1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

A Novel Metagenome-Derived Gene Cluster from Termite Hindgut:
Encoding Phosphotransferase System Components and High Glucose
Tolerant Glucosidase.

Gang Gao^{§,1}, An Wang^{§,1}, Bo-Liang Gong¹, Qing-Qing Li¹, Yu-Huan Liu¹, Zhu-Mei He^{*,1},
Gang Li^{*,1}

[§]These authors contributed equally to this work.
¹School of life sciences, Sun Yat-sen Univeristy, Guangzhou 510275, P. R. China
*corresponding author: lsshezm@mail.sysu.edu.cn; lsslig@mail.sysu.edu.cn

28

29 **Abstract**

30 Functional screening of a metagenomic library of termite hindgut identified an
31 overlapping gene cluster encoding the phosphotransferase system (PTS) components,
32 which consisted of a glucoside specific PTS enzyme II gene (*glu1923*) and a
33 glycoside hydrolase gene (*glu1392*). Hydrolytic experiments revealed that the
34 combined effect of Glu1923 and Glu1392 was responsible for the utilization of
35 glucosidic substrates in recombinant *Escherichia coli* (*E. coli*) strains. Yeast two
36 hybrid analysis suggested that there was an interaction between Glu1923 and Glu1392,
37 and the domain EIIA of Glu1923 played an important role for the interaction. In
38 addition, the hydrolase Glu1392 displayed hydrolysis ability toward cellobiose and
39 maltose, and had a very high tolerance to glucose with a K_i value of 2.25 M. These
40 properties make Glu1392 an interesting candidate in biotechnology applications after
41 further study.

42

43 **Keywords:** *Coptotermes formosanus* • Metagenomic library • Phosphotransferase
44 system • Glucosidase • Glucose tolerance

45

46 **Introduction**

47 In bacteria, glucoside sugar is most commonly taken up by the phosphoenolpyruvate
48 (PEP)-sugar phosphotransferase system (PTS) [1], and the utilization of glucoside
49 sugar can convert oligosaccharide into glucose and providing the carbon sources [2,

3]. The PTS, which consists of three key components including enzyme I (EI), heat-stable phosphocarrier protein (HPr), and enzyme II (EII), plays an important role in carbohydrates transport [4]. EI firstly autophosphorylates from PEP and transfers P to HPr protein, and then the phosphoryl group is delivered to the carbohydrates via EII. Finally, the phosphorylated carbohydrate was translocated into cytoplasm for specific PTS-dependent utilization [5]. To date, numbers of PTS genes/operons were identified by computational and functional analyses [6-8]. Further studies indicated that EI and HPr from different organisms were evolutionarily conserved [9]. However, the permease EII had substrate specificity, and **only a few** EIIs with unique structures or functions were identified until now. In addition, almost all of the PTS genes/operons reported so far were derived from cultured microorganisms, and little attention had been paid to discover novel PTS genes/operons from unculturable microorganisms, which account for over 99 % of microorganisms in the environment. Therefore, some efforts should be made to discover novel PTS genes/operons. The metagenomic approach, **featured by** direct cloning of DNA from environment samples and thereby accessing the potential of unculturable organisms, was a powerful tool for the isolation of novel PTS genes/operons. Also, the study on novel PTS genes/operons from unculturable microorganisms will help us to better understand how bacteria absorb and utilize sugar substrates.

Previous studies had isolated and analyzed many bacteria β -glucoside operons **that** regulated the utilization of β -glucosides [10,11]. In enteric bacteria *E. coli*, glucoside PTS operons contain all of the genetic elements for the uptake and utilization of

glucosides, but the operons are cryptic and transcriptionally silent in nature [9]. Thus, the wild-type strains are normally unable to utilize the glucoside sugar as a carbon source. The modulation of β -glucoside (such as arbutin and salicin) PTS operon *gluGFB* have been extensively studied, and the result showed that an adaptive mutation by the insertion of an insertion sequence (IS) to the regulation region can activate these cryptic PTS genes [12]. Recently, another *E. coli* strain with β -glucoside sugar (cellobiose) utilization phenotype was also produced by mutations in cryptic operons *chb* and *asc*, or by replacing the cryptic promoters of these endogenous operons with a constitutive promoter [13].

Termites are very efficient decomposers of lignocellulosic plant materials and their high digestion capacity mainly results from the hindgut-associated obligate symbionts [14]. Therefore, these symbiotic microorganisms in the hindgut may contain rich enzymes and PTS genes/operons related with cellulose degradation. To discover novel PTS components from unculturable microorganisms of termite hindgut and give a deep insight into the regulatory mechanism of the bacterial PTS operons, we here identified a novel PTS gene cluster encoding two PTS components from metagenomic library of hindguts inhabiting microbes of lower termite, *Coptotermes formosanus*. Also, we investigated the interaction of two proteins encoded by two glucoside PTS components. Finally, the recombinant glucosidase Glu1392 with high glucose tolerance was characterized after being overexpressed in *E. coli* and purified.

Materials and Methods

Strains, vectors and chemicals

E. coli DH5 α was used as the cloning host, and BL21 (DE3) was served as the expression host. The vector pET-32a (+) was used for protein expression. *Saccharomyces cerevisiae* AH109 strains, pGBKT7 and pGADT7 vectors were used for yeast two hybrid assay. All restriction endonucleases and ligases were purchased from TakaRa (Dalian, China). All other chemical and reagents were from Sangon (Shanghai, China) unless indicated otherwise.

Construction of metagenome library and screening of glucosidase

Wood-feeding *Coptotermes formosanus* termites were collected from Dinghu Mountain, Guangdong province of China (22°10'26" N, 112°31'15" E). Approximately 500 worker termites were dissected and the aqueous content of the hindgut were collected as previously reported [15]. A total genomic DNA from the hindgut was isolated using E.Z.N.A.TM Bacterial DNA isolation kit (Omega). Then, the genomic DNA was partially digested with *Eco*RI and the DNA fragments of 2-8 kb were purified by Omega Gel Extraction Kit (Omega) and inserted into pUC118 vector (TaKaRa), which had been previously digested with *Eco*RI and dephosphorylated with calf intestine alkaline phosphatase. Next, the recombinant plasmid was transformed into *E. coli* DH5 α and the transformants were cultured on lysogeny broth (LB) agar plate containing 0.75 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG), 0.1 % esculin hydrate, 0.25 % of ferricammonium citrate and 100 μ g/ml ampicillin at 37 °C overnight. Those colonies forming clear black halos were selected as positive colonies that have glucosidase

activity [16] and DNA inserts in plasmids were sequenced on ABI 377 DNA sequencer.

Sequence and structure analysis of the overlapping gene cluster

The open reading frames (ORFs) of DNA sequences were identified using ORF-finder program (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The predicted ORFs were submitted to the NCBI website (<http://blast.ncbi.nlm.nih.gov>) for similarity analysis and homology search. Structure-based sequence alignment was performed by using CLUSTAL W (www.phylogeny.fr) [17]. Protein sequence analyses were performed at the Expasy proteomics server (<http://www.expasy.org/>) [18].

Expression of the overlapping gene cluster and individual genes and purification of recombinant proteins

The individual gene *glu1923*, *glu1392*, and the overlapping gene cluster *glu3311* were both amplified via PCR and the reaction conditions were: 98 °C, 5 mins; 98 °C, 10 sec; 55 °C, 10 sec; 72 °C, 15 sec, repeated for 30 cycles; 72 °C, 5 mins. The PCR products were digested using *KpnI/EcoRI* (for *Glu1923*), *BamHI/HindIII* (for *Glu1392*), and *KpnI/HindIII* (for *Glu3311*), and then ligated to expression vector pET-32a (+). The recombinant vectors were transformed into *E. coli* BL21 (DE3), and the cells were plated on LB agar containing 100 µg/mL ampicillin. The transformants were grown in a 100-ml flask containing 10 mL LB medium (a final concentration of 100 µg/mL ampicillin) at 37 °C until an OD₆₀₀ reached to 1.0. Next, induction was initiated by adding IPTG to a final concentration of 1.0 mM and the cultures were incubated at 30 °C for 8 h (for *glu1392* and *glu3311* expression) or at 16°C for 12 h

(for *glu1923* expression) with shaking at 200 rpm. To purify expression products of *glu1392* and *glu3311*, cells were collected by centrifugation (8,000 g for 10 min at 4 °C). Finally, purification of the recombinant proteins was performed using a His·Bind Purification Kit (Novagen) according to the product manual. To purify recombinant Glu1923, the induced cells were resuspended in PBS buffer (NaCl 137 mM, KCl 2.7 mM, Na₂HPO₄ 10 mM, KH₂PO₄ 1.76 mM, pH 7.4) with 0.1% n-dodecyl-β-D-maltopyranoside (DDM) for protein solubilization. Then, the cells were harvested by centrifugation (8,000 g for 10 min at 4 °C) and sonicated with EDTA-lysozyme treatment. Ultimately, purification of the recombinant proteins was performed using a His·Bind Purification Kit (Novagen) according to the product manual.

SDS-PAGE and western blotting analyses of target proteins

Molecular mass of the denatured proteins was determined by SDS-PAGE. Proteins were visualized following Coomassie brilliant blue G-250 staining. The products of the intact overlapping gene *glu3111* were also analyzed by Native-PAGE. In addition, western blot analysis of target proteins was performed according to the standard protocol. Briefly, the recombinant proteins were separated on a 12 % SDS-PAGE gel, which was semi-dry transferred at 15 V for 30 min to 0.45 mm PVDF membrane (Bio-Rad, USA), immunoblotted with anti-His Tag mouse Monoclonal antibody (dilution to 1:5000) (BOSTER, China). Next, the IgG goat anti-mouse antibody conjugated with HRP (Horseradish peroxidase) was used as a secondary antibody (dilution to 1:5000), with a DAB (Diaminobenzidine) Kit used for visualization of the

protein band.

Hydrolysis activity of the recombinant proteins towards different substrates

The recombinant *E.coli* strains expressing *glu1923*, *glu1392*, and the overlapping gene cluster (*glu3311*) respectively were plated on LB agar plate containing 1.0 mM IPTG, 0.1 % esculin hydrate, 0.25 % of ferric ammonium citrate and 100 µg/ml ampicillin to test their hydrolysis activity according to the existence of black halo around corresponding colonies **or not**. Simultaneously, the recombinant expression vectors carrying different antibiotic resistance including pET-32a-glu1392 (with Ampicillin resistance) and pET-28a-glu1923 (with kanamycin resistance), were co-transformed to *E. coli* BL21 (DE3). Then, the co-transformants were also streaked on esculin plate for testing hydrolysis activity *in vivo*. In addition, abovementioned four recombinant strains were cultured on M9 minimal medium containing cellobiose, maltose, or trehalose for hydrolysis activity assay. M9 minimal medium was prepared according to the protocol described by Charusanti et al [19].

Interaction assay *in vivo* between Glu1392 and Glu1923 by yeast two hybrid method

To test **the** interactions between Glu1392 and Glu1923, the bait vector pGBKT7 containing *glu1392* and **the** prey vector pGADT7 inserted *glu1923*, were transferred into auxotrophic yeast strain AH109 by LiAc method according to the instruction manual of Clontech (Yeast Protocols Handbook, Protocol No. PT3024-1). Then, the positive transformants were selected from a nutritionally selective plate deficient in tryptophan and leucine (SD/-Leu-Trp) and streaked on SD/-Trp/-Leu/-His/-Ade plates

to further detect the protein interaction. The plates were incubated for 2-3 days at 30 °C for β -galactosidase activity assay via colony lift method [20]. In addition, to find out which domain of Glu1923 involved in the interaction with Glu1392, three deletion mutants containing different domains of Glu1923 were constructed by PCR and then ligated to the prey vector pGADT7, respectively. The first deletion mutant have the PTS-IIA domain deleted, this deletion mutant consisted of the first 464 residues at the N-terminal region of Glu1923. The second deletion mutant contained the PTS-IIC domain and was composed of 125–464 residues at the central region of Glu1923. The third deletion mutant included the PTS-IIC&IIA domain of Glu1923, which contained 125–641 residues at the C-terminal region of Glu1923. The prey vectors containing three deletion mutants were respectively transformed to auxotrophic yeast strain AH109 that harbouring bait vectors pGBKT7-*glu1392* for the interaction assay.

Enzymatic activity and substrate specificity assay

The enzymatic activity of Glu1392, Glu1923, and the mixture of Glu1392 and Glu1923 toward aryl-glucoside substrate p-nitrophenyl- β -D-glucopyranoside (pNPG, Sigma) were measured by following p-nitrophenol (pNP) release at 405 nm in a 96-well microtiter format based method [21]. Each enzyme solution contained 150 ng of purified enzyme, and the mixed enzyme solution was composed of purified Glu1392 (150 ng) and Glu1923 (150 ng). The 5 μ l (30 ug/ml) of enzyme solution was mixed with 0.1 mL of 5 mM pNPG solution in 0.1 M citrate phosphate buffer at pH 8.0. After incubation at 50°C for 20 min, the reaction was subsequently terminated by

adding 200 μ l of 1.0 M Na_2CO_3 . The pNP released from the mixture was measured in a microplate reader (Thermo Scientific Inc., U.S.A.). One unit of enzyme activity toward pNPG was defined as the amount of enzyme needed to produce 1 μ mol of pNP per minute under the assay condition. Each experiment was repeated for three times with duplicate replicates, and Duncan's test ($p < 0.05$, ANOVA) was used for significant difference analysis. Substrate specificity of the enzymes toward natural substrates including oligosaccharide cellobiose, maltose and trehalose was determined by incubating the purified enzyme with respective substrates (5 mM) in 0.1 M citrate phosphate buffer at 50 °C for 30 min and measuring the released glucose by the standard procedures [22].

Effects of pH and temperature on the activity and stability of Glu1392

The optimum pH of Glu1392 was determined using pNPG as substrate at 50 °C. The following buffers (100 mM) were used for the pH assays: glycine-HCl (pH 2.0-5.0), sodium acetate (pH 5.0-6.0), sodium phosphate (pH 6.0-8.0), Tris-HCl (pH 8.0-9.0), and glycine-NaOH (pH 9.0-11.0). Further study on the pH stability of Glu1392 was carried out by incubating at various pH systems (2-10) at 4 °C for 1 h, and the residual activity was measured according to the standard method mentioned above. The optimum temperature for the enzyme was determined by varying the temperature from 20 °C to 90 °C at pH 8.0. The thermostability was investigated by incubating Glu1392 alone in sodium phosphate at 20 - 90 °C for 1 h. After cooling on ice for 30 min, the residual activity of the reaction mixture was assayed under standard conditions.

Susceptibility of Glu1392 toward metal ions and chemical reagents

To determine the susceptibility of Glu1392 toward metal ions and chemical reagents, the purified enzyme was incubated with various metal ions (10 mM) and chemical reagents (1% or 10 %) in citrate phosphate buffer for 30 min at 50 °C. The corresponding enzyme activity in the presence of the metal ions and chemical reagents was measured under standard conditions using pNPG as substrate. The amount of enzymatic activity was calculated as a percentage of the activity comparing to that of the control.

Kinetic assays of Glu1392

The kinetic parameters of Glu1392 were obtained by measuring the rate of hydrolysis of pNPG and cellobiose at various concentrations (1–10 mM) in 0.1 M citrate phosphate buffer at 50 °C for 30 min. The enzymatic kinetic parameters, K_m and k_{cat} , were determined from the Lineweaver-Burk plots using the computer program.

Effect of glucose on enzyme activity

The inhibitory effect of glucose on enzyme activity was investigated according to the method reported by Uchima et al [23]. In brief, 24 µl of various concentrations of glucose solutions (0.2 M, 0.5 M, 0.8 M, 1.0 M, 1.5 M, 2.5 M, 3.0 M, 3.5 M) were pre-incubated with 1 µl of enzyme solution for 1 h at room temperature, respectively. Next, each pre-incubated solution was mixed with 25 µl of pNPG (5 mM and 7 mM) and 50 µl of citrate phosphate buffer (100 mM, pH 8.0), which contained the same concentration of glucose. The residual enzyme activity of the reaction mixture was measured under standard conditions. The relative activity was defined as the relative

value to the activity of the control without glucose.

Nucleotide sequence accession number

The nucleotide sequence data of the overlapping gene cluster have been submitted to the nucleotide sequence databases (GenBank) under accession number KM624528.

Results

Construction of metagenomic library and screening of glucosidase positive clones

A metagenomic library containing about 800,000 clones was constructed, and restriction analysis of 30 randomly picked clones revealed that the average DNA insert size was approximately 5.2 kb. Thus, the generated metagenomic library represented about 4.16 Gb of cloned DNA from the termite hindgut symbionts. Thirteen positive clones with clear black halo on esculin agar plate were identified from the metagenomic library, and one of them was finally selected for further study due to its high activity.

Sequence analysis of positive clone

Sequence analysis of the positive clone revealed the presence of two complete ORFs (named *glu1923* and *glu1392*), which encoded a protein (640 amino acids) of PtsG superfamily and a glucosidase (463 amino acids) of glycoside hydrolase family 1 (GH family 1), respectively. Moreover, Glu1923 showed the high identity of 68 % with glucoside-specific IIABC component of the PTS system from a bacterium enrichment culture clone P75-7G (accession no, HQ601969.1), and Glu1392 displayed 70 % identity with the β -glucosidase from *Reticulitermes flavipes* termite (Accession:

ADK12988.1). Also, the arrangement mode of *glu1923* and *glu1392* in the DNA insert of the positive clone accorded with that of the typical glucose specific PTS operon in *E. coli* (Figure 1A & Figure 1B) [24, 25]. However, their actual functions still needed to be further investigated. Interestingly, there was a four-nucleotide overlap between *glu1923* and *glu1392* (Figure 1B), which was never found in the PTS system. In addition, multiple sequence alignment of Glu1923 and Glu1392 with several of their closest homologues was performed, and Glu1392 and Glu1923 exhibited high identity with PTS associated glucosidase and EII permease, respectively, and their active site residues were also identified. These results preliminarily confirmed the function of Glu1923 and Glu1392.

Heterologous expression and purification of recombinant protein

The molecular mass of the induced proteins were determined by SDS-PAGE and Western blot analyses (Figure 2A & 2B). The results of Figure 2A revealed that the molecular mass of the recombinant Glu1392 and Glu1923 was about 65 kDa and 85 kDa, which agreed with their predicted size plus the N-terminal fusion protein of the expression vector (18 kDa). Also, the results of Western blot assay confirmed that *glu1392* and *glu1923* were successfully expressed in *E. coli* (Figure 2B). To investigate the biologically active form of the overlapping gene cluster *glu3311*, the Native-PAGE and Western blot analysis of *glu3311* was performed (Figure 2C). Through this separation, we tried to know whether *glu3311* expressed two individual monomeric proteins (Glu1392 and Glu1923) or a multimeric protein consisting of two monomeric proteins. The results of Figure 2C revealed that *glu3311* expressed two

individual monomeric proteins, and their sizes were consistent with those of Glu1392 and Glu1923. Therefore, the biologically active form of the overlapping gene cluster *glu3311* was two individual monomeric proteins instead of a multimeric protein.

Hydrolytic activity analysis of glucoside PTS components in *E. coli*

To test the hydrolytic activity of glucoside PTS components towards esculin hydrate in *E. coli*, the recombinant strains expressing *glu1392*, *glu1923*, and *glu3311* were cultured on LB screening plates, respectively. The results indicated that individual Glu1392 or Glu1923 was unable to hydrolyze esculin hydrate and thus no dark halo was formed on the plate. However, the recombinant strain expressing *glu3311* or co-expressing *glu1392* and *glu1923* displayed high hydrolytic activity towards esculin hydrate along with the formation of clear dark halo (Figure 3A), which suggested that the combined action of Glu1923 and Glu1392 was responsible for the utilization of glucosidic substrate esculin hydrate in *E. coli* cells. In addition, further studies demonstrated that the recombinant strain expressing *glu3311* or co-expressing *glu1392* and *glu1923* grew much faster on the M9 medium containing cellobiose or maltose as the sole carbon source than the wild type BL21 strain, or the recombinant strain expressing *glu1392* or *glu1923*, individually. The former formed the colonies on the plate only after two days, while the latter needed at least 1 week to form the colonies on the plate (Figure 3B & Figure 3C). Nevertheless, both of them could not grow on the M9 minimal medium containing trehalose as the sole carbon source (Figure 3D). These results showed that the glucoside PTS components composed of Glu1923 and Glu1392 could effectively utilize cellobiose (β -1,4 linkage) and maltose

(α -1,4 linkage) as a carbon source with the help of other PTS components (such as EI and HPr), but they could not hydrolyze trehalose (α -1,1 linkage).

Protein interaction analysis by yeast two hybrid method

In order to explore whether there was an interaction between Glu1923 and Glu1392, yeast two hybrid analysis was carried out. The results indicated that *S. cerevisiae* AH109 cells expressing two intact PTS components (*glu1923* and *glu1392*) could produce distinct blue color (Figure 4A-1), which suggested there was possibly an interaction between Glu1923 and Glu1392. Moreover, the conserved domain search revealed that *Glu1923* consisted of three domains including PTS-EIIB, PTS-EIIC, and PTS-EIIA (Figure 4B). Further studies showed that the Glu1923 mutant with PTS-EIIB domain deletion still had interaction with Glu1392 (Figure 4A-4). However, the Glu1923 mutant with PTS-EIIA domain deletion and the Glu1923 mutant with PTS-EIIA and PTS-EIIB domain deletion could not interact with Glu1392 (Figure 4A-2,3). Therefore, the domain PTS-EIIA of Glu1923 probably played an important role for the interaction of Glu1923 and Glu1392.

Enzymatic activity and substrate specificity of recombinant enzyme

Glu1392 yielded a high activity of 22 ± 1.7 U/mg for pNPG, and Glu1923 had almost no activity for pNPG. In addition, the substrate specificity of Glu1923, Glu1392 and the mixture of Glu1923 and Glu1392 was tested using three disaccharide glycoside substrates, including cellobiose, maltose and trehalose (Figure 5). Glu1392 displayed the highest hydrolysis activity towards cellobiose (β -1,4 linkage), high hydrolysis activity (92 %) towards maltose (α -1,4 linkage), and no activity for trehalose (α -1,1

linkage). However, the optimum substrate for Glu1392 was probably phosphorylated disaccharides instead of unphosphorylated disaccharides. Because Glu1392 exhibited 62% sequence identity with BglH of *Bacillus subtilis*, which had high 6-P- β -glucosidase activity [26]. Unfortunately, we were not able to carry out the corresponding experiments because the phosphorylated disaccharides were not commercially available. In addition, the mixture of Glu1392 and Glu1923 showed similar substrate specificity with Glu1392, which also confirmed once more that Glu1923 had no hydrolysis activity towards abovementioned three substrates.

Effects of pH and temperature on the enzymatic activity and stability of the recombinant Glu1392

As shown in Figure 6A, the maximum activity of Glu1392 towards pNPG was obtained at pH 8.0. Moreover, the enzyme was found to be stable in the pH range of 7.0-10.0, and more than 70 % of maximum activity was remained, confirming that Glu1392 had broader alkaline pH activity range. The optimal temperature for the enzyme was 50 °C (Figure 6B). The enzyme was highly stable below 50 °C, and its activity almost kept **unchanged** after incubation for 1 h (Figure 6B). When the temperature was more than 50 °C, the enzyme still had a moderate stability, and its residual activity was more than 60 % after incubation for 1 h for a temperature range of 20-80 °C.

Susceptibility of Glu1392 toward metal ions and chemical reagents

The effect of different metal ions and chemical reagents on the enzyme activity of Glu1392 was investigated by using pNPG as substrate (Table 1). Glu1392 displayed

high tolerance to some metal ions and chemical reagents, such as K^+ , Li^+ , Na^+ , Mg^{2+} , Mn^{2+} , Ca^{2+} , methanol, ethanol, isopropanol and triton X-100. However, the enzymatic activity was evidently inhibited in the presence of 10 mM of Fe^{3+} , Fe^{2+} , Hg^{2+} , Ni^{2+} , Zn^{2+} , Co^{2+} , 1 % SDS, and 1 % EDTA.

Kinetic parameters of recombinant enzyme

To investigate the kinetic parameters of Glu1392, the Michaelis-Menten constants (K_m), turnover numbers (k_{cat}), and catalytic efficiencies of Glu1392 for its synthetic substrate pNPG and its natural substrate cellobiose were determined. The k_{cat} and K_m values of the enzyme were 114.7/s and 290 μ M for pNPG, and 163.2/s and 134 μ M for cellobiose, respectively. The k_{cat}/K_m value of the enzyme for cellobiose was about 3-fold higher than that for pNPG, which clearly demonstrated that the catalytic efficiency of Glu1392 for cellobiose was much higher than that for pNPG.

Inhibitory effect of glucose on Glu1392

Effect of glucose on Glu1392 was investigated with various concentrations of glucose, and results showed that glucose, whose concentration was lower than 1 M, had only a little inhibition to the enzymatic activity, and more than 80 % maximum activity was remained. When glucose concentration was increased to 2 M, the residual activity of Glu1392 was about 60 %. Even when glucose concentration reached up to 3 M, Glu1392 still had 26 % residual activity, which suggested that it had very high tolerance to glucose. Interestingly, glucose could also effectively improve the thermostability of Glu1392. When incubated at 50 °C for 8 h, Glu1392 had only 45.5 % residual activity. However, when 0.5 M or 1.0 M glucose was presented, the residual

activity of Glu1392 reached up to 78 % and 69 % respectively. In addition, the inhibition constant (K_i) of Glu1392 was also investigated according to the intersections of the lines on Dixon plot analysis, and results indicated that the K_i value of Glu1392 was 2.28 M.

Discussion

The metagenomic approach has become an important tool to discover novel lignocellulases from hindgut-associated obligate symbionts of termites in the last decade, and several enzymes related to lignocellulose degradation were successfully identified from termites via metagenomic strategy. These enzymes included feruloyl esterases [27], carboxylesterase [28], hemicellulases [29], β -glucosidases [30]. However, all of these enzymes were isolated from the gut microbiote of the higher termites, and seldom attention was paid to lignocellulases from hindgut-associated obligate symbionts of lower termites. Furthermore, only single gene was obtained in the abovementioned studies. In the present study, a metagenomic library was constructed using hindgut of lower termite *C. formosanus*, and functional screening resulted in the identification of a novel overlapping gene cluster encoding PTS components involved in utilization of synthetic glucoside sugar (aesculin, pNPG) and natural disaccharide (cellobiose and maltose). To the best of our knowledge, this is the first report on the gene cluster encoding PTS components obtained from hindgut of lower termite *C. formosanus* via metagenomic strategy up to now. Therefore, this study isolated a novel gene cluster from uncultured microbes using function-based

metagenomic approach, and the study on the PTS-related gene cluster may bring up new knowledge to better understand the carbohydrate uptaking in microbials.

Here, the structure feature of the novel PTS gene cluster accorded with that of other microorganisms. However, there was a four-nucleotide overlap between two PTS components *glu1923* and *glu1392* (Figure 1A). These overlapping PTS operon components from uncultural microorganisms were never reported previously, and their regulation mechanism and interaction are worth further study. Moreover, the wild-type *E. coli* strains are normally unable to utilize the glucoside sugar as a carbon source because its glucoside PTS operons are cryptic and transcriptionally silent in nature [8]. To active the expression of some cryptic PTS genes in *E. coli*, the insertion of an IS to the regulation region was needed [31]. Also, mutations of cryptic operons or replacement of the cryptic promoters with a constitutive promoter were feasible methods[32]. In this study, glucoside sugar utilization phenotype was obtained by co-expressing of the two PTS components (*glu1392* and *glu1923*) instead of conventional mutation methods. The EII sugar permeases has been experimentally verified to be localized in the cytoplasmic membrane and typically have three subdomains. The domains of EIIA and EIIB were hydrophilic, while EIIC was hydrophobic and formed the sugar translocation channel [33]. Conserved domain analysis convinced that Glu1923 contained the three typical functional domains. Also, Yeast two hybrid analysis revealed a clear interaction between EII (Glu1923) and glycosyl hydrolase (Glu1392), which is not previously reported in the PTS system. Further studies indentified the domain EIIA of Glu1923 played an important role for

the interaction of the two proteins. This metagenome-derived novel PTS cluster we found here have a special overlapping gene structure, and the interaction between their encoding product may play a new insight to understand the relationships and interactions of the PTS components.

In addition, Glu1392 displayed hydrolysis ability toward cellobiose and maltose, and its maximal activity was obtained at pH 8.0, which was very different from those of most other β -glucosidases of termites, such as two β -glucosidases from the termite *Coptotermes formosanus* (pH optimum was 5.0) [34], β -glucosidase from the termite *Nasutitermes takasagonensis* (pH optimum was 6.0) [35], as well as β -glucosidase from a metagenomic library of termite *Globitermes sulphureus* gut (pH optimum was 6.0) [36]. The enzyme was relatively stable under a broad pH range (pH 5~10), showing good pH stability compared to other termite glucosidases, including a β -glucosidase (CfGlu1B) from the termite *Coptotermes formosanus* (enzyme lost its activity at pH 7.0) [34], β -glucosidase from the termite *Nasutitermes takasagonensis* (enzymatic activity was stable from pH 5.0 to 7.0) [35], and β -glucosidase from a metagenomic library of termite *Globitermes sulphureus* gut (more than 70% activity was observed from pH 5.5 to 8.5) [36]. Moreover, the optimum temperature of Glu1392 was 50 °C, and the enzyme was highly stable below 50 °C and moderately stable for a temperature range of 20 °C–80 °C. Because most applications of β -glucosidases require high optimal temperature (50 °C or above), interests in identifying novel β -glucosidases with high optimal temperature and good thermostability have been increasing in the last decade. Recently, several thermostable

β -glucosidases have been reported [36, 37, 38], and these β -glucosidases had higher temperature optimum (65-90 °C) and better thermostability than Glu1392. However, these enzymes had very low glucose tolerance with a K_i value of 0.60 mM [37] and 8 mM **respectively** [38], which limited to their wide application in biotechnology. Comparing to most of the other microbial β -glucosidaes (K_i values range from 0.35 to 100 mM), Glu1392 showed very high tolerance to glucose with a K_i value of 2.25 M. The high tolerance to glucose makes Glu1392 **having** great advantage in the hydrolysis of cellulosic biomass. During the process for the production of ethanol from cellulose, β -glucosidase is often inhibited by the end product glucose, and thus a β -glucosidase with high tolerance to glucose has additional advantages in conversion of cellulosic materials to fermentable sugars through relieving the inhibition of endoglucanase and cellobiohydrolase caused by the accumulation of cellobiose during the bioconversion process [39]. Furthermore, Glu1392 have a broad substrate specificity, this feature may **make it to** be used to degrade the plant cell wall more efficiently [40]. In view of **the** abovementioned enzymatic properties of Glu1392, this enzyme has good application potential for the production of ethanol from cellulose after further improving its temperature optimum and thermostability.

In conclusion, we here identified an overlapped gene cluster encoding novel PTS components from metagenomic library of guts inhabiting microbes of lower termite by function-based metagenomic screening strategy. This gene cluster contained a glucoside specific PTS enzyme II gene (*glu1923*) and a glycosyl hydrolase gene (*glu1392*). The collaboration of Glu1923 and Glu1329 resulted in utilization of

glucoside sugar in *E. coli*. Furthermore, the domain EIIA of Glu1923 plays an important role for the interaction of Glu1923 and Glu1392. In addition, the recombinant Glu1392 displayed a very high tolerance to glucose. These features make it a good candidate in biotechnology applications after further study. This study leads to a better understanding for the structure arrangement of the bacterial PTS operons, and also highlights the utility of functional metagenomic approach in discovering of novel enzymes or gene clusters.

Acknowledgements

This work was supported by the grant of National Science Foundation of China (31170117, 31270156), National marine research funds for public welfare projects of China (201205020), Science and Technology Plan Project in Guangdong Province (2012B010300021), National Science Foundation of Guangdong Province (S2012010010464).

References

- [1] D. Ajdić, V.T. Pham, Global transcriptional analysis of *Streptococcus mutans* sugar transporters using microarrays. *J. Bacteriol.* 189(2007) 5049-5059.
- [2] C.L. An, M.K. Kim, T.H. Kang, J. Kim, H. Kim, H.D. Yun, Cloning and biochemical analysis of β -glucoside utilization (bgl) operon without phosphotransferase system in *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Microbiol. Res.* 167(2012) 461-469.
- [3] M. Kricker, B.G. Hall, Directed evolution of cellobiose utilization in *Escherichia coli* K12. *Mol. Biol. Evol.* 1(1984)171-182.
- [4] P. Kotrba, M. Inui, H. Yukawa, Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J. Biosci. Bioeng.* 92(2001) 502-517.
- [5] J. Deutscher, C. Francke, P.W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 70(2006) 939-1031.
- [6] J. Plumbridge, Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. *Curr. Opin. Microbiol.* 5(2002) 187-193.
- [7] P. Joyet, H. Bouraoui, F.M. Aké, M. Derkaoui, A.C. Zébré, T.N. Cao et al., Transcription regulators controlled by interaction with enzyme IIB components of the phosphoenolpyruvate:

sugar phosphotransferase system. *Biochim. Biophys. Acta* 1834(2013) 1415-1424.

[8] W.J. Mitchell, P. Tewatia, P.G. Meaden, Genomic analysis of the phosphotransferase system in *Clostridium botulinum*. *J. Mol. Microbiol. Biotechnol.* 12(2007) 33-42.

[9] J.H. Tchieu, V. Norris, J.S. Edwards, M.H. Jr. Saier, The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 2001. 3(2001) 329-346.

[10] T.R. Raghunand, S. Mahadevan, The beta-glucoside genes of *Klebsiella aerogenes*: conservation and divergence in relation to the cryptic bgl genes of *Escherichia coli*. *FEMS Microbiol. Lett.* 223(2003) 267-274.

[11] D. Harwani, P. Zangoui, S. Mahadevan, The beta-glucoside (bgl) operon of *Escherichia coli* is involved in the regulation of oppA, encoding an oligopeptide transporter. *J. Bacteriol.* 194(2012) 90-99.

[12] L.L. Parke, B.G. Hall, Characterization and nucleotide sequence of the cryptic operon of *Escherichia coli* K12. *Genetics* 124(1990) 455-471.

[13] P. Vinuselvi, S.K. Lee, Engineering *Escherichia coli* for efficient cellobiose utilization. *Appl. Microbiol. Biotechnol.* 92(2011) 125-132.

[14] A. Tartar, M.M. Wheeler, X. Zhou, M.R. Coy, D.G. Boucias, M.E. Scharf, Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol. Biofuels* 2009. 2(2009) 25.

[15] F. Legendre, M.F. Whiting, C. Bordereau, E.M. Canello, T.A. Evans, P. Grandcolas, The phylogeny of termites (Dictyoptera : Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. *Mol. Phylogenet. Evol.* 48(2008) 615-627.

[16] M. Dion, L. Fourage, J.N. Hallet, B. Colas, Cloning and expression of a beta-glycosidase gene from *Thermus thermophilus*. Sequence and biochemical characterization of the encoded enzyme. *Glycoconj. J.* 16(1999) 27-37.

[17] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, et al., Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36(2008) W465-W469.

[18] B.L. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37(2009) D233-D238.

[19] P. Charusanti, T.M. Conrad, E.M. Knight, K. Venkataraman, N.L. Fong, B. Xie, et al., Genetic basis of growth adaptation of *Escherichia coli* after deletion of pgi, a major metabolic gene. *Plos Genet.* 2010. 6(2010) e1001186.

[20] N. Mockli, D. Auerbach, Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. *Biotechniques* 36(2004) 872-876.

[21] N.J. Parry, D.E. Beever, E. Owen, I. Vandenberghe, J. Van Beeumen, M.K. Bhat, Biochemical characterization and mechanism of action of a thermostable beta-glucosidase purified from *Thermoascus aurantiacus*. *Biochem. J.* 353(2001) 117-127.

[22] S. Mukherjee, S. Khowala, Secretion of cellobiase is mediated via vacuoles in *Termitomyces clypeatus*. *Biotechnol. Prog.* 18(2002) 1195-1200.

[23] C.A. Uchima, G. Tokuda, H. Watanabe, K. Kitamoto, M. Arioka, Heterologous expression in *Pichia pastoris* and characterization of an endogenous thermostable and high-glucose-tolerant beta-glucosidase from the Termite *Nasutitermes takasagoensis*. *Appl. Environ. Microbiol.* 78(2012) 4288-4293.

- 542 [24] Q. Chen, O. Amster-Choder, The different functions of BglF, the *E. coli* beta-glucoside
543 permease and sensor of the bgl system, have different structural requirements. *Biochemistry*
544 37(1998) 17 040-17047.
- 545 [25] J. Plumbridge, O. Pellegrini, Expression of the chitobiose operon of *Escherichia coli* is
546 regulated by three transcription factors: NagC, ChbR and CAP. *Mol. Microbiol.* 52(2004)
547 437-449.
- 548 [26] B. Setlow, A. Cabrera-Hernandez, R.M. Cabrera-Martinez, P. Setlow, Identification of
549 aryl-phospho-beta-D-glucosidases in *Bacillus subtilis*. *Arch. Microbiol.* 181(2004) 60-67.
- 550 [27] K. Rashamuse, T. Ronneburg, W. Sanyika, K. Mathiba, E. Mmutlane, D. Brady, Metagenomic
551 mining of feruloyl esterases from termite enteric flora. *Appl. Microbiol. Biotechnol.* 98(2014)
552 727-737.
- 553 [28] K. Rashamuse, N. Mabizela-Mokoena, T.W. Sanyika, B. Mabvakure, D. Brady, Accessing
554 Carboxylesterase Diversity from Termite Hindgut Symbionts through Metagenomics. *J. Mol.*
555 *Microbiol. Biotechnol.* 22(2012) 277-286.
- 556 [29] G. Bastien, G. Arnal, S. Bozonnet, S. Laguerre, F. Ferreira, R. Fauré, et al., Mining for
557 hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional
558 metagenomics. *Biotechnol. Biofuels* 2013. 6(2013) 78.
- 559 [30] Y.X. Wang, Y. Xu, J.M. Li, A novel extracellular ss-glucosidase from *Trichosporon asahii*:
560 yield prediction, evaluation and application for aroma enhancement of *Cabernet Sauvignon*. *J.*
561 *Food Sci.* 77(2012) M505-M515.
- 562 [31] T. Bahr, D. Lüttmann, W. März, B. Rak, B. Görke, Insight into bacterial phosphotransferase
563 system-mediated signaling by interspecies transplantation of a transcriptional regulator. *J.*
564 *Bacteriol.* 193(2011) 2013-2026.
- 565 [32] A.H. Kachroo, A.K. Kancherla, N.S. Singh, U. Varshney, S. Mahadevan, Mutations that alter
566 the regulation of the chb operon of *Escherichia coli* allow utilization of cellobiose. *Mol.*
567 *Microbiol.* 66(2007) 1382-1395.
- 568 [33] L. Lopian, Y. Elisha, A. Nussbaum-Shochat, O. Amster-choder, Spatial and temporal
569 organization of the *E. coli* PTS components. *EMBO J.* 29(2010) 3630-3645.
- 570 [34] T. Feng, H. Liu, Q. Xu, J. Sun, H. Shi, Identification and characterization of two endogenous
571 β -glucosidases from the termite *Coptotermes formosanus*. *Appl. Biochem. Biotechnol.*
572 176(2015) 2039-2052.
- 573 [35] C.A. Uchima, G. Tokuda, H. Watanabe, K. Kitamoto, M. Arioka, A novel glucose-tolerant
574 β -glucosidase from the salivary gland of the termite *Nasutitermes takasagoensis*. *J. Gen. Appl.*
575 *Microbiol.* 59(2013) 141-145.
- 576 [36] Q. Wang, C. Qian, X.Z. Zhang, N. Liu, X. Yan, Z. Zhou, Characterization of a novel
577 thermostable β -glucosidase from a metagenomic library of termite gut. *Enzyme Microb.*
578 *Technol.* 51(2012) 319-324.
- 579 [37] J. Krisch, O. Bencsik, T. Papp, C. Vágvölgyi, M. Takó, Characterization of a beta-glucosidase
580 with transgalactosylation capacity from the zygomycete *Rhizomucor miehei*. *Bioresour.*
581 *Technol.* 114(2012) 555-560.
- 582 [38] K. Pokusaeva, M. O'Connell-Motherway, A. Zomer, G.F. Fitzgerald, D. van Sinderen,
583 Characterization of two novel alpha-glucosidases from *Bifidobacterium breve* UCC2003. *Appl.*
584 *Environ. Microbiol.* 75(2009) 1135-1143.
- 585 [39] B. Benoliel, M.J. Pocas-Fonseca, F.A. Torres, L.M. de Moraes, Expression of a

Glucose-tolerant beta-glucosidase from *Humicola grisea* var. *thermoidea* in *Saccharomyces cerevisiae*. Appl. Biochem. Biotechnol. 160(2010) 2036-2044.

- [40] N. Palackal, C.S. Lyon, S. Zaidi, P. Luginbühl, P. Dupree, F. Goubet, et al., A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. Appl. Microbiol. Biotechnol. 74(2007) 113-124.

Figure legends

Fig.1. Two overlapping PTS cluster components derived from metagenomic library of termite hindgut (A) and other glucoside specific PTS operons in *E. coli* (B).

Fig.2. SDS-PAGE analysis (A) and Western Blot analysis (B) of recombinant Glu1392 and Glu1923, as well as Native-PAGE analysis and Western Blot analysis (C) of expression products of intact overlapping gene cluster *glu3311*. (A) Lanes: M, standard protein molecular mass markers (sizes in kilodaltons are indicated on the left); 1, recombinant Glu1392 from supernatant of *E. coli* BL21(DE3) cell lysates; 2, recombinant Glu1923 from supernatant of *E. coli* BL21(DE3) cell lysates; (B) Lanes: M, standard protein molecular mass markers; 1, recombinant Glu1392 from supernatant of *E. coli* BL21(DE3) cell lysates; 2, recombinant Glu1923 from supernatant of *E. coli* BL21(DE3) cell lysates; (C), Native-PAGE (Lane 1) and western blotting analysis (Lane 2) of the expression products of intact overlapping gene cluster *glu3311*. The target protein bands were indicated by arrows.

Fig.3. Hydrolytic activity analysis of glucoside PTS components in *E. coli* towards four substrates. (A) aesculin LB plate incubating for two days; (B) cellobiose M9 minimal plate incubating for one week; (C) maltose M9 minimal plate incubating for one week;

(D) trehalose M9 minimal plate incubating for one week. In abovementioned plates, 1, BL21 (DE3) strains containing vector *pet-32a-glu1923*; 2, BL21 (DE3) strain containing vector *pet-32a-glu1392*; 3, BL21 (DE3) strain contain two compatible vectors *pet-32a-glu1392* and *pet-28a-glu1923*; 4, BL21 (DE3) strain contain overlapping gene cluster *glu3311*.

Fig.4. Yeast two hybrid analysis of the interaction between Glu1392 and Glu1923.

(A) Yeast two hybrid assay. 1, interaction between Glu1392 and intact Glu1923; 2, interaction between Glu1392 and Glu1923 mutant deleted EIIA domain; 3, interaction between Glu1392 and Glu1923 mutant deleted EIIA and EIIB domains; 4, interaction between Glu1392 and Glu1923 mutant deleted EIIB domain; (B) The structure of *Glu1923* protein and its three deletion mutants.

Fig.5. The relative activity of Glu1392, Glu1923, and the mixture of Glu1392 and Glu1923 toward three substrates. The relative activity was defined as the relative value to the maximum activity. Data represent the means of three experiments and error bars represent standard deviation.

Fig.6. Effects of pH (A) and temperature (B) on the activity and stability of Glu1392 using pNPG as substrate. Data points are the average of triplicate measurements and error bars represent standard deviation.

633

634

635 **Table 1 Susceptibility of Glu1392 toward metal ions and chemical reagents**

Compounds	Concentrations	Residual activity (%)
K ⁺	10 mM	109.87±2.59
Li	10 mM	92.50±3.17
Na ⁺	10 mM	107.46±2.37
Fe ³⁺	10 mM	55.43±1.32
Fe ²⁺	10 mM	48.67±1.95
Mg ²⁺	10 mM	93.24±2.11
Mn ²⁺	10 mM	87.50±2.03
Hg ²⁺	10 mM	10.5±0.98
Ni ²⁺	10 mM	22.35±1.13
Zn ²⁺	10 mM	15.33±1.06
Co ²⁺	10 mM	25.55±1.98
Ca ²⁺	10 mM	81.23±2.23
Methanol	10 %	96.31±1.76
Ethanol	10 %	95.24±1.94
Isopropanol	10 %	98.52±2.17
TRItion X-100	1 %	90.94±2.00
SDS	1 %	43.01±2.49
EDTA	1 %	48.13±1.38

636

637

638

639

640

641

642

643

Figure 1

Figure 2

648

649 **Figure 3**

A

B

650

651 **Figure 4**

Figure 5

Figure 6