sequences from the NR database. The results were then analyzed using MEGAN software (version 4.6) (Huson *et al.*, 2007). Genes classified at the same taxonomic level were aggregated, and the classification results were visualized using the Krona add-in tool in Excel.

Prediction of functional regions of ORFs using PFAM and HMMER

To predict the functional regions of ORFs, protein FASTA sequences were submitted to Pfam using an e-value cutoff of 1.0. A personal email address was provided to receive the results of the functional regions from the HMMer website (https://www.ebi.ac.uk/Tools/hmmer/search/phmmer). After 2–3 days, the results were sent to the personal email address.

Prediction of protein spatial structure

The three-dimensional structure of proteins was predicted using Phyre2 software, based on homology modeling (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). Protein sequences were submitted to determine the secondary and tertiary structures, domain composition, and model quality of the protein.

Prediction of protein expression level using periscope software

Protein expression levels in *E. coli* were predicted using the free Periscope software, available at http://lightning.med.monash.edu/periscope/. The expression levels of soluble proteins were categorized into three levels: high, moderate, and low. The software also predicts the amount of soluble protein in mg/l.

Prediction of physical properties of proteins

The acid- or alkali-loving ability of proteins was predicted using AcalPred software (http://lin-group.cn/server/AcalPred). FASTA protein sequences were entered into the search box, and the software returned the acid- or alkali-loving ability of the protein in a few minutes. A maximum of 100 sequences can be submitted per prediction. The optimal enzyme activity temperature was predicted using TBI software (http://www.tbi.org.tw/tools/). FASTA amino acid sequences were input, and the results were available after a few minutes. Heat resistance was predicted at three levels: if the melting temperature (Tm) index is > 1, the predicted Tm is above 65°C; if the Tm index is < 0, the predicted Tm is below 55°C; and if the Tm index is between 0 and -1, the predicted Tm is between 55°C and 65°C.

RESULTS AND DISCUSSION

Taxonomic prediction of ORFs encoding 6-phospho-β-glucosidase

Using the KEGG database and MGA software, a total of 1718 ORFs were predicted to encode 6-phospho-β-glucosidase. A total of 1718 genes were identified

to encode 6-phospho-β-glucosidase, of which 1700 genes (98.85%) were annotated in the NR database. These genes were classified using MEGAN analysis. Among them, bacteria were the dominant group, with 1699 ORFs (99.94% of the total) classified as bacteria, while the remaining gene was from Eukaryota (0.06%). Within the bacterial group, 1633 ORFs (96.12%) were identified as belonging to three phyla: Proteobacteria, Firmicutes, and Actinobacteria. Proteobacteria was the most abundant bacterial phylum, comprising 1494 ORFs, (87.93%), which was 11.85 times greater than the abundance of Firmicutes 126 ORFs (7.42%) and Actinobacteria 13 ORFs (0.77%). The dominance of Proteobacteria suggests its significant role in the cleavage of phosphorylated cellobioses. This result aligns with studies by Pang *et al.* (2021), who found Proteobacteria and Acidobacteria to be the dominant phyla in soils surrounding sugarcane plants under continuous monoculture cultivation. Similarly, Haq *et al.* (2022) observed that Proteobacteria was the most dominant phylum in the bacterial community around fungal rot on dead birch trunks.

Table 1. Results of classification of the 1700 ORFs.

Sources	Gene number	Percentage (%)	Phylum	Class	Order	Family	Genus	Species
Bacteria	1699	99.94	3	5	12	19	44	3
Eukaryota	1	0.06	1	1	1	1	1	1
Sum	1700	100	4	6	13	20	45	4

At the class level, 1608 ORFs (94.64%) were classified, including Gammaproteobacteria 1472 ORFs (86.64%), Bacilli 93 ORFs (5.47%) and Clostridia 25 ORFs (1.47%), other classes represented less than 1%. The high abundance of Gammaproteobacteria in the soil microbial sample surrounding white-rot fungi suggests that the fungi created a selective environment conducive to the growth of this class. This finding is consistent with studies suggesting that Gammaproteobacteria thrive in environments affected by pollutants, such as those found in polluted soils (Zhang *et al.*, 2021). Furthermore, Gammaproteobacteria are known to survive under extreme conditions, such as high pressure or salinity, as found in deep-sea environments (Franco *et al.*, 2017).

The most dominant order was *Enterobacterales*, which accounted for 81.40%, followed by *Lactobacillales* (4.36%), *Clostridiales* (1.47%), and *Aeromonadales* (1.47%), with the remaining orders representing less than 1%. *Enterobacterales* are known to thrive in extreme environments, such as those polluted with heavy metals or containing antibiotics (Glushakova *et al.*, 2022). The results of this study also show that *Enterobacterales* dominate the soil environment surrounding white-rot fungi.

The most abundant family was *Enterobacteriaceae* (54.33%), followed by *Erwiniaceae* (6.24%) and *Yersiniaceae* (3.83%). Other families represented less than 3%. This finding aligns with previous research by Degelmann *et al.* (2009), who suggested that *Enterobacteriaceae*, as facultative anaerobic bacteria, participate in the anaerobic degradation of mono-

saccharides in forest soils, where competition for nutrients is high. At the genus level, 627 ORFs (36.90%) were identified, with *Enterobacter* (7.88%) and *Kluyvera* (6.18%) being the most common genera. At the species level, only 3 identified ORFs (0.18%) were classified, including *Enterobacter cancerogenus*, *Enterobacter cloacae*, and *Cedecea davisae*. The low number of identified species reflects the limitations of current databases (Table 1 and Figure 1).

Figure 1. Classification of ORFs into phylum, class and order level.

Analysis of functional domains and spatial structure of ORFs

Proteins typically consist of one or more functional regions known as domains, which are critical for understanding their function (Mistry *et al.*, 2021). To further explore the functions of enzymes, the functional domains of ORFs encoding 6-phospho-β-glucosidase were analyzed using the Pfam and HMMER databases. Of the 1718 predicted ORFs, 213 (12.40%) were complete genes, while 454 ORFs were missing the 3' end, 397 ORFs lacked the 5' end, and the remaining 654 ORFs lacked both ends. The complete ORFs encoding 6-phospho-β-glucosidase were further analyzed for their functional domains. Of the 213 complete ORFs, 210 contained functional domains: 152 ORFs (71.36%) had the GH1 domain, and 58 ORFs (27.23%) had the GH4 domain. The GH1 domains were 449–477 amino acids in length, interspersed with gaps of 3–10 amino acids, while GH4 domains were smaller, approximately 170 amino acids, with gaps of about 270 amino

acids. These findings are consistent with previous research showing that 6-phospho-β-glucosidase enzymes in bacteria typically belong to the GH1 family (Veldman *et al.,* 2020).

Combining the domain analysis results with taxonomic classification revealed that the majority of ORFs with GH1 and GH4 domains belonged to the Proteobacteria and Firmicutes phyla. Among the ORFs with the GH1 domain, 129 (84.87%) were from Proteobacteria, which was 8.6 times greater than the number from Firmicutes. Similarly, among the ORFs with the GH4 domain, 94.83% were from Proteobacteria, which was 18.34 times higher than that from Firmicutes. These findings highlight that the majority of genes encoding 6-phospho-β-glucosidase with the GH1 domain are derived from Proteobacteria (representative genes with GH1 and GH4 domains from different phyla are shown in Figure 2).

Since the structure of proteins tends to be more conserved than their amino acid sequences during evolution, the spatial structure of the five representative genes was predicted using Phyre2 software. The results showed that the secondary structure of the five genes consisted of 28-39% alpha helices, 12-14% beta strands, and 2-4% disordered regions. The spatial structure models of the five identified genes were based on the published 6-phospho-β-glucosidase template, with a coverage range of 88-95% and a confidence level of 100% (Table 2 and Figure 3).

Figure 2. Diagram showing the functional and phyla domains of representative proteins. (A) Proteins with the GH1 domain. (B) Proteins with the GH4 domain.

Table 2. Detailed template information of spatial structure model of the five proteins.

ORFs	6-phospho-β-glucosidase template
GL0494307	c6wgdB_crystal structure of a 6-phospho-beta-glucosidase from *B. licheniformis*
GL0182588	c6wgdB_crystal structure of a 6-phospho-beta-glucosidase from *B. licheniformis*
GL0335762	c1up7A_structure of the 6-phospho-beta glucosidase from *Thermotoga maritima*
GL0413390	c1up7A_structure of the 6-phospho-beta glucosidase from *T. maritima*
GL0436665	c5c3mB_crystal structure of gan4c, a GH4 6-phospho-glucosidase from *Geobacillus stearothermophilus*

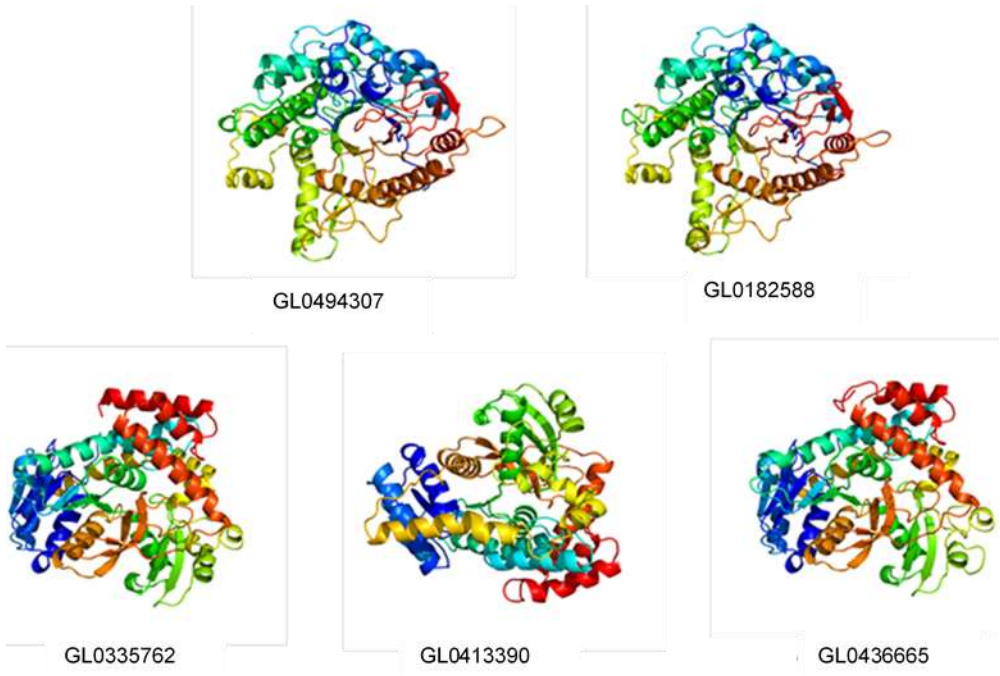

Figure 3. Structural model of the 5 proteins using Phyre2 software.

Prediction of expression levels of genes encoding the enzyme 6-phospho-β-glucosidase

E. coli is widely regarded as the most popular recombinant protein expression system. The expression of soluble proteins not only facilitates the purification of target proteins but also enhances the ability to obtain structurally intact and biologically active proteins. Therefore, to provide a foundation for experimental studies, the expression levels of soluble proteins were predicted using Periscope software.

A total of 210 domain-containing complete ORFs were further evaluated for expression in *E. coli*. ORFs with high expression levels are shown in Table 3. The gene expression analysis revealed that two ORFs containing the GH1 domain, with secondary structures consisting of 28-32% alpha helices, 12% beta strands, and 3% disordered regions, exhibited the highest expression levels, ranging from 904 to 4714 mg/l. Three ORFs containing the GH4 domain, with secondary structures consisting of 37-39% alpha helices, 12-14% beta strands, and 2-4% disordered regions, showed relatively high expression levels above 1000 mg/l. The predicted expression levels of ORFs were consistent with the experimental expression results of 6-phospho-β-glucosidase (Yang *et al.*, 2013; Yu *et al.,* 2013). According to Yang *et al.* (2013), in the presence of IPTG, 6-phosphoglucosidase from Black Liquor Sediment was mainly expressed in the soluble fraction, with a molecular mass consistent with its theoretical molecular mass. Therefore, these predictions can be trusted as a relative ranking of gene expression level.

Table 3. Prediction of expression level of 6-phospho-β-glucosidase proteins in *E. coli.*

No	Number of ORF	Domain	Phylum	Representative ORF	Expression level (mg/l)
1	152	GH1	Firmicutes	GL0494307	4714
			Proteobacteria	GL0182588	904
			Firmicutes	GL0393514	733
			Proteobacteria	GL0535097	649
			Proteobacteria	GL0262858	589
2	58	GH4	Proteobacteria	GL0335762	1549
			Proteobacteria	GL0413390	1092
			Proteobacteria	GL0436665	1078
			Proteobacteria	GL0280854	761
			Proteobacteria	GL0318352	35

Prediction of some physical properties of proteins

For the convenience of experimental studies, some physical properties of the five genes were also predicted. The AcalPred software was used to predict whether the proteins were acidic or alkaline. This free analytical system distinguishes the ability of enzymes to withstand acidic or alkaline environments. AcalPred is based on data from 150,000 different temperature-resistant proteins in the NCBI database, predicting heat tolerance based on similarity principles. After inputting the FASTA amino acid sequences of the five selected genes into AcalPred, the results were obtained as shown in Table 4. According to the predictions of this tool, an enzyme with a predictive index between 0.5 and 1 is considered alkaline, while an enzyme with a predictive index below 0.5 is acidic. The results indicated that all five selected genes encoded alkaline enzymes. This finding is consistent with previous studies, which showed that 6-phospho-β-glucosidase genes function effectively under neutral pH conditions (Yang *et al.*, 2013). Soil microorganisms often operate under very specific pH conditions, and this alkaline pH is suitable for the growth of Proteobacteria, particularly the Gammaproteobacteria class.

Table 4. Results for discriminating between the acidic and alkaline enzymes.

No	ID of genes	Type of enzyme	Probability	
			Acidic enzyme	Alkaline enzyme
1	GL0494307	Alkaline enzyme	0.326	0.674
2	GL0182588	Alkaline enzyme	0.469	0.531
3	GL0335762	Alkaline enzyme	0.151	0.849
4	GL0413390	Alkaline enzyme	0.112	0.888
5	GL0436665	Alkaline enzyme	0.069	0.931

The enzyme's heat resistance, according to TBI, is classified into three levels: if the Tm is greater than 1, the enzyme can withstand temperatures above 65°C; if Tm is between 0 and 1, the enzyme's heat resistance ranges from 55°C to 60°C; and if Tm is less than 0, the heat resistance is below 55°C. The results of the heat tolerance analysis for the five genes are shown in Table 5. The findings indicated that the optimal temperature for enzyme activity was relatively high, with four out of the five selected genes exhibiting heat resistance above 65°C. These results not only demonstrate the potential of these enzymes for industrial applications but also assist in determining the appropriate temperature and pH conditions for experimental research. According to Yang *et al.*, (2013) the optimal temperature for a novel metagenome-derived 6-phospho-β-glucosidase from black liquor sediment was found to be 37°C.

Table 5. Results for melting temperature prediction.

No	ID of genes	Tm	Temperature
1	GL0494307	1.044	Higher than 65°C
2	GL0182588	2.010	Higher than 65°C
3	GL0335762	0.964	55°C-65°C
4	GL0413390	1.008	Higher than 65°C
5	GL0436665	1.275	Higher than 65°C

CONCLUSION

We mined 1718 ORFs encoding 6-phospho-β-glucosidase from bacterial DNA metagenome data surrounding white-rot fungi based on the KEGG database. A total of 1699 ORFs (99.94%) were classified as bacterial, belonging to three common phyla: Proteobacteria, the most dominant phylum, with 1494 ORFs (87.93%), which was 11.85 times more abundant than the second phylum, Firmicutes (7.42%), and Actinobacteria (0.77%). The most prevalent class was Gammaproteobacteria, which accounted for 1472 ORFs (86.64%). Domain prediction revealed 152 ORFs encoding the GH1 domain (84.87% of Proteobacteria) and 58 ORFs with the GH4 domain (94.83% of Proteobacteria). The secondary structure of the representative proteins consisted of 28-39% alpha helices, 12-14% beta strands, and 2-4% disordered regions. The spatial structure model of the five identified proteins was based on the published 6-phospho-β-glucosidase template, with coverage ranging from 88% to 95% and a confidence level of 100%. Proteins containing the GH4 domain exhibited secondary structures with 37-39% alpha helices, 12-14% beta strands, and 2-4% disordered regions, and showed relatively high expression levels above 1000 mg/l. The 6-phospho-β-glucosidase enzymes were alkaline, and their optimal temperature for activity was relatively high.

ACKNOWLEDGMENTS

This study was supported by the grant from the Bilaterial International Project, code: NĐT.50.GER/18, from Ministry of Science and Technology (MOST), Vietnam and Federal Ministry of Education and Research, Germany; using the facilities in National Key Laboratory of Gene Technology,

Institute of Biotechnology, Vietnam Academy of Science and Technology (VAST), Vietnam.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Degelmann D. M., Kolb S., Dumont M., Murrell J. C., and Drake H. L. (2009). Enterobacteriaceae facilitate the anaerobic degradation of glucose by a forest soil. *FEMS Microbiology Ecology, 68*(3), 312–319. http://doi.org/10.1111/j.1574-6941.2009.00681.x

Deutscher J., Francke C., and Postma P.W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews, 70*(4), 939-1031. http://doi.org/10.1128/MMBR.00024-06

Franco D. C., Signori C. N., Duarte R. T., Nakayama C. R., Campos L. S., and Pellizari V. H. (2017). High prevalence of Gammaproteobacteria in the sediments of Admiralty Bay and north Bransfield Basin, Northwestern Antarctic Peninsula. *Frontiers in Microbiology, 8,* 153. http://doi.org/10.3389/fmicb.2017.00153.

Glushakova A. M., Kachalkin A. V., Prokof'eva T. V., and Lysak L. V. (2022). Enterobacteriaceae in soils and atmospheric dust aerosol accumulations of Moscow city. *Current Research in Microbial Sciences, 3,* 100-124. http://doi.org/10.1016/j.crmicr.2022.100124.

Haq I. U., Hillmann B., Moran M., Willard S., Knights D., Fixen K. R., *et al.* (2022). Bacterial communities associated with wood rot fungi that use distinct decomposition mechanisms. *ISME Communications, 2*(1), 26. http://doi.org/10.1038/s43705-022-00108-5.

Huson D. H., Auch A.F., Qi J., and Schuster S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research, 17*(3), 377-386. http://doi.org/10.1101/gr.5969107.

Janusz G., Pawlik A., Sulej J., Swiderska-Burek U., Jarosz-Wilkolazka A., and Paszczynski A. (2017). Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiology Reviews, 41*(6), 941-962. http://doi.org/10.1093/femsre/fux049.

Folman L. B., Klein Gunnewiek P. J., Boddy L., and de Boer W. (2008). Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. *FEMS Microbiology Ecology, 63*(2), 181-191. http://doi.org/10.1111/j.1574-6941.2007.00425.x.

Le T. T. H., Nguyen T. B., Nguyen H. D., Nguyen H. D., Le N. G., Dao T. K., *et al.* (2022). De novo metagenomic analysis of microbial community contributing in lignocellulose degradation in humus samples harvested from cuc phuong tropical forest in vietnam. *Diversity, 14*(3), 220. https://doi.org/10.3390/d14030220

Mhete M., Eze P. N., Rahube T. O., and Akinyemi F. O. (2020). Soil properties influence bacterial abundance and diversity under different land-use regimes in semi-arid environments. *Scientific African, 7*(e00678), https://doi.org/10.1016/j. sciaf.2019.e00246.

Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G. A., Sonnhammer E. L. L, *el al.* (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*(D1), D412-D419. http://doi.org/10.1093/nar/gkaa913.

Monteiro L. M. O., Vici A. C., Pinheiro M. P., Heinen P. R., de Oliveira A. H. C., Ward R. J., *et al.* (2020). A highly glucose tolerant ß-glucosidase from *Malbranchea pulchella* (MpBg3) enables cellulose saccharification. *Scientific Reports, 10*(1), 6998. http://doi.org/10.1038/s41598-020-63972-y.

Pang Z., Dong F., Liu Q., Lin W., Hu C., and Yuan Z. (2021). Soil metagenomics reveals effects of continuous sugarcane cropping on the structure and functional pathway of rhizospheric microbial community. *Frontiers in*

Microbiology, 12, 627569. http://doi.org/10.3389/fmicb.2021.627569.

Pinar O., and Rodríguez-Couto S. (2024). Biologically active secondary metabolites from white-rot fungi. *Frontiers in Chemistry, 12,* 1363354. http://doi.org/10.3389/fchem.2024.1363354.

Pusztahelyi T., Holb I. J., and Pócsi I. (2015). Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science, 6*, 573. http://doi.org/10.3389/fpls.2015.00573.

Veldman W., Liberato M. V., Almeida V. M., Souza V. P., Frutuoso M. A., Marana S. R., *et al.* (2020). X-ray structure, bioinformatics analysis, and substrate specificity of a 6-phospho-β-glucosidase glycoside hydrolase 1 enzyme from *Bacillus licheniformis*. *Journal of Chemical Information and Modeling, 60*(12), 6392-6407. http://doi.org/10.1021/acs.jcim.0c00759.

Yu W. L., Jiang Y. L., Pikis A., Cheng W., Bai X. H., Ren Y. M., *et al.* (2013). Structural insights into the substrate specificity of a6-phospho-β-glucosidase BglA-2 from Streptococcus pneumoniae TIGR4. *Journal of Biological Chemistry, 288*(21), 14949-14958. http://doi.org/10.1074/jbc.M113.454751.

Yang C., Niu Y., Li C., Zhu D., Wang W., Liu X., *et al.* (2013). Characterization of a novel metagenome-derived 6-phospho-β-glucosidase from black liquor sediment. *Applied and Environmental Microbiology, 79*(7), 2121-2127. http://doi.org/10.1128/aem.03528-12.

Zhang L., Fu Q., Li W., Wang B., Yin X., Liu S., *et al.* (2017). Identification and characterization of a novel β-glucosidase via metagenomic analysis of *Bursaphelenchus xylophilus* and its microbial flora. *Scientific Reports, 7*(1), 14850. http://doi.org/10.1038/s41598-017-14073-w.

Zhang Q., Zhang Z., Lu T., Yu Y., Penuelas J., Zhu Y. G., *et al.* (2021). Gammaproteobacteria, a core taxon in the guts of soil fauna, are potential responders to environmental concentrations of soil pollutants. *Microbiome, 9*(1), 196. http://doi.org/10.1186/s40168-021-01150-6.