UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOLOGIA CELULAR

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (BIOLOGIA MOLECULAR)

BIOPROSPECÇÃO DA MICROBIOTA INTESTINAL DE *SYNTHERMES WHEELERI* UTILIZANDO O METAGENOMA BACTERIANO PARA MINERAR ENZIMAS CAPAZES DE CONVERTER LIGNOCELULOSE EM QUÍMICOS COM ALTO VALOR AGREGADO.

RAUL ALCÂNTARA TEIXEIRA LIMA

BRASÍLIA-DF

2023

RAUL ALCÂNTARA TEIXEIRA LIMA

BIOPROSPECÇÃO DA MICROBIOTA INTESTINAL DE *SYNTHERMES WHEELERI* UTILIZANDO O METAGENOMA BACTERIANO PARA MINERAR ENZIMAS CAPAZES DE CONVERTER LIGNOCELULOSE EM QUÍMICOS COM ALTO VALOR AGREGADO.

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular) da Universidade de Brasília, como requisito parcial para obtenção do título de Doutor em Biologia Molecular.

Orientador

Prof. Dr. Ricardo Henrique Krüger

BRASÍLIA-DF

2023

Dedico este trabalho à minha mãe, Marta Dias, e à minha irmã, Erica Mônica, que sempre me apoiaram nessa caminhada. Dedico também a mim mesmo, por todo o esforço pessoal colocado em ação. E também dedico a toda a comunidade cientifica envolvida na luta contra a Pandemia da COVID-19.

AGRADECIMENTOS

Gostaria de agradecer ao meu orientador, Prof. Dr Ricardo Henrique Krüger, que desde o mestrado sempre foi muito solicito em querer me ajudar no trabalho proposto, e nunca mediu esforços para que eu pudesse desfrutar das melhores condições possíveis para desempenhar as atividades. Além disso, sempre foi muito humano e empático, com todas as causas pessoais que eu passei e precisei de apoio. Obrigado por todas as dicas e conselhos. O senhor vai estar sempre no meu coração, porque além de Professor, pelo qual tenho extremo respeito e admiração, eu ganhei um amigo.

Quero agradecer à UnB, mais especificamente, ao Instituto de Ciências Biológicas, por ter ofertado o Edital de seleção de doutorado que possibilitou o meu ingresso na Pós-graduação *stricto senso,* no Programa de Ciências Biológicas (Biologia Molecular) – PPGBM. Aos funcionários, servidores e terceirizados, que sempre nos ajudaram direta ou indiretamente durante o percurso, desde o seu Antônio que deixava os corredores sempre limpos e agradáveis até à Luciana, por exemplo, que sempre deu todo suporte burocrático necessário. Sou grato também aos funcionários do Restaurante Universitário que ajudaram na alimentação do dia a dia.

Quero agradecer aos meus colegas de laboratório que me apoiaram e compartilharam de seu conhecimento técnico-científico, e contribuíram para que eu pudesse reproduzir os experimentos, e além disso se tornaram amigos pessoais para a vida. São eles: Jéssica Pinheiro, Jovane Dantas, Gabrielle Rosa, Juliana Peixoto, Carla Vizzotto, Jéssica Fernandes, Helena Ipê, Tayná Diniz, Regina Sartori, Rafaela Silveira, Otávio Pinto, entre outros que posso eventualmente ter esquecido.

Sou muito grato aos meus amigos pessoais que fiz morando em Brasília, porque eles estiveram comigo também nessa caminhada e contribuíram para o meu progresso. São eles: Andreza Bona, Wilson Coelho, Ana Clara, Anderson, Wembesom Mendes, Alisson Bolina, Brendo, Tereza Cristina, Rafael Ícaro.

Gostaria de agradecer à minha família que me forneceu todo o apoio para a caminhada de sucesso.

Gostaria de agradecer também a Deus que sempre foi minha maior força e me manteve firme para não desistir em momentos de adversidade.

E também sou grato às instituições de fomento que forneceram o apoio financeiro indispensável para que eu pudesse trabalhar de forma tranquila. O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que forneceu a minha bolsa, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) que fomentaram projetos de pesquisa e participações em eventos científicos, ao Programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular) – PPGBM por fomentar a participação em eventos científicos.

"Entenda os seus medos, mas jamais deixe que eles sufoquem os seus sonhos."

Lewis Carroll, Alice no País das Maravilhas, 1865.

RESUMO

LIMA, RAT. "Bioprospecção da microbiota intestinal de *Synthermes wheeleri* utilizando o metagenoma bacteriano para minerar enzimas capazes de converter lignocelulose em químicos com alto valor agregado". 163 folhas. Tese de Doutorado – Instituto de Biologia, Universidade de Brasília, Brasília, 2023.

Os cupins consomem aproximadamente 3 a 7 bilhões de toneladas de materiais lignocelulósicos por ano e, portanto, representam um dos decompositores de lignocelulose mais prolíficos e eficientes da Terra. A bioconversão de polissacarídeos da parede celular por cupins é um processo altamente coordenado alcançado pelos simbiontes microbianos residentes no intestino. Neste trabalho objetivamos utilizar o potencial biotecnológico desses microrganismos para a obtenção de enzimas capazes de converter os polissacarídeos em produtos químicos de alto valor agregado e entender como as GHs se distribuem no instestino da espécie. O processo foi realizado com o metagenoma intestinal bacteriano de *Syntermes wheeleri*, uma espécie endêmica de cupins do cerrado brasileiro. Aqui desenvolvemos análises de bioinformática integrada que sugeriu grupos de proteínas em árvores filogenéticas com potencial de inovação. Os domínios dessas proteínas foram distribuídos em Glicosil Hidrolases (GHs) - 3, 5, 9 e 10 com representantes dos filos Firmicutes, Proteobacteria, Bacteroidetes e Spirochaeta, entre outros. Com o objetivo de caracterizar e avaliar o potencial biotecnológico utilizamos *E. coli* BL21(DE3) e os sistemas de plasmídeos pET para sintetizar alguns deles. Os resultados bioquímicos demonstraram 40-50°C como a melhor temperatura de atividade para Exo8574 (exoglucanase) e Bgl7226 (β-glicosidase), além de apresentarem pH ácido e básico como pH ótimo, respectivamente. O dicroísmo circular demonstrou a dependência da estrutura secundária pelo pH de acordo com a mudança nas quantidades de α-hélices e folhas β. Essas características proporcionaram as melhores condições para a sacarificação. Este processo pode ser usado em diferentes biomassas, como cana-de-açúcar e espiga de milho, usando o tratamento térmico como um processo de pré-tratamento ambientalmente sustentável, em comparação com o pré-tratamento químico, para melhorar o acesso de proteínas aos polissacarídeos. Nossos resultados ajudarão a elucidar a capacidade das enzimas selecionadas do metagenoma de bactérias intestinais de *S. wheeleri* em processos de biotecnologia de bioconversão de lignocelulose.

Palavras-chave: glicosil hidrolases; metagenoma; *Syntermes wheeleri*; lignocelulose.

ABSTRACT

LIMA, RAT. "Bioprospecting the gut microbiota of *Synthermes wheeleri* using the bacterial metagenome to mine enzymes capable of converting lignocellulose into chemicals with high added value". 163 sheets. PhD Thesis - Institute of Biology, University of Brasília, Brasília, 2023.

Termites consume approximately 3 to 7 billion tons of lignocellulosic materials per year and therefore represent one of the most prolific and efficient lignocellulose decomposers on Earth. The bioconversion of cell wall polysaccharides by termites is a highly coordinated process achieved by microbial symbionts residing in the gut. In this work we aimed to use the biotechnological potential of these microorganisms to obtain enzymes capable of converting polysaccharides into high value-added chemicals and to understand how GHs are distributed in the gut of the species. The process was carried out with the bacterial gut metagenome of *Syntermes wheeleri*, an endemic termite species from the Brazilian cerrado. Here we developed integrated bioinformatics analyses that suggested groups of proteins in phylogenetic trees with potential for innovation. The domains of these proteins were distributed in Glycosyl Hydrolases (GHs) - 3, 5, 9 and 10 with representatives of the phyla Firmicutes, Proteobacteria, Bacteroidetes and Spirochaeta, among others. In order to characterize and evaluate their biotechnological potential, we used E. coli BL21(DE3) and pET plasmid systems to synthesize some of them. The biochemical results showed 40-50°C as the best activity temperature for Exo8574 (exoglucanase) and Bgl7226 (β-glucosidase), as well as acidic and basic pH as the optimum pH, respectively. Circular dichroism showed the dependence of the secondary structure on pH according to the change in the quantities of α-helices and β-sheets. These characteristics provided the best conditions for saccharification. This process can be used on different biomasses, such as sugarcane and corncob, using heat treatment as an environmentally sustainable pretreatment process, compared to chemical pretreatment, to improve protein access to polysaccharides. Our results will help elucidate the capacity of selected enzymes from the S. wheeleri gut bacteria metagenome in lignocellulose bioconversion biotechnology processes.

Keywords: glycosyl hydrolases; metagenome; *Syntermes wheeleri*; lignocelulose.

LISTA DE FIGURAS

Capítulo 1

Capítulo 2

Capítulo 3

Capítulo 4

LISTA DE TABELAS

Capítulo 2

Capítulo 3

Capítulo 4

LISTA DE SIGLAS

µg: micrograma

µL: microlitros

µM: micromolar

16 rRNA – RNA ribossomal 16S

cm – centímetro

Da: daltons

EC: *Enzyme Comission* (Comissão de enzimas)

EUA – Estados Unidos da América

GH: *Glicosil hydrolase* (Glicosil hidrolases)

h: hora

IPTG: Isopropyl β-D-1-thiogalactopyranoside (isopropil β-D-1 tiogalactopiranosídeo)

kDa: kiloDalton

M: molar

min: minutos

mM: milimolar

mPa: millipascal

NCBI: *National Center for Biotechnology Information (Centro Nacional de Informações sobre Biotecnologia)*

nm: nanomolar

OD: *optical density* (densidade óptica)

pNP - *p-Nitrophenyl*

pNPG - *p-Nitrophenyl β-D-cellobioside*

r.p.m: rotações por minuto

s: segundo

SUMÁRIO

Capítulo 1

Revisão da literatura

Justificativa

Objetivos

1.0 Revisão da literatura

1.1 Estratégias de bioprospecção – explorando a biodiversidade

A bioprospecção permite que os pesquisadores explorem a biodiversidade e identifiquem novas moléculas, enzimas e micróbios relevantes para pesquisas e aplicações industriais (Beattie et al., 2011; Afifahussain et al., 2022). As estratégias de bioprospecção podem ser subdivididas em microrganismos (metagenômica), enzimas (transcriptômica/proteômica) ou moléculas individuais (metabolômica). A bioprospecção microbiana explora uma vasta diversidade de microrganismos, que constituem mais de dois terços das formas de vida globais, presentes em uma ampla gama de ambientes, incluindo os extremos (Becker & Wittmann, 2020).

Um dos esforços mais fundamentais da biologia tem sido descrever a composição do mundo vivo. Séculos de pesquisa produziram uma imagem cada vez mais detalhada das espécies que habitam nosso planeta e seus respectivos papéis nos ecossistemas globais. Os microrganismos concentram a segunda maior biomassa mundial ficando atrás somente das plantas, sendo que as bactérias, os fungos e as Archaeas são os maiores representantes com um total de 70, 12 e 7 Gigatoneladas de carbono (Gt C), respectivamente **(Figura 1)** (Bar-On, Phillips & Milo, 2018).

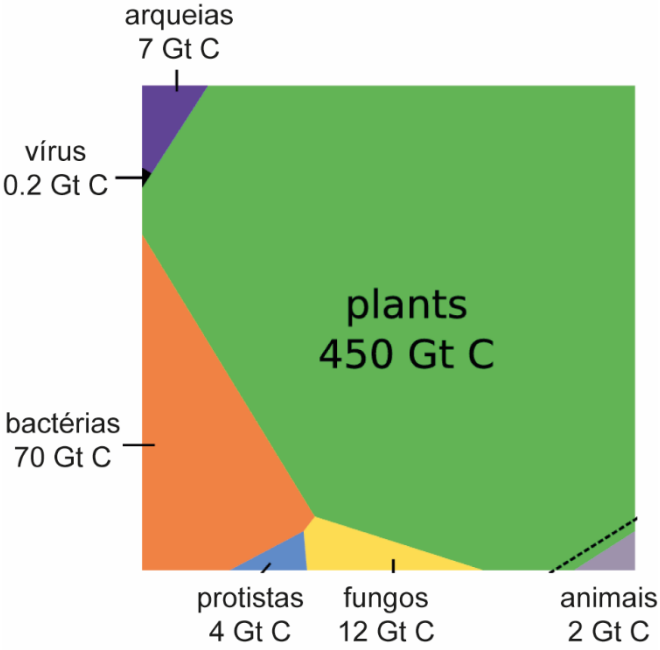


Figura 1. Representação gráfica da distribuição global de biomassa por táxons. Biomassas absolutas de diferentes táxons são representadas usando um diagrama de Voronoi, com a área de cada célula sendo proporcional à taxa global de biomassa (a forma específica de cada polígono não tem significado). Gt C (Gigatoneladas de Carbono). Adaptado de (Bar-On et al., 2018).

Os instrumentos metagenômicos fornecem informações até mesmo de micróbios de difícil cultivo, sendo uma fonte rica de novas vias metabólicas, enzimas e seus produtos catalisados **(Figura 2)**. A combinação de tecnologias baseadas em ômicas permite que os pesquisadores conectem genes com vias enzimáticas e biossíntese de metabólitos, alimentando a descoberta de novas moléculas para diversas aplicações (Bansal et al., 2022).

A bioprospecção microbiana é uma terminologia nova, mas a humanidade tem usado micróbios para seu benefício por muitos anos, desde a implementação de fermento para panificação até a descoberta de novas moléculas para tratar doenças (Perez Rojo et al., 2023). O recente surgimento da análise metagenômica desencadeou o potencial da bioprospecção microbiana. Estima-se que existam

mais de 10^{16} micróbios em apenas 1 tonelada de solo, dos quais 85% a 99% são de difícil cultivo (Curtis & Sloan, 2005; Lok, 2015). Foi previsto que a Terra hospeda um trilhão de espécies (10^{12}) de microrganismos, dos quais apenas 0,001% foram descobertos até agora (Locey & Lennon, 2016).

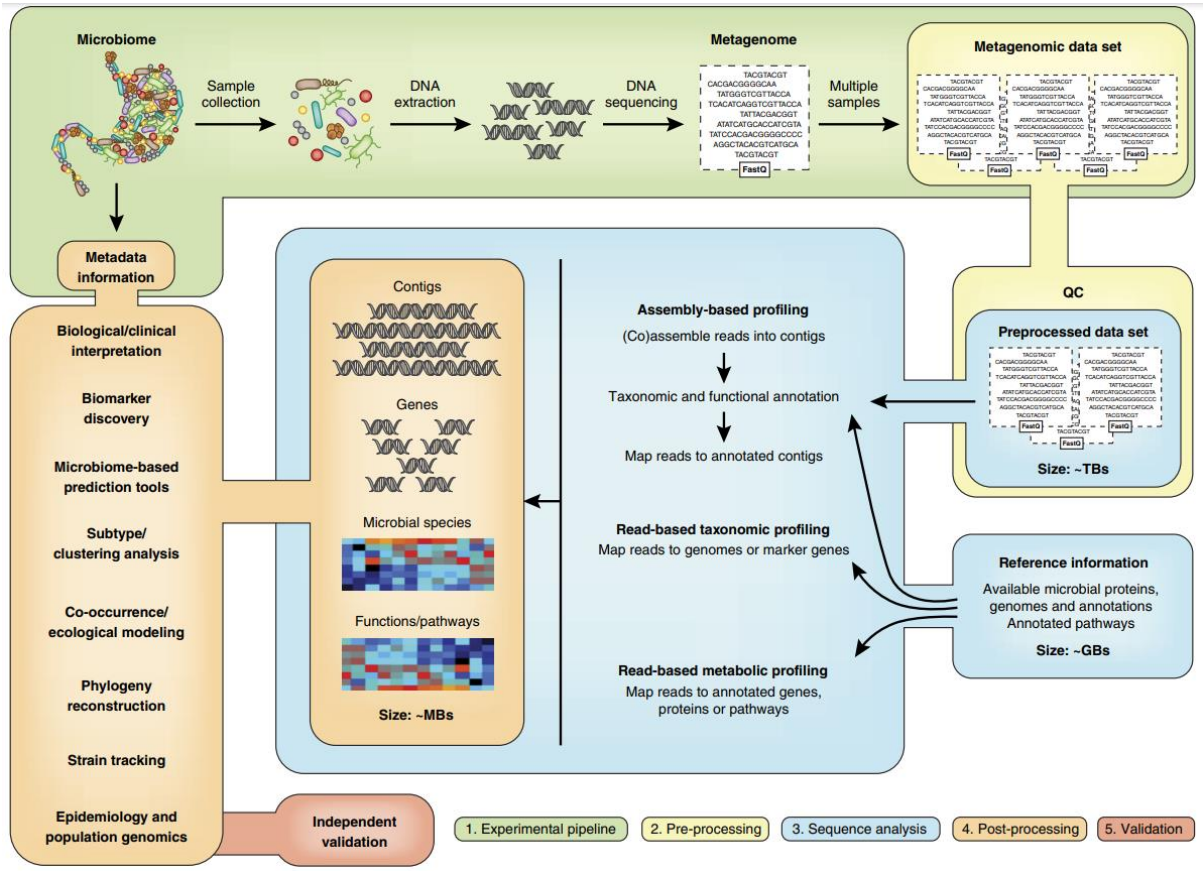


Figura 2. **Resumo de um fluxo de trabalho de metagenômica. Etapa (1):** desenho do estudo e protocolo experimental. **Etapa (2):** pré-processamento computacional. As etapas de controle de qualidade computacional (QC) minimizam vieses ou artefatos de sequência fundamentais, como remoção de adaptadores de sequenciamento, ajuste de qualidade, remoção de duplicatas de sequenciamento (usando, por exemplo, ferramentas FastQC, Trimmomatic121 ou Picard). As sequências de DNA estranhas ou não-alvo também são filtradas e as amostras são subamostradas para normalizar os números lidos se a diversidade de táxons ou funções for comparada. **Etapa (3):** análise das sequências. **Etapa (4):** pós-processamento. Várias técnicas estatísticas multivariadas podem ser usadas para interpretar os dados. **Etapa (5):** validação. Conclusões de dados biológicos de alta dimensão são suscetíveis a vieses de estudo, portanto, análises de acompanhamento são vitais. (Quince et al., 2017).

Os microrganismos conduzem os ciclos biogeoquímicos globais, sustentam as redes alimentares e sustentam a saúde de animais e plantas. Sua imensa

diversidade filogenética, metabólica e funcional representa um rico potencial de descoberta de novos táxons, enzimas e compostos bioquímicos, incluindo produtos naturais (Pye et al., 2017). Diferentes ambientes já foram explorados para novos micróbios, enzimas e vias metabólicas seguindo uma abordagem metagenômica, entre eles: solo, água, seres vivos (animais vertebrados e invertebrados), por exemplo (Corander et al., 2022; Roux & Emerson, 2022; Lima et al., 2020; Paoli et al., 2022; Thorn et al., 2023).

1.2 Bioprospecção microbiana do intestino de cupim

Os cupins são um clado de baratas eussociais (ordem Blattodea; anteriormente Isoptera) com aproximadamente 3000 espécies descritas (Evangelista et al., 2019). Eles são considerados importantes engenheiros de ecossistemas em habitats tropicais devido às significativas alterações ambientais que promovem. Sua maior contribuição para a funcionalidade do ecossistema é através da decomposição promovida por diferentes grupos alimentares que consomem madeira, grama, serapilheira e matéria orgânica do solo (Quince et al., 2017; Oliveira et al., 2022).

Todos os cupins se alimentam de lignocelulose derivada de plantas, uma mistura altamente abundante, mas recalcitrante, de celulose, hemicelulose e lignina. Eles dependem da microbiota gastrointestinal mutualística para mediar a

hidrólise e fermentação da lignocelulose, resultando na produção de acetato, hidrogênio (H_2) e metano (Li & Greening, 2022).

Os cupins digerem de 74% a 99% da celulose da dieta por hora por meio de sua microbiota intestinal simbiótica, tendo assim um impacto significativo no ciclo de carbono em florestas (sub)tropicais. Com interesses biotecnológicos em destaque, a última década de estudos ômicos relacionados ao intestino de cupins identificou, em abundância, enzimas ativas de carboidratos (*CAZymes*) envolvidas na desconstrução da biomassa vegetal, o que avançou significativamente a compreensão da digestão simbiótica nesses ecossistemas intestinais (Li et al., 2019).

Para os cupins "inferiores" que se alimentam de madeira, a digestão das partículas de madeira é realizada por um sistema celulolítico duplo que combina enzimas do hospedeiro e de seus simbiontes intestinais, incluindo flagelados celulolíticos e bactérias. Para os cupins "superiores" que se alimentam de madeira (família Termitidae), que são totalmente livres de flagelados intestinais, a biomassa vegetal é digerida em seu trato intestinal por meio de associações com a microbiota "procariótica", que é hipotetizada e cada vez mais corroborada por um papel na digestão simbiótica para os hospedeiros dos cupins **(Figura 3)** (Li & Greening, 2022).

Alguns genes que codificam enzimas que degradam a lignocelulose foram caracterizados a partir de microrganismos de cupins, contudo a maioria deles foram isolados de microrganismos de cultura. O perfil de filogenia 16S rRNA de amostras de intestino de cupins superiores revelou famílias altamente diversas de bactérias, incluindo várias espécies consideradas de difícil cultivo. Com isso, elas se tornam

uma fonte potencial para a descoberta de novos genes que codificam enzimas que hidrolisem a lignocelulose (Nimchua et al., 2012; Brune, 2014; Zhou et al., 2023).

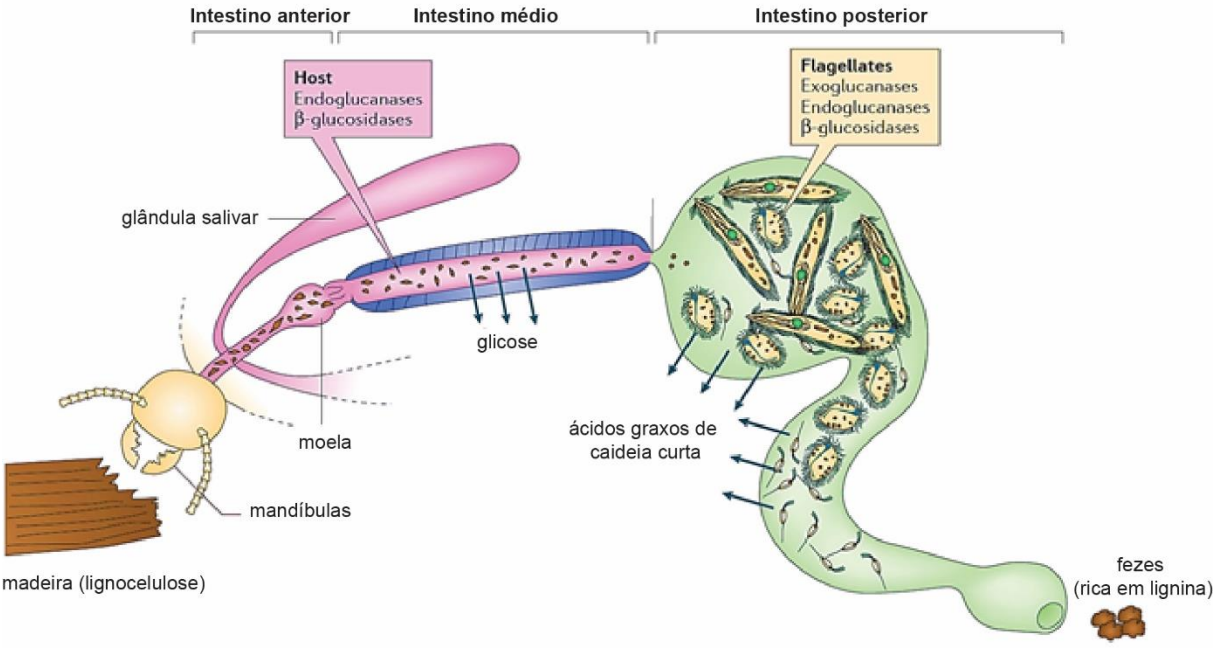


Figura 3. **Esquema ilustrativo representando o intestino dos cupins e as enzimas celulases agindo na lignocelulose.** Os cupins são divididos da boca ao ânus em intestino anterior, intestino médio e intestino posterior. A glicose liberada no intestino médio é absorvida ali mesmo pelo epitélio, enquanto a madeira pré-digerida segue para o intestino posterior, onde nos cupins inferiores receberá a ação enzimática de protozoários flagelados e nos superiores de bactérias. Adaptado de (Brune, 2014).

1.2.1 *Syntermes wheeleri* – espécie de estudo

O *Syntermes wheeleri* é uma espécie neotropical de cupim "superior" que faz parte da família dos Termitidae (Syntermitinae), os quais se alimentam de serapilheira (folhas secas) e se destacam pelo seu tamanho, com o soldado de *S. wheeleri* medindo 2 cm, aproximadamente, enquanto das demais espécies medem de 0,3 a 1 cm. Os operários da espécie são divididos em dois instares: o terceiro

cujos operários possuem cabeça clara e o quarto cujos operários possuem cabeça escura **(Figura 4)** (Constantino, 1995).

Os operários são os responsáveis pela digestão e alimentação dos outros componentes da colônia. Os membros do quarto instar forrageiam, principalmente à noite, coletando folhas mortas que são armazenadas dentro do cupinzeiro e servem para alimentação de ambos os instares. O intestino dessa espécie é dividido em faringe, esôfago, papo, moela, mesêntero, segmento misto, primeiro proctodeo (P1), válvula entérica (P2), terceiro proctodeo (P3), cólon (P4) e reto (P5). No Brasil, a distribuição de *S. wheeleri* concentra-se no Cerrado (Constantino, 1995).

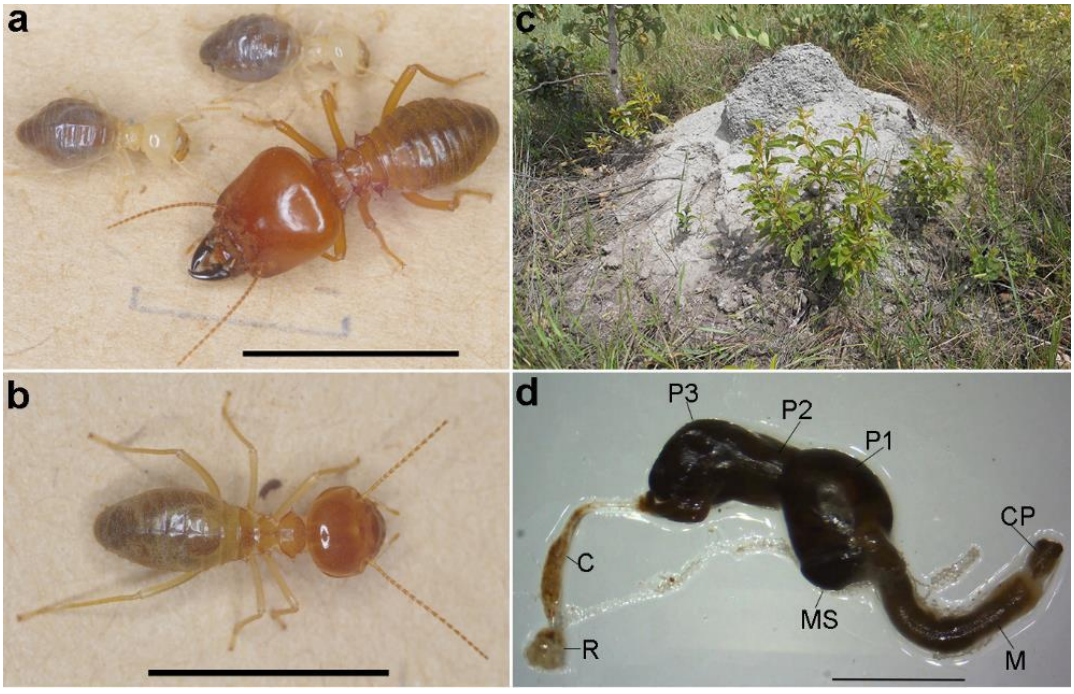


Figure 4. Os operários do cupim neotropical *Syntermes wheeleri*: (a) soldado e trabalhador de terceiro instar com cabeça branca, **(b)** trabalhador de quarto instar de cabeça vermelha, **(c)** o cupinzeiro e **(d)** compartimentos intestinais; papo (CP); intestino médio (M); segmento misto (MS); primeiro segmento proctodeal (P1); valva entérica (P2); terceiro segmento proctodeal (pança-P3); cólon (C); reto (R). Barras de escala, 1,0 cm em a, b e c; e 0,3 cm em d. (Santana et al., 2015).

1.3 Lignocelulose – a base alimentar dos cupins

A lignocelulose é um recurso de biomassa renovável abundante na Terra, com uma produção anual de mais de 20 bilhões de toneladas, incluindo madeira, plantas, resíduos agrícolas, resíduos florestais e resíduos sólidos urbanos, que podem ser convertidos em produtos ecológicos de alto valor agregado (López-Mondéjar et al., 2019; Liu et al., 2020). Os combustíveis fósseis e outros recursos não renováveis tem se tornado escassos. Nesse contexto, a utilização de energia renovável e seus derivados como substitutos podem efetivamente reduzir a dependência de energia não renovável, alcançando assim o desenvolvimento sustentável e os benefícios econômicos (Wu et al., 2022).

No entanto, a utilização em larga escala de lignocelulose ainda é um desafio. A hidrólise enzimática de celulose e hemicelulose é uma etapa fundamental no processo de transformação, contudo, a complexa estrutura reticulada da lignocelulose dificulta o processo ao inibir as interações enzimáticas, e apresentar ligações inespecíficas entre enzimas e a lignina, o que inibe o processo de hidrólise (Liu et al., 2016; Shirkavand et al., 2016). Esse tem sido um dos maiores desafios da área nos últimos anos, e para solucionar esse problema, os pré-tratamentos da biomassa lignocelulósica vem sendo umas das melhores opções industriais visando a desconstrução da estrutura reticulada, diminuindo o conteúdo de lignina e a cristalinidade da celulose, por exemplo (Yoo et al., 2020).

A biomassa lignocelulósica é composta principalmente por uma rede altamente complexa de homopolissacarídeo celulose (40-60% do peso seco), vários

heteropolissacarídeos hemicelulósicos (20-35% do peso seco) e o polímero aromático lignina (15-40% do peso seco) **(Figura 5)** (Zoghlami & Päes, 2019). Os principais pré-tratamentos utilizados constituem métodos físicos e químicos mas apresentam desvantagens ecológicas quando comparados ao biológico.

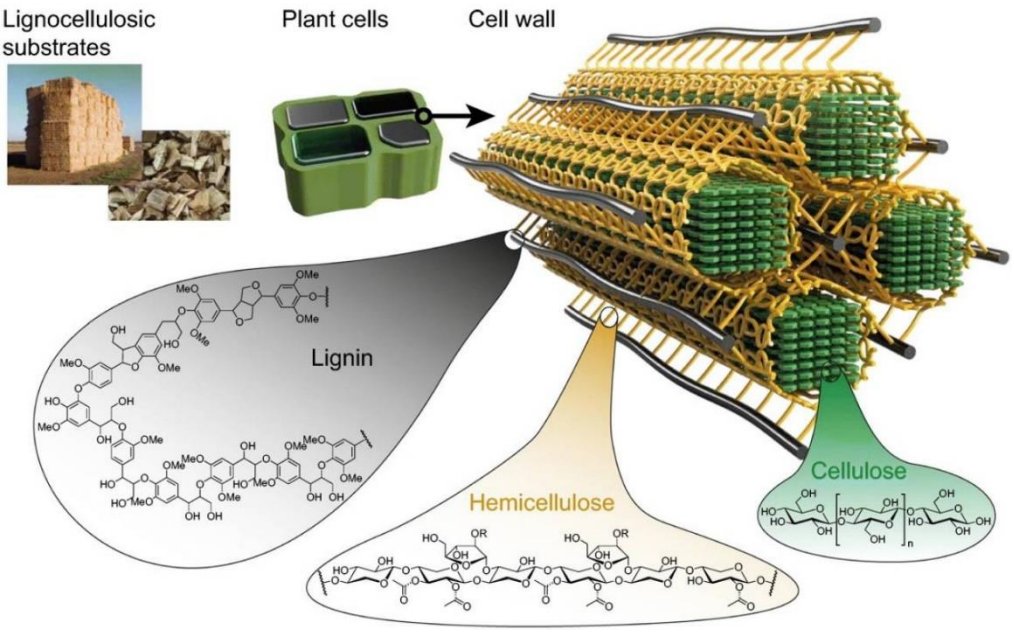


Figura 5. **Modelo esquemático do arranjo de celulose, hemicelulose e lignina em fontes lignocelulósicas** (Brethauer et al., 2020).

Os tratamentos físicos incluem prensa mecânica, tratamento ultrassônico, tratamento por micro-ondas, contudo consomem muita energia e geralmente são usados como método auxiliar. Já os tratamentos químicos têm a vantagem da alta eficiência, como, por exemplo, a palha de arroz tratada com hidróxido de potássio e ureia que podem obter 92,38% de rendimento de açúcar após 32,47% de lignina ter sido removido nas condições ideais. Apesar disso, os reagentes químicos tradicionais apresentam algumas desvantagens, como a recuperação incompleta

de polissacarídeos, a possível geração de inibidores e riscos de poluição ambiental, afetando assim a eficiência da produção **(Figura 6)** (Wang et al., 2021; Tarrsini et al., 2021).

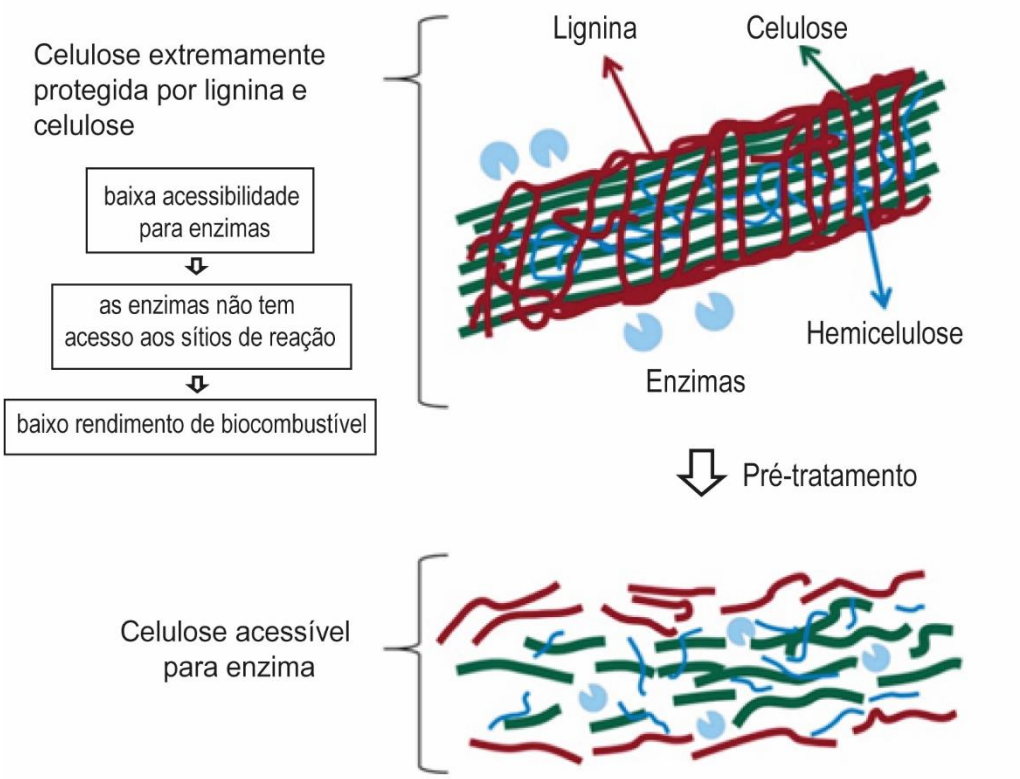


Figura 6. **Função do pré-tratamento na lignocelulose** (Adaptado de Shirkavand et al., 2016).

Os pré-tratamentos tradicionais, como exemplificado anteriormente, requerem equipamentos especiais, além de produzirem inibidores que afetam o processo. Relativamente falando, os métodos biológicos consomem menos energia e não produzem inibidores durante o processo de tratamento porque é realizado em condições amenas, apesar de demandarem um maior período de ação e perca de atividade sobre o substrato (Shirkavand et al., 2016).

1.3.1 Celulose

Entre todos os biopolímeros, a celulose, um polímero de monômeros de glicose ligados por ligações do tipo β(1→4), é o recurso natural mais abundante do planeta, com uma produção anual de cerca de 1,5 trilhão de toneladas, e é uma fonte quase inesgotável de matéria-prima para a fabricação de bioprodutos ecológicos e sustentáveis (Dassanayake et al., 2018; Baghaei & Skrifvars, 2020).

A celulose é produzida principalmente pela fotossíntese das plantas, sendo em sua maioria em forma de fibras. O grau em que a celulose é polimerizada é determinado pela fonte. As mais abundantes são as madeiras macias e duras, havendo também fontes como plantas anuais, ervas daninhas e bambus, assim como animais como o organismo marinho *Tunicat*, que produz a celulose animal tunicina; microrganismos; algas (Heinze & Liebert, 2012; Zhao et al., 2017; Strnad & Zemljič, 2023).

A celulose é rica em grupos hidroxila, os quais realizam ligações de hidrogênio intramoleculares e intermoleculares e possibilitam modificações químicas adicionais e ajuste de propriedades **(Figura 7)**. A ligação glicosídica β-1,4, uma unidade repetitiva na estrutura da celulose, é composta por dois anéis de glicose anidra ($(C_6H_{10}O_5)_n$; n = 10.000 a 15.000, onde n é determinado pelo material de origem da celulose) que estão ligados juntos por oxigênio ligado covalentemente a C1 de um anel de glicose e C4 do anel adjacente (1 ➔ 4 ligações). Além disso, a ligação de hidrogênio intracadeia entre o oxigênio no anel molecular

próximo e o grupo OH estabiliza a ligação da estrutura e gera uma estrutura linear (Baghaei & Skrifvars., 2020; Muddasar et al., 2022).

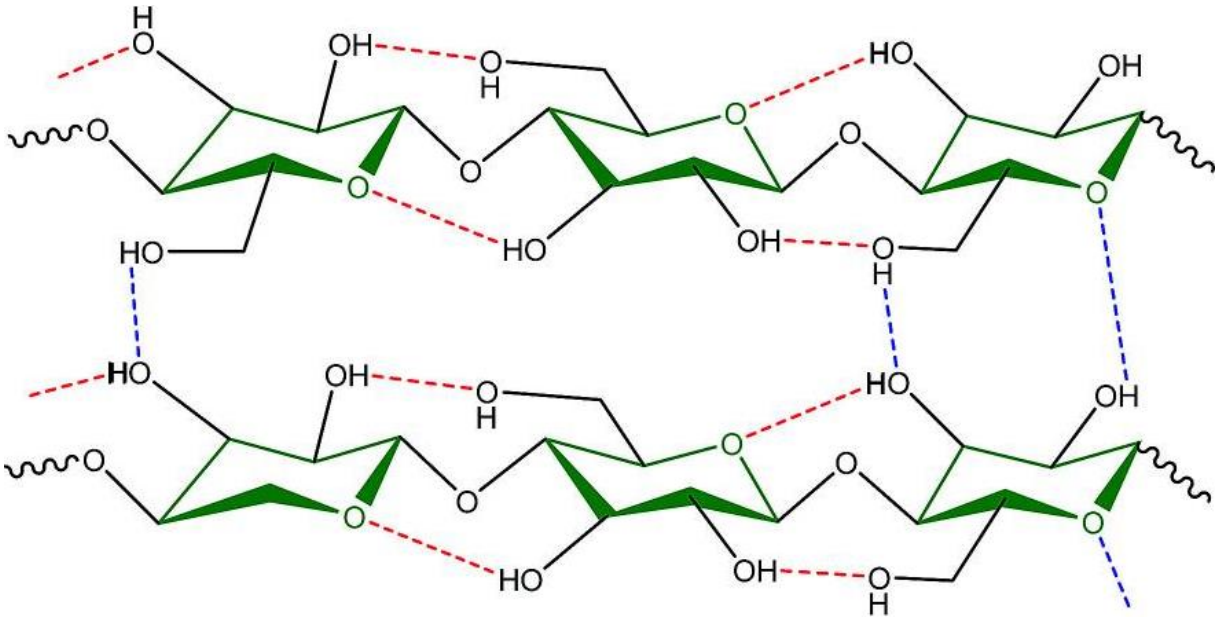


Figura 7. **Estrutura da celulose.** Ligações de hidrogênio intra e intermoleculares na estrutura molecular da celulose. As unidades de anidroglicose são ligadas por ligações 1,4-β-glicosídicas (Shaghaleh et al., 2018).

Existem pelo menos quatro diferentes estruturas polimórficas supramoleculares de celulose. A celulose nativa tipo I é a mais cristalina, e é composta pelas fases cristalinas Iα (triclínico, ou seja, todos os eixos cristalográficos com ângulos diferentes de 90°) e Iβ (monoclínica, ou seja, dois eixos cristalográficos com ângulos de 90° e um diferente de 90°). Algas e bactérias geralmente exibem a estrutura Iα, enquanto a parede celular dos tunicados e das plantas superiores são tipicamente caracterizados pela estrutura Iβ. A celulose tipo II é formada pelo chamado tratamento de mercerização com hidróxido de sódio, ou durante o processo de regeneração na produção de fibras de viscose. O tratamento com diamina da celulose I ou II dá o tipo III1 ou III2, respectivamente, e o tipo IV é

produzido por tratamento térmico do tipo III em glicerol (Pu et al., 2008; Muddasar et al., 2022).

1.3.2 Hemicelulose

A hemicelulose é um polissacarídeo composto de diferentes tipos de monossacarídeos, como pentoses (β-D-xilose e α-L-arabinose), hexoses (β-D-manose, β-D-glicose e α-D-galactose) e / ou ácidos urônicos (α-D-glicurônico, α-D-4-O-metilgalacturônico e α-D-ácido galacturônico) **(Figura 8)**. O conteúdo e a composição da hemicelulose variam muito com as espécies de plantas, maturidade, madeira inicial e tardia, tipo de célula e posição morfológica (Scheller & Ulvskov, 2010; Yuan et al., 2021).

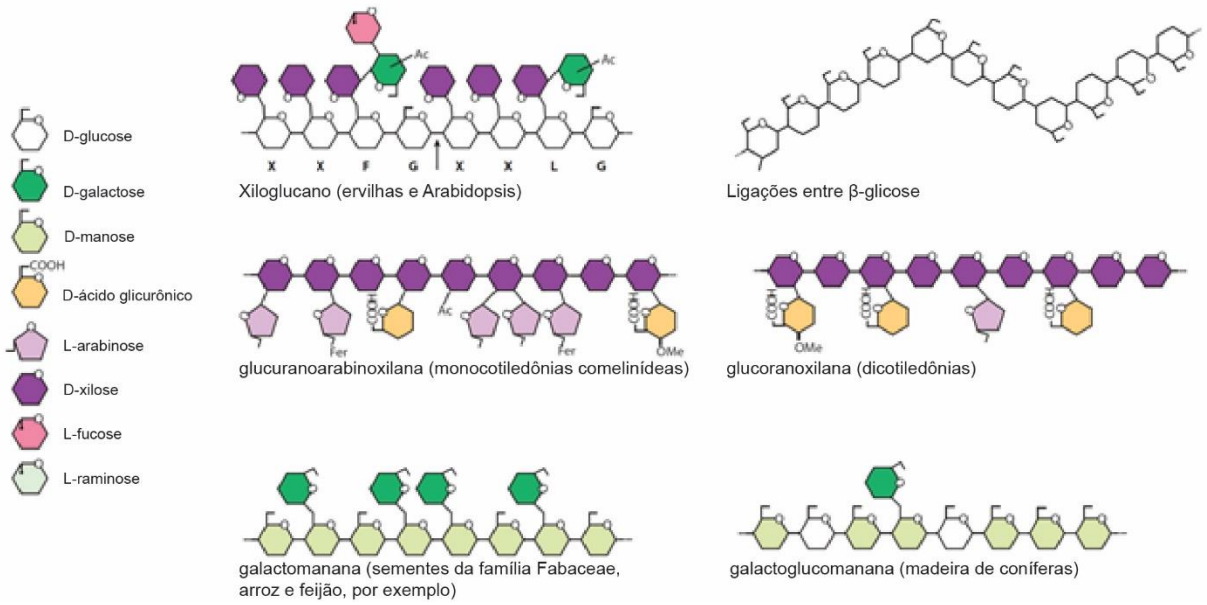


Figura 8. Ilustração esquemática dos tipos de hemicelulose encontrados nas paredes celulares dos vegetais. A hemicelulose é formada por unidades de pentoses, hexoses e/ou ácidos urônicos. Adaptado de (Scheller & Ulvskov, 2010).

A pirólise, como autohidrólise, hidrólise ácida diluída, hidrólise catalítica e hidrólise enzimática, é uma importante estratégia para obtenção de hemicelulose e seus derivados para a fabricação de materiais e produtos químicos com importância comercial. Os diferentes processos utilizados permitem a obtenção de diferentes concentrações de monossacarídeos e rendimentos (Yuan et al., 2021).

A hemicelulose armazenada em sementes, por exemplo, é usada diretamente como produtos na indústria de alimentos: gomas de guar e alfarroba (galactomananas), goma konjac (glucomanana) e goma de tamarindo (xiloglucana). Além disso, as hemiceluloses conferem propriedades importantes a muitos alimentos e rações. Na indústria de panificação, os arabinoxilanos insolúveis afetam a qualidade do cozimento. Os β-(1→3,1→4)-glucanos e arabinoxilanos são compostos antinutricionais bem conhecidos na alimentação animal e podem causar problemas de filtragem e turvação na indústria cervejeira devido à sua viscosidade Para aliviar esses problemas, enzimas que degradam a hemicelulose são adicionadas à ração e são usadas nas indústrias de panificação e cervejaria.

Por outro lado, o β-(1→3,1→4)-glucano tem um efeito documentado de redução do colesterol em humanos hipercolesterolêmicos e a ingestão diária de β-(1→3,1→4)-glucanos é recomendado pela *Food and Drug Administration* dos EUA. Na produção de biocombustíveis celulósicos, as hemiceluloses afetam a sacarificação da biomassa, e os açúcares liberados, principalmente as pentoses, são menos desejáveis para a fermentação do que as hexoses. Além disso tem sido estudada como possível bio-adsorvente de poluentes (Keogh et al., 2003; Yuan et al., 2021; Xiang et al., 2022).

1.3.3 Lignina

A lignina é um dos principais elementos não-carboidratos heterogêneos da lignocelulose cercado por ligações de celulose e hemicelulose, o que fornece hidrofobicidade e resistência às plantas. É um biopolímero aromático com unidades fenilpropanóides derivadas principalmente de álcoois monolignóis sinaptil, coniferil, e coumaril **(Figura 9)**. É um dos principais fatores responsáveis pela recalcitrância da biomassa (Singhvi & Kim, 2021).

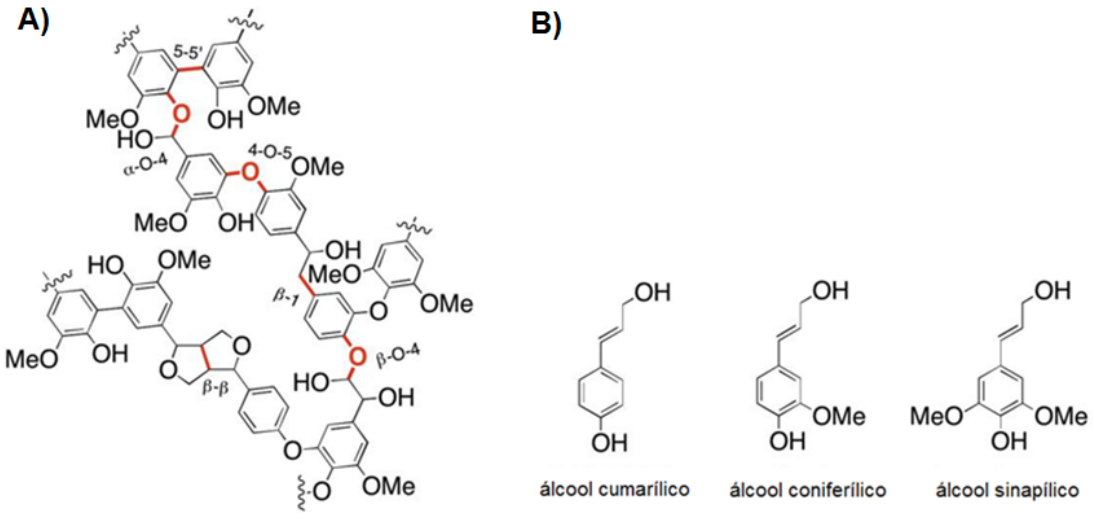


Figura 9. Estrutura da Lignina. (A) Possível estrutura da lignina; **(B)** três estruturas primárias de álcool na lignina. Adaptado de (Dai, Patti & Sato, 2016).

A quantidade e a composição da lignina variam muito entre e dentro das espécies de plantas, tecidos, tipos de células e camadas da parede celular e são influenciadas por fatores de desenvolvimento e ambientais. Até o momento, metabólitos de onze classes diferentes foram descobertos como blocos de construção de lignina que ocorrem naturalmente (Vanholme et al., 2019.)

1.4 Enzimas envolvidas na degradação de lignocelulose

A lignocelulose é constituída principalmente de celulose, com uma composição de aproximadamente 51%, seguida por hemicelulose e lignina, que perfazem cerca de 31% e 28%, respectivamente, que juntamente com proteínas, lipídios, pectina, açúcares solúveis e minerais são os componentes menores. A lignocelulose pode ser degradada em açúcares simples que podem ser utilizados para a fabricação de biocombustíveis e inúmeros produtos de valor agregado usando várias enzimas ativas de carboidratos (*CAZymes*) (Pauly & Keegstra, 2008; Chettri et al., 2020; Wardman et al., 2022).

CAZymes são a família de enzimas que atuam em carboidratos para sua biotransformação, ou seja, síntese, metabolismo, juntamente com sua modificação e transporte. Elas foram organizadas em várias classes considerando suas atividades catalíticas e similaridade de sequências de aminoácidos. As classes incluem hidrolases de glicosídeos (GHs), esterases de carboidratos (CEs), liases de polissacarídeos (PLs), glicosiltransferases (GTs) e atividades auxiliares (AA). Para manter as informações cada vez maiores sobre esta classe de enzimas, um sistema de banco de dados online foi criado - *CAZy* (http://www.cazy.org/) (Lombard et al., 2013; André et al., 2014; Chettri et al., 2020).

Recentemente, uma nova classe de *CAZyme*, as monooxigenases de polissacarídeos líticos (LPMOs), foram identificadas e classificadas na família das AA. Dentre essas classes, a família GH são as principais enzimas que promovem a degradação da biomassa vegetal. Elas utilizam fonte de açúcar para a geração de

biocombustíveis, como bioetanol e produtos de valor agregado, como biossurfactantes e ácidos orgânicos, dentre outros. Também é utilizado na extração de suco de frutas, branqueamento de papel, produção de detergente e processamento têxtil, para citar alguns (Paulino et al., 2017).

As GHs são definidas como um grupo generalizado de enzimas que hidrolisam a ligação glicosídica entre dois ou mais carboidratos ou entre um carboidrato e uma fração não carboidrato. Especificamente, as GHs ativas sobre polissacarídeos são as O-glicosidases, classificadas no grupo EC 3.2.1 juntamente com as S-glicosidases, enquanto as N-glicosidases pertencem ao grupo EC 3.2.2. Referindo-se ao banco de dados *CAZy* e com base nas semelhanças nas sequências de aminoácidos, as GHs são agrupados em 184 famílias das quais 24 são celulases (1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 16, 26, 30, 39, 44, 45, 48, 51, 116, 124, 131, 148, 175 e 180) - http://www.cazy.org/Glycoside-Hydrolases.html último acesso em Setembro de 2023 (Lombard et al. 2013).

Já as ligninases trabalham removendo o escudo de lignina existente quando o complexo de lignina-carboidrato cobre as fibras de celulose. As ligninases que compreendem um conjunto de enzimas não hidrolíticas (lacases, lignina-peroxidases e manganês peroxidases) atacam o escudo de lignina através da geração de radicais livres altamente reativos e inespecíficos que quebram as ligações carbono-carbono e éter na estrutura da lignina (Pollegioni et al., 2015; Atiwesh et al., 2022).

Por sua vez a rede de hemicelulose heterogênea é degradada por hemicelulases, que constituem um grande grupo de enzimas hidrolíticas e não

hidrolíticas com diferentes especificidades que funcionam cooperativamente. Elas compreendem várias enzimas de ação endo e exo que atacam a cadeia principal da hemicelulose, além de uma variedade de enzimas desramificadoras, por exemplo, α-arabinofuranosidase, ácido ferúlico esterase, acetil xilano esterase e α-glucuronidase que clivam diferentes grupos ligados à cadeia principal. A ação de pectinases (poligalacturonase, pectina esterase e pectato liase) na remoção de substâncias pécticas associadas à celulose na estrutura da parede celular são necessárias em algumas biomassas ricas em pectina (Moreira & Filho, 2016). A **Figura 10** mostra resumidamente a ação de enzimas na degradação da lignocelulose.

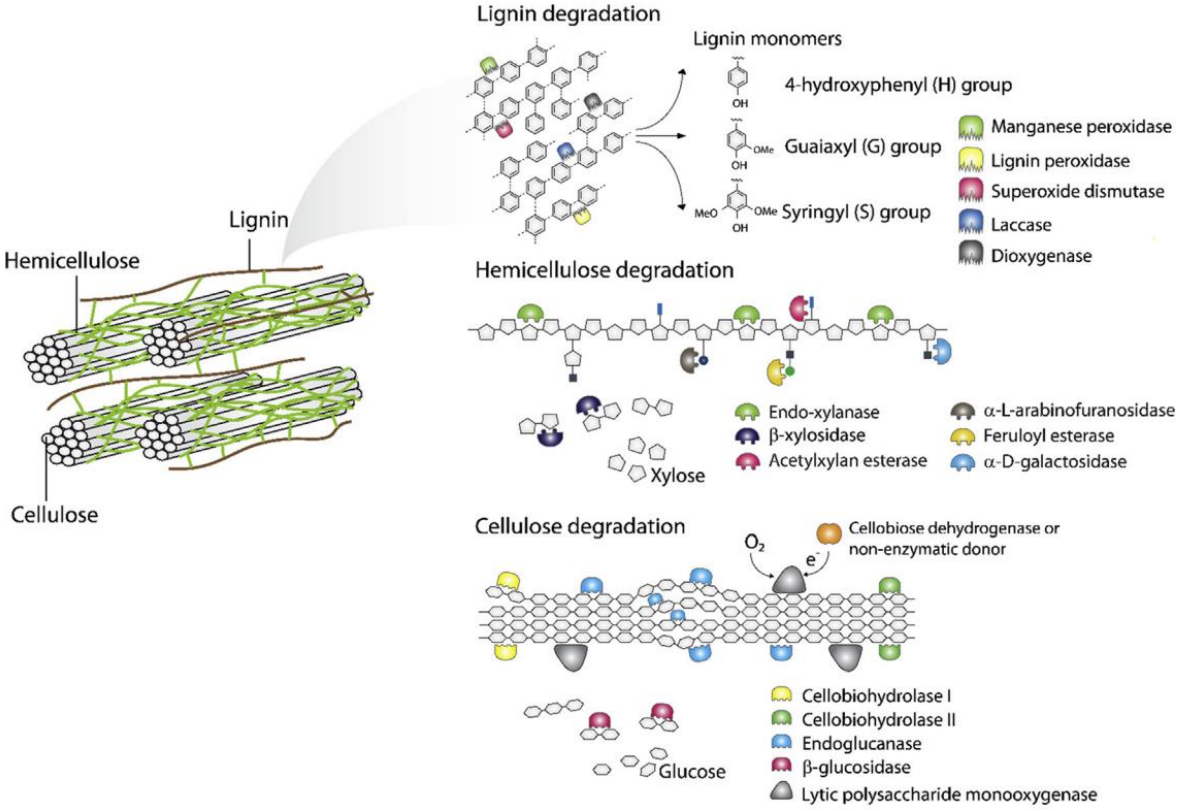


Figura 10. Diagramas esquemáticos da degradação da lignocelulose pelas ações sinérgicas das glicosil hidrolases e atividades auxiliares. Componentes não enzimáticos auxiliares trabalham para criar locais mais acessíveis para celulases centrais com glicosil hidrolases hidrolíticas acessórias e enzimas de ramificação, que atuam cooperativamente para a despolimerização de lignocelulose (lignina, hemicelulose e polímeros de celulose) (Champreda et al., 2019).

1.4.1 Celulases

A celulose é um polímero cristalino não ramificado, que recebe a ação de várias celulases para degradá-la eficientemente. Elas hidrolisam as ligações β-1,4-D-glucana, liberando celo-oligossacarídeos, celobiose ou glicose. A degradação completa da celulose é realizada por um complexo enzimático, que inclui endo-β-1,4-glucanases (EG EC 3.2.1.4), celobiohidrolases (CBH EC 3.2.1.91 e EC 3.2.1.176) e β- glicosidases (BGL EC 3.2.1.21) (Contreras et al., 2020; Bajpai, 2023).

As endoglucanases hidrolisam a cadeia de glicose internamente, liberando celo-oligossacarídeos e expondo extremidades livres redutoras e não redutoras adicionais. As CBHs hidrolisam as extremidades expostas e liberam celobiose, um forte inibidor de CBH, que posteriormente será convertido em monômeros de glicose. Por outro lado, as CBHs tornam as regiões cristalinas da matriz de celulose em amorfa, o que facilita o acesso ao substrato para as EGs. E finalmente, as β-glicosidases clivam as celobiose gerando monômeros de glicose. Assim, a degradação da celulose é influenciada pela ação sinérgica das enzimas descritas (Glasgow et al., 2021; Bajpai, 2023).

A celulase é muito importante na pesquisa acadêmica e industrial devido às suas amplas aplicações. Está em alta demanda e cobre aproximadamente 20% do mercado global de enzima. Elas são muito importante do ponto de vista da aplicação industrial, pois são usadas em várias indústrias, como papel e celulose, lavanderia, têxtil, detergentes, alimentos e rações, biocombustível, cervejaria e vinho, entre outras. Novazyme, DSM, DuPont e Amano Enzymes Inc, por exemplo,

são participantes proeminentes na produção de enzimas celulases em todo o mundo. O tamanho do mercado global de enzimas foi avaliado em US$ 9,9 bilhões em 2019, US$ 10,6 bilhões em 2020 e a previsão de receita em 2027 será de US$ 14,9 bilhões com base na estimativa de 2019 (https://www.grandviewresearch.com/industry-analysis/enzymes-industry). Poucas indústrias estão vendendo sua formulação de produtos em domínio aberto, enquanto a maioria está ocultando/ou não divulgando suas formulações de coquetéis de enzimas e vendendo sob licenças exclusivas (Jain & Agrawal, 2020; Singh et al., 2021).

Estima-se que o mercado global de celulase cresça a uma CAGR (*Compound annual growth rate* - taxa de crescimento anual composto) de 6,30% durante o período de previsão (2023-2030). Espera-se que a crescente demanda por celulase na indústria de alimentos e bebidas para produção de sucos de frutas, cerveja e pão impulsione o crescimento do mercado. Além disso, espera-se que o crescimento nas indústrias têxtil e de papel e celulose contribua para o crescimento do mercado. O tamanho do mercado foi avaliado em US$ 2,00 bilhões em 2022 e deve chegar a US$ 3,10 milhões em 2030. A América do Norte deteve a maior participação de mercado em 2020, devido à alta demanda na indústria de alimentos e bebidas (Patil, 2023).

1.5 Produção de enzimas por expressão heteróloga

Uma proteína codificada por DNA recombinante e expressa em um hospedeiro heterólogo é denominada proteína recombinante. A escolha de um hospedeiro heterólogo adequado para a expressão de genes heterólogos é uma etapa importante. Os pesquisadores tem usado diferentes organismos procarióticos e eucarióticos como hospedeiros heterólogos, por exemplo a *Escherichia coli*, *Bacillus subtillis*, *Pseudomonas fluorescens*, fungos filamentosos, leveduras, células de insetos, células de mamíferos e células vegetais (Young et al., 2012; Pouresmaeil & Azizi-Dargahlou, 2023).

A seleção de sistemas de expressão adequados depende de muitos critérios, como as instalações do laboratório e experiência local, custo de produção final, as características da proteína recombinante, a característica intrínseca do hospedeiro de expressão, a fonte do gene heterólogo, fácil otimização da expressão heteróloga, considerações regulatórias, o objetivo da produção de proteína recombinante (Pouresmaeil & Azizi-Dargahlou, 2023).

De acordo com o site da *marketsandmarkets* https://www.marketsandmarkets.com/, a produção de proteínas recombinantes em células de mamíferos teve o maior rendimento entre outros sistemas em 2021, e o rendimento deste sistema foi igual a 41,70% da receita total de proteínas recombinantes. Além disso, a expressão baseada nos sistemas bacterianos ficou em segundo lugar em termos de rendimento (Karyolaimos et al. 2019; Pouresmaeil & Azizi-Dargahlou, 2023) **Figura 11**.

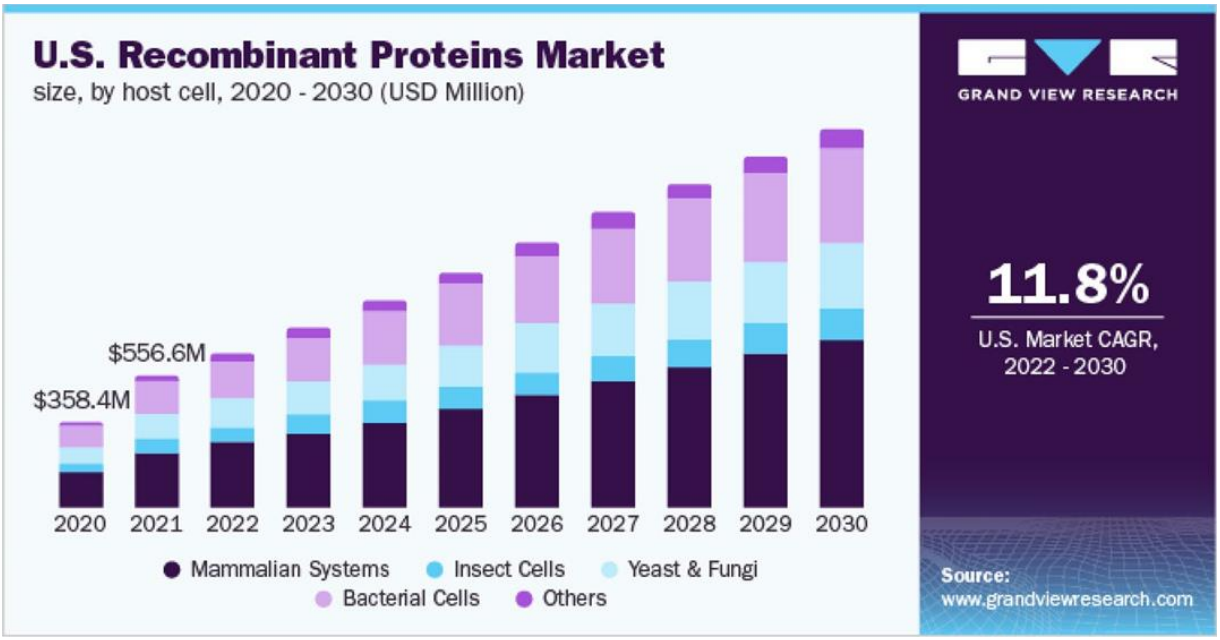


Figura 11. Comparação de hospedeiros heterólogos por renda. No total, a receita de proteínas recombinantes que são produzidas em todos os hospedeiros heterólogos é expressiva, alcançando entre 2020 e 2021 o valor de 358,4 e 556,6 milhões de dólares, respectivamente. Além disso, o sistema de expressão baseado em bactérias é o segundo mais lucrativo depois das células de mamíferos (Karyolaimos et al., 2019).

No entanto, desvantagens como alto custo e rendimento proteico relativamente baixo são mais comuns em sistemas de expressão de mamíferos do que em sistemas de expressão procarióticos. Os sistemas de expressão procariótica são hospedeiros adequados para a expressão de proteínas recombinantes, e esses organismos são conhecidos como hospedeiros acessíveis e mais simples. O sistema de expressão *E. coli* é o sistema de expressão bacteriana mais amplamente utilizado e é a primeira escolha para pesquisa de laboratório (Koehn & Hunt, 2009; Portolano et al. 2014; Pouresmaeil & Azizi-Dargahlou, 2023).

Várias cepas de *E. coli* são usadas para produção de proteína recombinante, dependendo das propriedades intrínsecas da proteína de interesse. *E. coli* K12 e BL21 (DE3) e muitos de seus derivados são os hospedeiros mais comumente usados

para expressão heteróloga. As cepas de *E. coli* diferem com base em características específicas adquiridas por meio da engenharia genética de seus genomas.

2.0 JUSTIFICATIVA

Atualmente, o grande problema enfrentado pela humanidade advém principalmente da escassez de formas convencionais de energia e o consequente impacto na sustentabilidade ambiental. Portanto, a população global, que passará de 8.03 bilhões em 2023 para 9.01 em 2037, começou a recuperar a forma alternativa de energia da biomassa lignocelulósica, como resíduos agrícolas e florestais. A biomassa lignocelulósica derivada de resíduos é uma alternativa às matérias-primas tradicionais à base de petróleo. A quantidade anual global de biomassa lignocelulósica obtida de resíduos como palhas e lascas de madeira pode chegar a 150 bilhões de toneladas (Ng et al., 2020; Vigneswari et al., 2021; Worldometers, 2023).

Essa ampla quantidade de rejeitos lignocelulósicos produzidos anualmente podem ser utilizados para a obtenção de produtos amplamente consumidos pela humanidade, e dessa forma contribuir para uma economia mundial verde, quando a utilização de fontes não-renováveis será mínima. O processo sustentável e ecológico é necessário para um planeta mais saudável, contudo, o investimento de tecnologias promissoras e que tragam resultados satisfatórios para competir com as atuais se faz necessário. Nesse cenário as enzimas celulases são peça fundamental, já que participam da bioconversão de lignocelulose em moléculas que podem ser utilizadas nas mais diversas aplicações industriais: biocombustíveis,

alimentação animal e humana, farmacêutica, biorremediação, produção de biopolímeros, etc (Shabbirahmed et al., 2022).

Essas enzimas podem ser obtidas de diferentes fontes, desde insetos, ruminantes, plantas a microrganismos, por exemplo. Nesse trabalho foi realizada uma bioprospecção metagenômica no intestino do cupim *Syntermes wheeleri*, espécie endêmica do cerrado brasileiro. Essa espécie é conhecida por ter uma dieta baseada em folhas secas do solo (serapilheira), e por fazer parte dos cupins superiores, os quais possuem bactérias responsáveis em seu trato intestinal para degradar esse material ingerido (Scharf, 2020).

Visto a capacidade dos cupins em degradar lignocelulose (3 a 7 bilhões de toneladas por ano) junto com a ausência de trabalhos nessa área, com essa espécie em específico, e o amplo mercado de celulases que chegou a US$ 2,00 bilhões em 2022, pensamos em minerar enzimas a partir do metagenoma bacteriano do intestino de *S. wheeleri* que fossem capazes de agir na desconstrução da lignocelulose, e dessa forma caracterizá-las a fim de ver seu potencial biotecnológico de aplicação (Tokuda et al., 2014; Patil, 2023).

3.0 OBJETIVOS

Objetivo geral

O presente trabalho tem por objetivo geral a produção e a caracterização de enzimas bioconversoras de lignocelulose a partir do metagenoma bacteriano do intestino do cupim *Syntermes wheeleri*, utilizando *Escherichia coli* como sistema de expressão heteróloga, e análise do potencial biotecnológico.

Objetivos específicos

a) Identificar a participação das Glicosil hidrolases nos proctodeal 1 e 3 de *Syntermes wheeleri*;

b) Selecionar enzimas celulases no banco de dados ggKbase por meio do *Enzyme comission*;

c) Realizar análises filogenéticas e metagenômicas a fim de selecionar enzimas com potencial biotecnológico;

d) Produzir as proteínas por expressão heteróloga em *Escherichia coli* BL21(DE3) utilizando *pET-24 a* (+);

e) Purificar as proteínas por processo de Cromatografia;

f) Realizar a caracterização bioquímica com a utilização de substratos sintéticos, como pNPs;

g) Realizar a caracterização biofísica com o auxílio de um equipamento de Dicroísmo Circular;

4.0 REFERÊNCIAS BIBLIOGRÁFICAS

André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., & Remaud-Siméon, M. (2014). CAZyme discovery and design for sweet dreams. *Current opinion in chemical biology*, *19*, 17–24. https://doi.org/10.1016/j.cbpa.2013.11.014

Atiwesh, G., Parrish, C. C., Banoub, J., & Le, T. T. (2022). Lignin degradation by microorganisms: A review. *Biotechnology progress*, *38*(2), e3226. https://doi.org/10.1002/btpr.3226

Baghaei, B., & Skrifvars, M. (2020). All-Cellulose Composites: A Review of Recent Studies on Structure, Properties and Applications. *Molecules (Basel, Switzerland)*, *25*(12), 2836. https://doi.org/10.3390/molecules25122836

Baghaei, B., & Skrifvars, M. (2020). All-Cellulose Composites: A Review of Recent Studies on Structure, Properties and Applications. *Molecules (Basel, Switzerland)*, *25*(12), 2836. https://doi.org/10.3390/molecules25122836

Bajpai, P. (2023). Cellulases in the biofuel industry. *Elsevier*.

Bansal M., Tiwari N., Sharma J. G. (2022). “Chapter 3 - revolution in microbial bioprospecting via the development of omics-based technologies,” in *Bioprospecting of microbial diversity*. Editors Verma P., Shah M. P. (Netherlands: Elsevier;)

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Beattie, A. J., Hay, M., Magnusson, B., de Nys, R., Smeathers, J., & Vincent, J. F. (2011). Ecology and bioprospecting. *Austral ecology*, *36*(3), 341–356. https://doi.org/10.1111/j.1442-9993.2010.02170.x

Becker, J., & Wittmann, C. (2020). Microbial production of extremolytes - high-value active ingredients for nutrition, health care, and well-being. *Current opinion in biotechnology*, *65*, 118–128. https://doi.org/10.1016/j.copbio.2020.02.010

Brethauer, S., Shahab, R. L., & Studer, M. H. (2020). Impacts of biofilms on the conversion of cellulose. *Applied microbiology and biotechnology*, *104*(12), 5201–5212. https://doi.org/10.1007/s00253-020-10595-y

Brune A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nature reviews. Microbiology*, *12*(3), 168–180. https://doi.org/10.1038/nrmicro3182

Champreda, V., Mhuantong, W., Lekakarn, H., Bunterngsook, B., Kanokratana, P., Zhao, X. Q., Zhang, F., Inoue, H., Fujii, T., & Eurwilaichitr, L. (2019). Designing cellulolytic enzyme systems for biorefinery: From nature to application. *Journal of bioscience and bioengineering*, *128*(6), 637–654. https://doi.org/10.1016/j.jbiosc.2019.05.007

Chettri, D., Verma, A. K., & Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology reports (Amsterdam, Netherlands)*, *28*, e00525. https://doi.org/10.1016/j.btre.2020.e00525

Constantino, R. (1995). Revision of the Neotropical termite genus *Syntermes Holmgren* (Isoptera: Termitidae). *The University of Kansas Science Bulletin*, *55*, 455–518. https://www.biodiversitylibrary.org/part/781

Contreras, F., Pramanik, S., M. Rozhkova, A., N. Zorov, I., Korotkova, O., P. Sinitsyn, A., Schwaneberg, U., & D. Davari, M. (2020). Engineering Robust Cellulases for Tailored Lignocellulosic Degradation Cocktails. *International Journal of Molecular Sciences*, *21*(5), 1589. https://doi.org/10.3390/ijms21051589

Corander, J., Hanage, W. P., & Pensar, J. (2022). Causal discovery for the microbiome. *The Lancet. Microbe*, *3*(11), e881–e887. https://doi.org/10.1016/S2666-5247(22)00186-0

Culligan, E. P., Sleator, R. D., Marchesi, J. R., & Hill, C. (2014). Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence*, *5*(3), 399–412. https://doi.org/10.4161/viru.27208

Curtis, T. P., & Sloan, W. T. (2005). Microbiology. Exploring microbial diversity--a vast below. *Science (New York, N.Y.)*, *309*(5739), 1331–1333. https://doi.org/10.1126/science.1118176

Dai, J., Patti, A. F., & Saito, K. (2016). Recent developments in chemical degradation of lignin: catalytic oxidation and ionic liquids. *Tetrahedron Letters*, *57*(45), 4945-4951. https://doi.org/10.1016/j.tetlet.2016.09.084

Dassanayake, R. S., Acharya, S., & Abidi, N. (2018). Biopolymer-based materials from polysaccharides: Properties, processing, characterization and sorption applications. *Advanced sorption process applications*, 1-24.

Evangelista, D. A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M. K., Legendre, F., Liu, S., Machida, R., Misof, B., Peters, R. S., Podsiadlowski, L., Rust, J., Schuette, K., Tollenaar, W., Ware, J. L., Wappler, T., Zhou, X., Meusemann, K., & Simon, S. (2019). An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proceedings. Biological sciences*, *286*(1895), 20182076. https://doi.org/10.1098/rspb.2018.2076

Glasgow, E., Vander Meulen, K., Kuch, N., & Fox, B. G. (2021). Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy. *Current opinion in biotechnology*, *67*, 141–148. https://doi.org/10.1016/j.copbio.2020.12.020

Heinze T., Liebert T. (2012). Celluloses and Polyoses/Hemicelluloses. *Polymer Science: A Comprehensive Reference*. Elsevier. 83–152.

https://www.worldometers.info/world-population/world-population-projections/

Hussain, N., Baqar, Z., Mumtaz, M., El-Sappah, A. H., Show, P. L., Iqbal, H. M., ... & Bilal, M. (2022). Bioprospecting fungal-derived value-added bioproducts for sustainable pharmaceutical applications. *Sustainable Chemistry and Pharmacy*, *29*, 100755. https://doi.org/10.1016/j.scp.2022.100755

Jain, L., & Agrawal, D. (2020). Biofuel cellulases: diversity, distribution and industrial outlook. In *Microbial Fermentation and Enzyme Technology* (pp. 283-298). CRC Press.

Karyolaimos, A., Ampah-Korsah, H., Hillenaar, T., Mestre Borras, A., Dolata, K. M., Sievers, S., Riedel, K., Daniels, R., & de Gier, J. W. (2019). Enhancing Recombinant Protein Yields in the *E. coli* Periplasm by Combining Signal Peptide and Production

Rate Screening. *Frontiers in microbiology*, *10*, 1511. https://doi.org/10.3389/fmicb.2019.01511

Keogh, G. F., Cooper, G. J., Mulvey, T. B., McArdle, B. H., Coles, G. D., Monro, J. A., & Poppitt, S. D. (2003). Randomized controlled crossover study of the effect of a highly beta-glucan-enriched barley on cardiovascular disease risk factors in mildly hypercholesterolemic men. *The American journal of clinical nutrition*, *78*(4), 711–718. https://doi.org/10.1093/ajcn/78.4.711

Klemm, D., Heublein, B., Fink, H. P., & Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie (International ed. in English)*, *44*(22), 3358–3393. https://doi.org/10.1002/anie.200460587

Koehn, J., & Hunt, I. (2009). High-Throughput Protein Production (HTPP): a review of enabling technologies to expedite protein production. *Methods in molecular biology (Clifton, N.J.)*, *498*, 1–18. https://doi.org/10.1007/978-1-59745-196-3_1

Li, H., & Greening, C. (2022). Termite-engineered microbial communities of termite nest structures: a new dimension to the extended phenotype. *FEMS microbiology reviews*, *46*(6), fuac034. https://doi.org/10.1093/femsre/fuac034

Liu, H., Sun, J., Leu, S. Y., & Chen, S. (2016). Toward a fundamental understanding of cellulase-lignin interactions in the whole slurry enzymatic saccharification process. *Biofuels, Bioproducts and Biorefining*, *10*(5), 648-663. https://doi.org/10.1002/bbb.1670

Liu, N., Li, H., Chevrette, M. G., Zhang, L., Cao, L., Zhou, H., Zhou, X., Zhou, Z., Pope, P. B., Currie, C. R., Huang, Y., & Wang, Q. (2019). Functional metagenomics reveals

abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *The ISME journal*, *13*(1), 104–117. https://doi.org/10.1038/s41396-018-0255-1

Liu, Y. J., Li, B., Feng, Y., & Cui, Q. (2020). Consolidated bio-saccharification: Leading lignocellulose bioconversion into the real world. *Biotechnology advances*, *40*, 107535. https://doi.org/10.1016/j.biotechadv.2020.107535

Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5970–5975. https://doi.org/10.1073/pnas.1521291113

Lok C. (2015). Mining the microbial dark matter. *Nature*, *522*(7556), 270–273. https://doi.org/10.1038/522270a

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*, *42*(Database issue), D490–D495. https://doi.org/10.1093/nar/gkt1178

López-Mondéjar, R., Algora, C., & Baldrian, P. (2019). Lignocellulolytic systems of soil bacteria: A vast and diverse toolbox for biotechnological conversion processes. *Biotechnology advances*, *37*(6), 107374. https://doi.org/10.1016/j.biotechadv.2019.03.013

Moreira, L. R., & Filho, E. X. (2016). Insights into the mechanism of enzymatic hydrolysis of xylan. *Applied microbiology and biotechnology*, *100*(12), 5205–5214. https://doi.org/10.1007/s00253-016-7555-z

Muddasar, M., Beaucamp, A., Culebras, M., & Collins, M. N. (2022). Cellulose: Characteristics and applications for rechargeable batteries. *International journal of biological macromolecules*, *219*, 788–803. https://doi.org/10.1016/j.ijbiomac.2022.08.026

Ng, H. S., Kee, P. E., Yim, H. S., Chen, P. T., Wei, Y. H., & Chi-Wei Lan, J. (2020). Recent advances on the sustainable approaches for conversion and reutilization of food wastes to valuable bioproducts. *Bioresource technology*, *302*, 122889. https://doi.org/10.1016/j.biortech.2020.122889

Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., & Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *Journal of microbiology and biotechnology*, *22*(4), 462–469. https://doi.org/10.4014/jmb.1108.08037

Oliveira, M. H., Viana-Junior, A. B., Rolim, M. D. S. L., Eloi, I., Andrade, M. R. L., Souza, J. J. L. L., & Bezerra-Gusmão, M. A. (2022). The Impact of *Constrictotermes cyphergaster* (Termitidae: Nasutitermitinae) Termites on Semiarid Ecosystems in Brazil: A Review of Current Research. *Insects*, *13*(8), 704. https://doi.org/10.3390/insects13080704

Paoli, L., Ruscheweyh, H. J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., Lotti, A., Clayssen, Q., Salazar, G., Milanese, A., Carlström, C. I., Papadopoulou, C., Gehrig, D., Karasikov, M., Mustafa, H., Larralde, M., Carroll, L. M., Sánchez, P., Zayed, A. A., Cronin, D. R., … Sunagawa, S. (2022). Biosynthetic potential of the global ocean microbiome. *Nature*, *607*(7917), 111–118. https://doi.org/10.1038/s41586-022-04862-3

Patil, M. (2023, june 29). *Global Market Analysis on Cellulase market, Chlorogenic Acid market, Spandex Fibermarket forecasted till 2030*. Recuperado de https://www.einnews.com/pr_news/642111523/global-market-analysis-on-cellulase-market-chlorogenic-acid-market-spandex-fibermarket-forecasted-till-2030.

Paulino, B. N., Pessôa, M. G., Molina, G., Kaupert Neto, A. A., Oliveira, J. V. C., Mano, M. C. R., & Pastore, G. M. (2017). Biotechnological production of value-added compounds by ustilaginomycetous yeasts. *Applied microbiology and biotechnology*, *101*(21), 7789–7809. https://doi.org/10.1007/s00253-017-8516-x

Pauly, M., & Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant journal : for cell and molecular biology*, *54*(4), 559–568. https://doi.org/10.1111/j.1365-313X.2008.03463.x

Perez Rojo, F., Pillow, J. J., & Kaur, P. (2023). Bioprospecting microbes and enzymes for the production of pterocarpans and coumestans. *Frontiers in bioengineering and biotechnology*, *11*, 1154779. https://doi.org/10.3389/fbioe.2023.1154779

Pollegioni, L., Tonin, F., & Rosini, E. (2015). Lignin-degrading enzymes. *The FEBS journal*, *282*(7), 1190-1213. https://doi.org/10.1111/febs.13224

Pouresmaeil, M., & Azizi-Dargahlou, S. (2023). Factors involved in heterologous expression of proteins in E. coli host. *Archives of microbiology*, *205*(5), 212. https://doi.org/10.1007/s00203-023-03541-9

Pu, Y., Zhang, D., Singh, P. M., & Ragauskas, A. J. (2008). The new forestry biofuels sector. *Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy*, *2*(1), 58-73. https://doi.org/10.1002/bbb.48

Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H., & Linington, R. G. (2017). Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(22), 5601–5606. https://doi.org/10.1073/pnas.1614680114

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Roux, S., & Emerson, J. B. (2022). Diversity in the soil virosphere: to infinity and beyond?. *Trends in microbiology*, *30*(11), 1025–1035. https://doi.org/10.1016/j.tim.2022.05.003

Santana, R. H., Catão, E. C., Lopes, F. A., Constantino, R., Barreto, C. C., & Krüger, R. H. (2015). The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. *Microbial ecology*, *70*(2), 545–556. https://doi.org/10.1007/s00248-015-0581-z

Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual review of plant biology*, *61*, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315

Shabbirahmed, A. M., Haldar, D., Dey, P., Patel, A. K., Singhania, R. R., Dong, C. D., & Purkait, M. K. (2022). Sugarcane bagasse into value-added products: a

review. *Environmental science and pollution research international*, *29*(42), 62785–62806. https://doi.org/10.1007/s11356-022-21889-1

Shaghaleh, H., Xu, X., & Wang, S. (2018). Current progress in production of biopolymeric materials based on cellulose, cellulose nanofibers, and cellulose derivatives. *RSC advances*, *8*(2), 825–842. https://doi.org/10.1039/c7ra11157f

Shirkavand, E., Baroutian, S., Gapes, D. J., & Young, B. R. (2016). Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment–A review. *Renewable and Sustainable Energy Reviews*, *54*, 217-234. https://doi.org/10.1016/j.rser.2015.10.003

Singh, A., Bajar, S., Devi, A., & Pant, D. (2021). An overview on the recent developments in fungal cellulase production and their industrial applications. *Bioresource Technology Reports*, *14*, 100652. https://doi.org/10.1016/j.biteb.2021.100652

Singhvi, M., & Kim, B. S. (2021). Lignin valorization using biological approach. *Biotechnology and applied biochemistry*, *68*(3), 459–468. https://doi.org/10.1002/bab.1995

Strnad, S., & Zemljič, L. F. (2023). Cellulose-Chitosan Functional Biocomposites. *Polymers*, *15*(2), 425. https://doi.org/10.3390/polym15020425

Tarrsini, M., Ng, Q. H., Teoh, Y. P., Shuit, S. H., Ooi, Z. X., & Kunasundari, B. (2021). Structural and composition modification of Harum Manis mango (Mangifera indica) leaves via chemical pretreatment for bioethanol production. *Biomass Conversion and Biorefinery*, 1-13. https://doi.org/10.1007/s13399-021-01469-y

Thorn, C. S., Maness, R. W., Hulke, J. M., Delmore, K. E., & Criscione, C. D. (2023). Population genomics of helminth parasites. *Journal of helminthology*, *97*, e29. https://doi.org/10.1017/S0022149X23000123

Vanholme, R., De Meester, B., Ralph, J., & Boerjan, W. (2019). Lignin biosynthesis and its integration into metabolism. *Current opinion in biotechnology*, *56*, 230–239. https://doi.org/10.1016/j.copbio.2019.02.018

Vigneswari, S., Noor, M. S. M., Amelia, T. S. M., Balakrishnan, K., Adnan, A., Bhubalan, K., Amirul, A. A., & Ramakrishna, S. (2021). Recent Advances in the Biosynthesis of Polyhydroxyalkanoates from Lignocellulosic Feedstocks. *Life (Basel, Switzerland)*, *11*(8), 807. https://doi.org/10.3390/life11080807

Wang, W., Tan, X., Imtiaz, M., Wang, Q., Miao, C., Yuan, Z., & Zhuang, X. (2021). Rice straw pretreatment with KOH/urea for enhancing sugar yield and ethanol production at low temperature. *Industrial Crops and Products*, *170*, 113776. https://doi.org/10.1016/j.indcrop.2021.113776

Wardman, J. F., Bains, R. K., Rahfeld, P., & Withers, S. G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nature reviews. Microbiology*, *20*(9), 542–556. https://doi.org/10.1038/s41579-022-00712-1

Worldometers *World Population Projections*. *(Julho de* 2022).

Wu, Z., Peng, K., Zhang, Y., Wang, M., Yong, C., Chen, L., Qu, P., Huang, H., Sun, E., & Pan, M. (2022). Lignocellulose dissociation with biological pretreatment towards the biochemical platform: A review. *Materials today. Bio*, *16*, 100445. https://doi.org/10.1016/j.mtbio.2022.100445

Xiang, Z., Tang, N., Jin, X., & Gao, W. (2022). Fabrications and applications of hemicellulose-based bio-adsorbents. *Carbohydrate polymers*, *278*, 118945. https://doi.org/10.1016/j.carbpol.2021.118945

Yoo, C. G., Meng, X., Pu, Y., & Ragauskas, A. J. (2020). The critical role of lignin in lignocellulosic biomass conversion and recent pretreatment strategies: A comprehensive review. *Bioresource technology*, *301*, 122784. https://doi.org/10.1016/j.biortech.2020.122784

Young, C. L., Britton, Z. T., & Robinson, A. S. (2012). Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology journal*, *7*(5), 620–634. https://doi.org/10.1002/biot.201100155

Yuan, Q., Liu, S., Ma, M. G., Ji, X. X., Choi, S. E., & Si, C. (2021). The Kinetics Studies on Hydrolysis of Hemicellulose. *Frontiers in chemistry*, *9*, 781291. https://doi.org/10.3389/fchem.2021.781291

Zhao, Y., Moser, C., Lindström, M. E., Henriksson, G., & Li, J. (2017). Cellulose Nanofibers from Softwood, Hardwood, and Tunicate: Preparation-Structure-Film Performance Interrelation. *ACS applied materials & interfaces*, *9*(15), 13508–13519. https://doi.org/10.1021/acsami.7b01738

Zhou, Y., Staver, A. C., & Davies, A. B. (2023). Species-level termite methane production rates. *Ecology*, *104*(2), e3905. https://doi.org/10.1002/ecy.3905

Zoghlami, A., & Paës, G. (2019). Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. *Frontiers in chemistry*, *7*, 874. https://doi.org/10.3389/fchem.2019.00874

Capítulo 2

The microbial metabolism in the gut of the litter-feeding termite *Syntermes wheeleri* (Isoptera: Termitidae) reveals functional compartmentalization of Glycosyl hydrolases[1]

[1] Manuscrito em revisão para ser submetido em periódico cientifico.

1.0 Abstract

Termite ecosystem function and evolutionary history are tightly bound to symbiosis with microbes, which has transformed the species into an important player in carbon cycling in tropical areas. *Syntermes wheeleri* (Termitidae: Syntermitinae) is a litter-feeding termite widespread in the Brazilian Cerrado. The guts of *S. wheeleri* workers possess a unique microbiota and peculiar anatomy: the first proctodeal segment (P1) is more voluminous than the third proctodeal segment (P3), which is the most dilated segment in most termites. To better understand the role of the enlarged P1 segment in microbial community assembly and function, we used metagenomics and metatranscriptomics to describe microbial metabolism in the whole gut of *S. wheeleri*, especially the ones involved in carbohydrates breakdown, the CAZymes, as well as the P1 and P3 compartments. Our results showed a vast microbial metabolic repertoire and the potential to assist in lignocellulose digestion, with GH diversification, and supply nutrients to the host. These findings add to our understanding of factors governing gut microbiota establishment in termites, which appear to select microbes based on functional groups and efficiency in degrading food substrates under gut conditions, rather than microbial lineages.

2.0 Introduction

The evolutionary history, diversity, and niche occupation of termites place them amongst the most successful animals on Earth (Mayhew, 2007). Their gut microbiota is one of the main factors responsible for this success, with strong evidence demonstrating the importance of termite gut microbes in niche diversification (Inward et al., 2007). The functions of termite gut microbes range from providing the host with nutrients (e.g., carbon, nitrogen, vitamins) to defense against pathogens (Bignell, 2011). To provide a suitable environment for such a diverse range of microbial activity, the hindgut is divided into the following segments, which possess different physicochemical properties: first and third proctodeal segments (P1 and P3, respectively), enteric valve, colon, and rectum. The P3 segment (paunch) is dilated and capable of hosting an enormous number of microbial cells (Noirot, 1995, 2001; Bignell, 2011). In some species, primarily soil and litter feeders, P1 is also dilated and densely populated by microbes (Noirot, 2001).

The main drivers of community structure in termite gut microbes are host phylogeny and diet (Abdul Rahman et al., 2015; Mikaelyan et al., 2015). The microbiota also varies according to gut compartment, likely because the physicochemical properties of each gut segment selects for different microbes (Brune, 1998; Köhler et al., 2012; Rossmassler et al., 2015; Mikaelyan et al., 2017).

Recently, Mikaelyan et al. (2017) demonstrated that microbial communities from the same gut compartment of different termite species within the same feeding guild are more similar to each other than to microbial communities from different gut compartments within the same species. A similar trend was observed in

functional potential of the microbiota in a study by Rossmassler et al. (2015), where the microbiota in gut compartments of different species within the same feeding guild clustered apart from the microbiota in other gut compartments of the same host. Both studies noted that among the exceptions was *Cornitermes* sp., a litter-feeding termite of Syntermitinae (Termitidae). A striking characteristic of Syntermitinae gut anatomy is the large P1 compartment, which led Noirot (2001) to propose this group as a separate clade from Nasutitermitinae.

In a previous work we described the gut microbiota composition of *Syntermes wheeleri* (Isoptera: Termitidae, Syntermitinae), another litter-feeding termite of the Syntermitinae subfamily (Santana et al., 2015). The microbiota of *S. wheeleri* also differs markedly compared to that of other species, containing a high proportion (>70%) of Firmicutes (Santana et al., 2015).

Since the reported patterns of gut microbial community structure and function of Syntermitinae species differed between previous studies, it is crucial to examine members of this clade thoroughly to better understand its evolution. Here, we take another step towards unraveling the ecosystem function of *S. wheeleri* gut microbiota. We describe its metabolism using high-throughput metagenome and metatranscriptome sequencing of the microbiota in the whole gut and in the largest compartments, P1 and P3. Considering the order of food passage through the gut and the greater P1 volume, we hypothesized that (1) microbial metabolism in P1 specializes in the digestion of complex molecules and in the defense against pathogens and toxins; and (2) microbial metabolism in P3 specializes in processing the simpler molecules yielded by P1.

3.0 Material and Methods

3.1 Sampling, processing, and sequencing

S. wheeleri workers were sampled in a pristine Cerrado area (cerrado *stricto sensu)* near a conservation unit (-15.636149, -47.856199) from three different mounds (i.e., three biological replicates). Immediately after sampling (≤ 3 hours), termites were surface-sterilized with 70% ethanol and degutted with sterile forceps. Segments P1 and P3 were separated with sterile blades and placed in microcentrifuge tubes with LifeGuard Soil Preservation Solution (MO BIO). The co-extraction of RNA and DNA was performed with the RNA PowerSoil Total RNA Isolation Kit and DNA Elution Accessory Kit (MO BIO) following the manufacturer's instructions. To obtain sufficient material and diminish biases (e.g., differences in processing time between samples and replicates, and metabolic state of the termites), samples were taken on 3 different days (during May and June 2016) at different times in the morning.

The DNA samples were paired-end (2×101 bp) sequenced by Macrogen Inc. (Seoul, Korea) in one lane of the Illumina HiSeq 2000 sequencer after library preparation with TruSeq SBS Kit v3. RNA samples were cleaned with RNeasy PowerClean Pro Cleanup Kit (MO BIO) and treated with TURBO DNA-free™ Kit (Thermo Fisher Scientific) according to the provided protocol. The samples were paired-end (2×101 bp) sequenced in one lane of the Illumina HiSeq 2500 sequencer after bacterial rRNA removal with the Illumina Ribo-Zero rRNA Removal Kit (Bacteria), followed by library preparation with the TruSeq RNA Sample Prep Kit v2, and sequenced with TruSeq Rapid SBS Kit v4.

3.2 Sequence analysis

The pipeline used in this study for sequencing analysis was: (1) quality filtering and removal of contaminant sequences, (2) scaffold assembly, (3) read mapping to assembled scaffolds, (4) gene prediction, and (5) annotation and prediction of metabolic pathways. The raw sequence files of the metagenome and the metatranscriptome (32 data sets) were submitted to the NCBI Short Read Archive (accession number SRP145533).

3.3 Quality filtering and assembly

Sequencing reads were preprocessed to remove the last base and contaminant sequences with BBDuk (BBMap - Bushnell B. - sourceforge.net/projects/bbmap/ - version 36.28) and quality filtered with Sickle (https://github.com/najoshi/sickle) using default parameters. Contigs were assembled with IDBA-UD using default parameters. Scaffolds >1000 base pairs were annotated and submitted to ggKbase repository (http://ggkbase.berkeley.edu), where they were binned and taxonomically classified. Taxonomic classification was performed by BLAST alignment against UniRef (Suzek et al., 2007) in conjunction with an internal genome database. The taxonomy of a single scaffold was based on the lowest taxonomic rank that showed agreement across ≥50% of the proteins on the scaffold.

3.4 Annotation

We used Prodigal software (Hyatt et al., 2010) to predict CDSs (CoDing Sequence) in the assembled scaffolds. The predicted proteins were annotated

according to the method of Hug et al. (2013) using USEARCH (Edgar, 2010) for similarity searches in the KEGG, UniRef100 (Suzek et al., 2007), and UniProt (The UniProt, 2018) databases. Reciprocal best hits were determined by forward and reverse searches against the KEGG and UniRef100 databases and filtered with a minimum 300-bit score as the threshold. Searches following only the forward direction were filtered with a minimum 60-bit score as the threshold.

CAZyme domains were identified performing a similarity search in dbCAN database (Yin et al., 2012) using hmmscan (HMMER3 version 3.1; hmmer.org). The results were filtered based on e-value and size using the script *hmmscan-parser.sh* (http://csbl.bmb.uga.edu/dbCAN/download.php). The relative abundance of domains in each sample was calculated as domain count per total domains predicted.

3.5 Metabolic pathway prediction

Enzyme class (EC) numbers from each sample were retrieved from the features table at the ggKbase repository. Metabolic pathways were inferred based on EC number mapping against the MetaCyc database (http://metacyc.org/) (Caspi et al., 2016) using MinPath (Minimal set of Pathways) (Ye and Doak, 2009). MinPath is a parsimony approach to reconstruct metabolic pathways in a sample based on the minimal pathways necessary to explain the predicted enzymes. Hierarchical classification of pathways was carried out according to MetaCyc database ontology.

3.6 Mapping and coverage calculation

The resulting SAM file from bowtie2 mapping (Langmead and Salzberg, 2012) was converted to BAM and sorted using SAMtools (view and sort options, respectively) (Li et al., 2009). A BED file was built using the scaffold coordinates of each enzyme. The coverage of these locations was extracted from the mapping file using BEDtools (coverage –hist) (Quinlan and Hall, 2010). The coverage for each enzyme was then calculated based on the histogram as follows: sum[coverage* (bases with that coverage/gene length)]. The relative abundance of metabolic pathways was estimated based on the average coverage of the enzymes engaged in the pathway.

3.7 Comparison of functional potential between P1 and P3

The matrix with the relative abundances of the pathways in each sample (normalized by the number of reads per million reads) was loaded in R version 3.3.1 (R Development Core Team, 2010). Nonmetric multidimensional scaling was calculated using metaMDS based on the Bray-Curtis dissimilarity (Bray and Curtis, 1957) in package vegan in R (Oksanen et al. 2016). Output was plotted using ggplot2 (Wickham 2009).

GH family abundance in P1, P3, and WG samples were calculated with R (prcomp function) (R Development Core Team, 2010) and plotted using factoextra R package (Kassambara, Mundt 2016). The relative abundance of metabolic pathways in P1 and P3 was compared using STatistical Analysis of Metagenomic Profiles (STAMP) software (Parks et al., 2014). The two-groups comparison was carried out with the following parameters: Welch's non-parametric t-test (statistical

test), two-sided (type), DP: Welch's inverted 0.95 (confidence interval method), no multiple test correction, and p-value filter >0.05.

To test whether differences in the number of genome equivalents in each sample could have introduced biases in the comparisons of gene abundances between P1 and P3 compartments, the number of genome equivalents in each sample was estimated using MicrobeCensus (Nayfach and Pollard, 2015) and compared by Welch's two-sample t-test.

3.8 Metatranscriptomic analysis

3.8.1 Quality, rRNA, and contaminant filtering

Sequencing reads were preprocessed as previously described for the metagenomic sequences. Prokaryotic and eukaryotic rRNA were filtered from metatranscriptomic data with SortMeRNA (Kopylova et al., 2012) using multiple databases and default parameters.

3.8.2 Mapping and coverage calculation

The remaining reads were mapped against CDSs of the assembled metagenomic scaffolds with bowtie2 (Langmead and Salzberg, 2012) using default parameters. Prior to mapping, all CDSs of termite origin were removed. To identify these sequences, a BLAST search (Altschul et al., 1990) was performed against a database of all termite nucleotide sequences available at NCBI.

Coverage calculations and abundance estimates of active metabolic pathways in the *S. wheeleri* gut followed the steps described for metagenomic analysis.

4.0 Results

4.1 Taxonomy

Taxonomy assignment of predicted genes in the metagenome showed 73%–83% of genes from Bacteria, 0.5%–3% from Metazoa, 0.4%–1% from Archaea, and 0.04%–0.1% from Fungi. The bacterial community structure determined by 16S ribosomal RNA (rRNA) gene pyrosequencing (amplicon sequencing) (Santana et al., 2015) and whole genome sequencing (WGS) produced similar results (Fig. 1).

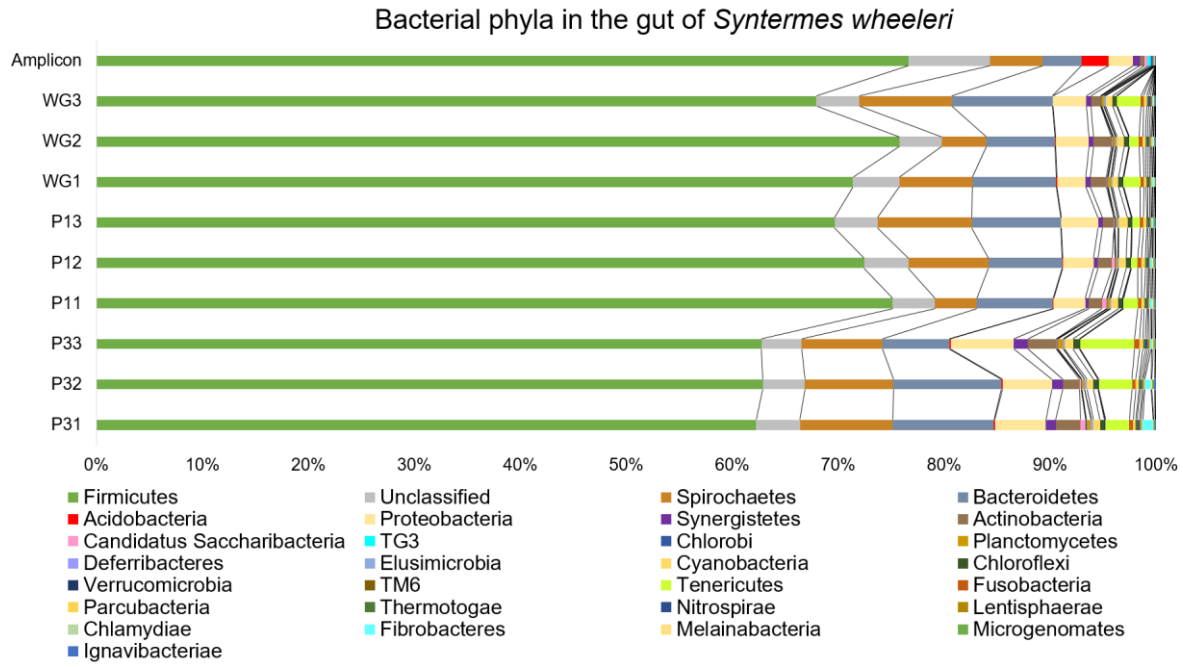


Fig 1. **Bacterial taxonomic profile of the gut microbiota metagenome from *Syntermes wheeleri***. The bacterial community structure was determined by 16S rRNA gene pyrosequencing (Amplicon) of the gut microbiota (from Santana *et al*., 2015) and whole genome sequencing of microbiota from the whole gut, first proctodeal segment (P1), and third proctodeal segment (P3), P11, P12 and P13 are biological replicates of P1 microbiota; P31, P32 and P33 are biological replicates of P3 microbiota; and WG1, WG2 and WG3 are biological replicates of whole gut microbiota. Only phyla with a relative abundance >0.1% are presented.

WGS data indicated that Firmicutes was the most abundant phylum in each sample, accounting for 67%–75% of the genes in whole gut (WG) samples, 69%–75% of the genes in P1 samples, and 62% of the genes in P3 samples. The second most abundant phylum varied across biological replicates (P11, P12, and P13 are

replicates of the P1 segment; P31, P32, and P33 are replicates of the P3 segment; WG1, WG2 and WG3 are replicates of the WG). Bacteroidetes was the second most abundant phylum in the P11, P31, P32, and WG samples, and Spirochaetes was the second most abundant in the P12, P13, and P33 samples. The third and fourth most abundant phyla were Proteobacteria (2.7%–4.8%), and Tenericutes (0.6%–5.1%). The relative abundances of these phyla were higher in P3 (4.8%–5.9% and 2.2%–5.1%, respectively) than in P1 (2.9%–3.4% and 0.6%–1.4%, respectively). In contrast, 16S rRNA data indicated that Acidobacteria and Proteobacteria were the third and fourth most abundant phyla, respectively. Other phyla accounted for <1% of genes, except for Actinobacteria (1.2%).

Most of the least abundant phyla (those accounting for <0.1% of genes) belong to the Candidate Phyla Radiation (CPR), recently described by Hug et al. (2016) and were detected only by WGS. The TG3 phylum was detected only by amplicon sequencing. The percentage of sequences for which taxonomy could not be assigned was 4% using WGS and 8% using amplicon sequencing.

4.2 Microbial metabolic potential

Metagenomic analysis of *S. wheeleri* gut microbiota showed higher abundances of metabolic pathways from the Biosynthesis and Degradation categories of the MetaCyc database in all samples. The most abundant Biosynthesis subcategories involved cofactor, amino acid, and secondary metabolite biosynthesis, and the most abundant Degradation subcategories involved carbohydrate, amino acid, inorganic nutrient, and nucleotide degradation.

In the Detoxification category, antibiotic resistance represented the most abundant pathways, which were more frequent in P1 than in P3 samples. In the Respiration subcategory, aerobic metabolism pathways were more abundant than anaerobic pathways in all samples except for P1 samples. In the Fermentation subcategory, pyruvate degradation represented the most abundant pathways in all samples, and in the Generation of Precursor Metabolites and Energy, tricarboxylic acid cycle, glycolysis, and chemoautotrophic energy metabolism represented most abundant pathways. Table 1 summarizes the most abundant pathways in metabolic categories relevant to the termite-microbiota association.

Table 1. Most of the abundant Degradation and Assimilation pathways in the metagenome of Syntermes wheeleri gut microbiota.

	Metabolic Pathway	**P11**	**P12**	**P13**	**P31**	**P32**	**P33**	**WG1**	**WG2**	**WG3**
Aromatic Compounds[b]	tetramethylpyrazine		+[a]				+			+
	gallate	+	+	+		+	+	+	+	
	2,4,6-trinitrotoluene		+	+	+	+	+	+	+	+
	p-cumate	+	+						+	
	vanillin and vanillate	+	+	+	+	+	+	+	+	+
Polymers [b]	cellulose and hemicellulose (cellulosome)	+	+	+	+	+	+		+	
	chitin	+	+		+			+	+	+
	glycogen	+	+	+	+	+	+	+	+	+
	muropeptide		+	+					+	+
	starch	+	+	+	+	+	+	+	+	+
Fermentation	acetoin biosynthesis*[1]	+	+	+				+	+	+
	pyruvate fermentation to acetate	+	+		+	+	+	+	+	+
	pyruvate fermentation to acetone	+	+	+	+		+	+		+
	pyruvate fermentation to butanoate	+		+	+		+			+

	pyruvate fermentation to ethanol*[3]			+	+	+	+	+	+	+
	pyruvate fermentation to propionate		+			+			+	
Generation of Precursor Metabolites and Energy	3-methylbutanol biosynthesis*[1]	+	+		+			+	+	
	ethylene biosynthesis*[1]		+	+					+	+
	chitin degradation to ethanol				+	+	+	+	+	
	glycolysis	+	+			+	+		+	
	hydrogen production*[1]	+	+	+	+	+	+	+	+	+
	methyl-coenzyme M oxidation to CO2	+				+	+	+		
	TCA cycle*[1]		+	+				+		+
Inorganic Nutrients Metabolism	Fe(II) oxidation*[3]	+	+	+	+	+	+		+	+
	cyanate degradation		+	+	+	+		+	+	+
	nitrate reduction (assimilatory)				+	+	+			+
	nitrate reduction (dissimilatory)*[1]	+	+	+						+
	selenate reduction	+						+	+	
	sulfate reduction (assimilatory)		+	+						+
	sulfite oxidation*[1]	+	+	+					+	
Detoxification	arginine dependent acid resistance*[1]	+	+	+						
	triclosan resistance	+	+	+	+	+	+	+	+	+
	vancomycin resistance	+	+	+			+		+	+
	arsenate detoxification (glutaredoxin)*[3]				+	+	+	+	+	+
	ascorbate glutathione cycle*[1]	+	+	+	+	+		+	+	+
	superoxide radicals degradation	+	+	+	+		+	+	+	+
C1 Compounds Assimilation	reductive acetyl-coenzyme A pathway			+	+	+	+	+		+
	reductive TCA cycles	+				+	+		+	+
	CO_2 fixation into oxaloacetate (anapleurotic)	+	+	+		+		+	+	+
	reductive monocarboxylic acid cycle*[3]	+	+	+	+	+	+	+	+	+

	formaldehyde assimilation	+	+		+		+			+	+	
	formate reduction to 5,10-methylenetetrahydrofolate	+		+	+	+	+	+				

P11, P12, and P13 are biological replicates of the first proctodeal segment (P1) microbiota; P31, P32 and P33 are biological replicates of the third proctodeal segment (P3) microbiota; and WG1, WG2, WG3 are biological replicates of the whole gut microbiota.

[a] The plus sign (+) indicates samples in which the pathway was among the five most abundant pathways
[b] Lignocellulose degradation
[*1] Significantly more abundant in P1
[*3] Significantly more abundant in P3

4.3 Potential for carbohydrate degradation

Relative abundances of carbohydrate-active enzyme (CAZyme) domains is presented in Fig. 2A. The relative abundances of glycosyltransferases (GTs) and cellulosome components (cohesin, dockerin, and S-layer homology [SLH] domains) were higher in P1 samples, whereas abundances of glycoside hydrolases (GHs) were higher in P3 samples. PCA of the relative abundance of GH families showed that samples from the same gut segment did not cluster together (Fig. 2B), and P1 and P3 microbiota from the same biological replicates were placed in different dimensions. The GH families separating the P31 and P32 samples were GH 2, 3, 10, 23, 28, 31, 36, 43, 51, 78, and 105; and GH 5, 9, 18, and 94, separated P11. The only significant difference in relative abundance of GH families between P1 and P3 samples was the higher abundance of GH 43 in P3.

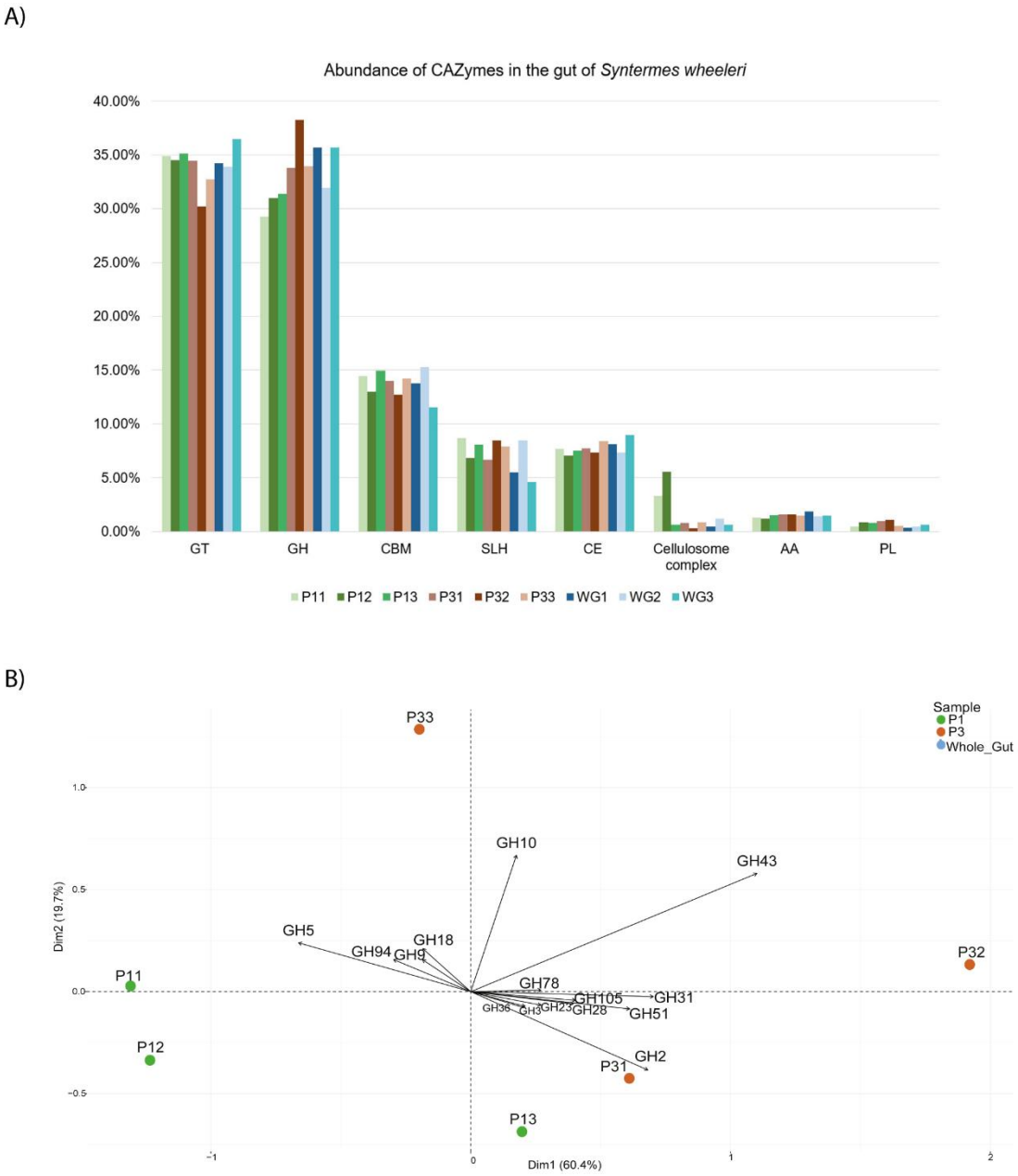


Fig. 2 Carbohydrate degradation potential of Syntermes wheeleri gut microbiota. Carbohydrate-active enzyme (CAZyme) classes detected in the metagenome. A Bars represent the relative abundance of the S-layer homology (SLH) domain, polysaccharide lyase (PL), carbohydrate esterase (CE), auxiliary activities(AA), carbohydrate-binding module (CBM), glycosyltransferase (GT), and glycosyl hydrolase (GH) families in the first and third proctodeal segments (P1 and P3, respectively) and in the whole gut (WG). **B** Results of principal component analysis show the 15 GH families that contribute the most to principal components. P11, P12, and P13 are biological replicates of P1 microbiota; P31, P32, and P33 are biological replicates of P3 microbiota; and WG1, WG2, and WG3 are biological replicates of WG microbiota, 590x320mm (300 x 300 DPI).

The 10 most abundant GH families were similar between P1, P3, and WG microbiota. On average, they were (in decreasingly order): GH 13, 109, 10, 3, 5, 43, 23, 2, 4, and 31 in P1 samples; GH 43, 13, 10, 109, 3, 2, 31, 23, 51, and 5 in P3 sample; and GH 13, 109, 10, 3, 43, 23, 5, 2, 4, and 51 in WG samples (Supporting Information Table S6). The differences were the absence of GH 4 among the most abundant in P3 and the absence of GH 51 among the most abundant in P1. The main functions of the detected GH families are shown in Supporting Information Table S6.

4.4 Microbial metatranscriptome

We successfully obtained high-quality RNA from seven (P11, P12, P13, P33, WG1, WG2 and WG3) out of the nine samples. Therefore, we did not sequence the metatranscriptome of the P31 and P32 samples. The purpose of metatranscriptome sequencing was to confirm the presence and activity (gene expression) of metabolic pathways inferred from the metagenome of *S. wheeleri* gut microbiota. Of the 902 pathways predicted by metagenome analysis, only 159 (17%) did not show gene expression. The CAZY genes suggest functional compartmentalization with families exclusively expressed to segments P1 or P3. The P1 segment microbiome express CAZY families: GH5, GH8, GH39, GH57, GT4, GT19, and GT32. We assigned GH5 transcripts to Clostridiales (i.e., putative endoglucanase, β-xylosidase, and α-amylase). Meanwhile, the P3 segment express CAZY families: GH26, GH28, GH30, GH32, GH63, GH66, GH88, GH105, GH112, and PL12, mainly from Acholeplasmatales source. The CAZY families uniquely expressed in the P3 segment showed similarities with database genes (i.e., β-mannanase, polygalacturonases, unsaturated glucuronyl hydrolases, D-galactosyl-beta-1-4-L-rhamnose

phosphorylase, glucan endo-1,6-beta-glucosidase, sucrose-6-phosphate hydrolase, alpha,alpha-trehalase, and heparinase II/III). The presence or absence of GH metatranscripts per gut segment, P1 or P3, can be seen in Fig. 3.

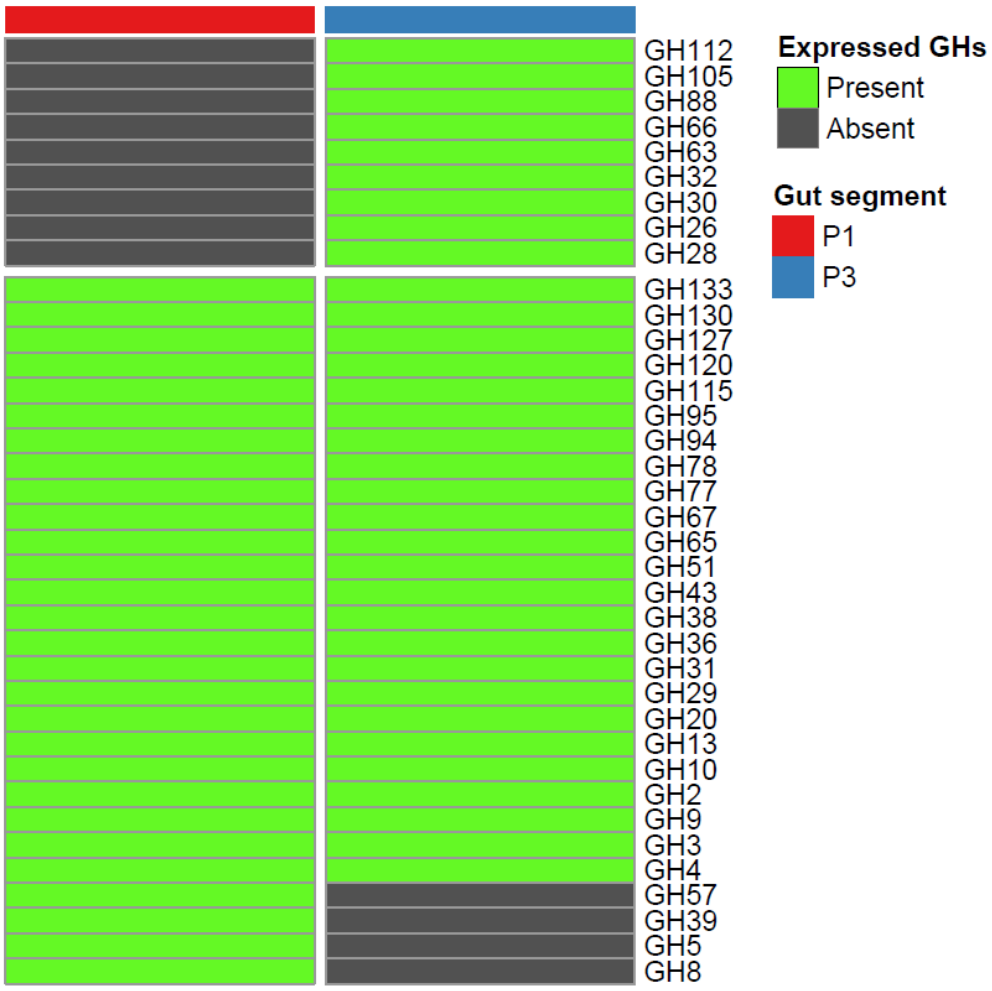


Fig. 3 Heatmap representing the transcripts of Glycosyl Hydrolases (GHs) in the P1 and P3 compartments of the Syntermes wheeleri gut. In green the presence and in gray the absence of that transcript in the metatranscriptome analyses.

5.0 Discussion

Taxonomic profile

In the present study, the gut microbiota taxonomic profile was similar regardless of the technique and phylogenetic marker chosen. The few differences observed include the finding of Acidobacteria among the most abundant phyla using 16S rRNA pyrosequencing, and Tenericutes among the most abundant phyla using WGS. These differences may be related to the resolution of each technique and differences in sequence databases. For example, we detected bacteria of the recently discovered Candidate Phyla Radiation (CPR) (Hug et al., 2016), which were not detected in our previous study of *S. wheeleri* gut microbiota (Santana et al., 2015), which might be due to recent updates in the sequence database.

Our WGS results showed that Firmicutes was the most abundant phylum in each sample. The high abundance of Firmicutes in the alkaline P1 segment in termite species with varied feeding habits has been observed in other studies (Bignell and Eggleton, 1995; Köhler et al., 2012). However, its high abundance in P3 was previously observed primarily in humus and soil feeders (Mikaelyan et al., 2017). In another litter-feeding termite of Syntermitinae, *Cornitermes* sp., the relative abundance of Firmicutes in P3 was significantly lower, resembling a pattern found in wood eaters (Mikaelyan et al., 2017). Therefore, the taxonomic profiles of P1 and P3 microbiota of *S. wheeleri* are similar to those of humus and soil feeders.

Acholeplasma species were the most abundant representatives of Tenericutes and found primarily in P3. These saprophytic bacteria are commonly found in the insect gut (Tully et al., 1987). Although, their presence in the termite gut

was not detected in previous high-throughput sequencing analyses (Köhler et al., 2012; Dietrich et al., 2014), those studies analyzed only P3 fluid. The exopolysaccharide biosynthesis potential and presence of fibronectin-binding protein in the binned sequences of this phylum suggest that these bacteria attach to the gut wall. The presence of superoxide dismutase genes (involved in protection against oxidative stress) also suggest host cell attachment, since the gut wall has a greater load of oxygen; however, it is also possible that superoxide dismutases are present in Tenericutes to cope with its own hydrogen peroxide production (Lazarev et al., 2011). Tenericutes bacteria in the gut of *S. wheeleri* may be involved in nitrogen fixation since their genomic sequences encode nitrogen-fixing enzymes.

Microbial metabolism

Microbes in symbiosis with insects provide vitamins and amino acids, and fuel the host's energy metabolism. The most abundant pathways detected in *S. wheeleri* microbiota are related to these functions. It is important to stress that the presence of metabolic pathways was inferred based solely on EC numbers. One limitation with this approach is that several reactions catalyzed by enzymes do not have an assigned EC number. In addition, because we used *in silico* prediction, it is possible that some of the pathways described here are incomplete.

Metatranscriptomic analysis showed that most of predicted metabolic pathways are active, indicating that the results of metagenomic analysis likely represent active members of the community. Differences in the most abundant pathways between these analyses probably reflect the metabolic state of the microbiota when the termites were sampled. We were not able to obtain high-

quality RNA from the P31 and P32 samples; therefore, the results discussed below are from the metagenomic analysis, unless otherwise stated.

Carbohydrate metabolism

Considering the passage of the food through the gut, we hypothesized that the P1 microbiota specializes in the degradation of more complex molecules, compared with the P3 microbiota. If our hypothesis was correct, we would expect to find higher relative abundances of enzymes responsible for carbohydrate polymer degradation. However, the abundance of genes encoding enzymes acting on cellulose and hemicellulose degradation was similar between compartments. Differences in chitin degradation pathways between compartments appears to reflect the different mechanisms used by microbes to digest chitin rather than differences in the ability to digest this molecule, since the degradation mechanisms rely on the compartment's microbial community composition and physicochemical properties.

PCA analysis of GH families showed that biological replicates of the same gut compartment did not cluster together; however, different gut compartments from the same sample were placed in different dimensions, which suggests differences in carbohydrate hydrolysis between compartments. For example, the GH families that contribute most to the separation of gut compartments of the first biological replicate (samples P11 and P31) point to a different set of domains involved in lignocellulose degradation in each compartment (cellulose digestion in P11 and xylan digestion in P31). Because most GH families include enzymes with different functions, it is difficult to determine which process predominates in each hindgut

compartment. Higher relative abundance of cellulosome components in P1, however, hints at the main lignocellulose degradation mechanisms in this compartment.

The most abundant GH families found in *S. wheeleri* gut microbiota are abundant in termites of different feeding guilds (Poulsen et al., 2014). For example, GH 4, 13, and 51 are among the most abundant families in cow dung feeders; GH 5 and 10 and are the most abundant families in wood feeders (He et al., 2013; Poulsen et al., 2014); and GH 2, 3, and 43 are the most abundant families in a fungus-growing, litter-feeding termite (Poulsen et al., 2014). These findings may reflect the food substrate of *S. wheeleri*, which is dry grass without signs of prior microbial decomposition (as found in cow dung) and is not cellulose-rich like wood.

Recent studies have shown the importance of GHs in carbohydrate metabolism and how they differ according to diet and termite species (Calusinska et al., 2020). Prokaryotes contribute most of the CAZymes used by higher termites to break down polysaccharides into smaller molecules. Cellulases are part of a group of enzymes within the GHs that play an important role in the deconstruction of lignocellulose. The results found in the *in silico* experiments not only contributed to a correlation between phylogeny and carbohydrate metabolism, but also facilitated the choice of potential enzymes that may be used for biotechnological applications, since it guides the best options by the highest metatranscriptome values that may indicate a possible greater presence of that enzyme in that location (Warnecke et al., 2007; Arora et al., 2022).

6.0 Conclusion

The microbial community that lives in symbiosis in the intestines of various termite species gives them the ability to participate in the metabolism of various complex molecules. Lignocellulose is one of the main sources of food for these insects, ranging from wood to leaf litter, for example. The diet they have been subjected to over the years has diversified the microbiota of these eusocial beings. The lower termites have the help of flagellated prokaryotes for the metabolism of lignocellulose, while the higher termites have lost them, giving space to prokaryotes that have come to play a leading role, as is the case with *Syntermes wheeleri*, the termite in this study. This survey found that Firmicutes, Spirochaetes, Bacteroidetes and Proteobacteria play a large part in the microbial diversity of the P1 and P3 segments. GHs, which are important in the bioconversion of carbohydrates, were among the main CAZymes analyzed in both intestinal segments. In addition, it was possible to observe the compartmentalization of function according to the GHs present more in one segment than another, such as greater metabolization of cellulose in P1 and hemicellulose in P3. These results made it possible to understand the specialization of prokaryotes in carbohydrate metabolism in the *S. wheeleri* species, thus allowing a more accurate look at the choice of potential enzymes for biotechnological application.

7.0 References

Abdul Rahman, N., Parks, D. H., Willner, D. L., Engelbrektson, A. L., Goffredi, S. K., Warnecke, F., Scheffrahn, R. H., & Hugenholtz, P. (2015). A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome*, *3*, 5. https://doi.org/10.1186/s40168-015-0067-8.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arora, J., Kinjo, Y., Šobotník, J., Buček, A., Clitheroe, C., Stiblik, P., Roisin, Y., Žifčáková, L., Park, Y. C., Kim, K. Y., Sillam-Dussès, D., Hervé, V., Lo, N., Tokuda, G., Brune, A., & Bourguignon, T. (2022). The functional evolution of termite gut microbiota. *Microbiome*, *10*(1), 78. https://doi.org/10.1186/s40168-022-01258-3

Bignell, D.E. (2010). Morphology, Physiology, Biochemistry and Functional Design of the Termite Gut: An Evolutionary Wonderland. In: Bignell, D., Roisin, Y., Lo, N. (eds) Biology of Termites: a Modern Synthesis. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-3977-4_14

Bignell, D.E., Eggleton, P. On the elevated intestinal pH of higher termites (Isoptera: Termitidae). *Ins. Soc* **42**, 57–69 (1995). https://doi.org/10.1007/BF01245699

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs*, *27*(4), 326-349. https://doi.org/10.2307/1942268

Brune, A. (1998). Termite guts: the world's smallest bioreactors. *Trends in Biotechnology*, *16*(1), 16-21. https://doi.org/10.1016/S0167-7799(97)01151-7

Calusinska, M., Marynowska, M., Bertucci, M., Untereiner, B., Klimek, D., Goux, X., Sillam-Dussès, D., Gawron, P., Halder, R., Wilmes, P., Ferrer, P., Gerin, P., Roisin, Y., & Delfosse, P. (2020). Integrative omics analysis of the termite gut system adaptation to Miscanthus diet identifies lignocellulose degradation enzymes. *Communications biology*, *3*(1), 275. https://doi.org/10.1038/s42003-020-1004-3

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., & Karp, P. D. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, *44*(D1), D471–D480. https://doi.org/10.1093/nar/gkv1164

Dietrich, C., Köhler, T., & Brune, A. (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Applied and environmental microbiology*, *80*(7), 2261–2269. https://doi.org/10.1128/AEM.04206-13

Edgar R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., Kyrpides, N. C., Warnecke, F., Tringe, S. G., & Hugenholtz, P. (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PloS one*, *8*(4), e61126. https://doi.org/10.1371/journal.pone.0061126

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman,

D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature microbiology*, *1*, 16048. https://doi.org/10.1038/nmicrobiol.2016.48

Hug, L. A., Castelle, C. J., Wrighton, K. C., Thomas, B. C., Sharon, I., Frischkorn, K. R., Williams, K. H., Tringe, S. G., & Banfield, J. F. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome*, *1*(1), 22. https://doi.org/10.1186/2049-2618-1-22

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Inward, D. J., Vogler, A. P., & Eggleton, P. (2007). A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Molecular phylogenetics and evolution*, *44*(3), 953–967. https://doi.org/10.1016/j.ympev.2007.05.014

Kassambara, A. and Mundt, F. (2020) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=factoextra

Köhler, T., Dietrich, C., Scheffrahn, R. H., & Brune, A. (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (Nasutitermes spp.). *Applied and environmental microbiology*, *78*(13), 4691–4701. https://doi.org/10.1128/AEM.00683-12

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lazarev, V. N., Levitskii, S. A., Basovskii, Y. I., Chukin, M. M., Akopian, T. A., Vereshchagin, V. V., Kostrjukova, E. S., Kovaleva, G. Y., Kazanov, M. D., Malko, D. B., Vitreschak, A. G., Sernova, N. V., Gelfand, M. S., Demina, I. A., Serebryakova, M. V., Galyamina, M. A., Vtyurin, N. N., Rogov, S. I., Alexeev, D. G., Ladygina, V. G., … Govorun, V. M. (2011). Complete genome and proteome of Acholeplasma laidlawii. *Journal of bacteriology*, *193*(18), 4943–4953. https://doi.org/10.1128/JB.05059-11

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Mayhew P. J. (2007). Why are there so many insect species? Perspectives from fossils and phylogenies. *Biological reviews of the Cambridge Philosophical Society*, *82*(3), 425–454. https://doi.org/10.1111/j.1469-185X.2007.00018.x

Mikaelyan, A., Dietrich, C., Köhler, T., Poulsen, M., Sillam-Dussès, D., & Brune, A. (2015). Diet is the primary determinant of bacterial community structure in the guts of higher termites. *Molecular ecology*, *24*(20), 5284–5295. https://doi.org/10.1111/mec.13376

Mikaelyan, A., Meuser, K., & Brune, A. (2017). Microenvironmental heterogeneity of gut compartments drives bacterial community structure in wood- and humus-

feeding higher termites. *FEMS microbiology ecology*, *93*(1), fiw210. https://doi.org/10.1093/femsec/fiw210

Nayfach, S., & Pollard, K. S. (2015). Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome biology*, *16*(1), 51. https://doi.org/10.1186/s13059-015-0611-7

Noirot, C. (1995, September). The gut of termites (Isoptera). Comparative anatomy, systematics, phylogeny. I. Lower termites. In Annales de la Société Entomologique de France (ns) (Vol. 31, No. 3, pp. 197-226). Taylor & Francis. https://doi.org/10.1080/21686351.1995.12278468

Noirot, C. (2001). The gut of Termites (Isoptera) comparative anatomy, systematics, phylogeny. II. Higher Termites (Termitidae). In *Annales de la Société entomologique de France* (Vol. 37, No. 4, pp. 431-471).

Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P. M., Hauser, F., Pan, H., Yang, Z., Sonnenberg, A. S., de Beer, Z. W., Zhang, Y., Wingfield, M. J., Grimmelikhuijzen, C. J., de Vries, R. P., Korb, J., … Zhang, G. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(40), 14500–14505. https://doi.org/10.1073/pnas.1319718111

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Development Core Team (2010) R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J. O., Scheffrahn, R. H., Sillam-Dussès, D., & Brune, A. (2015). Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome*, *3*, 56. https://doi.org/10.1186/s40168-015-0118-1

Santana, R. H., Catão, E. C., Lopes, F. A., Constantino, R., Barreto, C. C., & Krüger, R. H. (2015). The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. *Microbial ecology*, *70*(2), 545–556. https://doi.org/10.1007/s00248-015-0581-z

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., & Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics (Oxford, England)*, *23*(10), 1282–1288. https://doi.org/10.1093/bioinformatics/btm098

Tully, J. G., Rose, D. L., Whitcomb, R. F., Hackett, K. J., Clark, T. B., Henegar, R. B., Clark, E., Carle, P., & Bove, J. M. (1987). Characterization of some new insect-derived acholeplasmas. *Israel journal of medical sciences*, *23*(6), 699–703.

UniProt Consortium T. (2018). UniProt: the universal protein knowledgebase. *Nucleic acids research*, *46*(5), 2699. https://doi.org/10.1093/nar/gky092

Wickham, H. (2009) Ggplot2: Elegant Graphics for Data Analysis. 2nd Edition, Springer, New York. https://doi.org/10.1007/978-0-387-98141-3

Ye, Y., & Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS computational biology*, *5*(8), e1000465. https://doi.org/10.1371/journal.pcbi.1000465

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research*, *40*(Web Server issue), W445–W451. https://doi.org/10.1093/nar/gks479

Capítulo 3

Characterization of a cellobiohydrolase Family-9 Glycoside Hydrolase bioprospected from the termite *Syntermes wheeleri* gut bacteria metagenome[1]

[1] Manuscrito em revisão a ser submetido no periódico *Biotechnol Lett*.

1.0 Abstract

Objectives Glycosyl hydrolases (GH) are enzymes involved in the degradation of lignocellulose and important for biorefinery industries that aim at the sustainable exploitation of lignocellulosic residues in order to generate value-added products. Glycosyl hydrolases of family 9 (GH9) are related to the degradation of plant material by bacteria. Here we show the characterization of a GH9 cellobiohydrolase (Exo8574) derived from the bacterial metagenome of the termite *Syntermes wheeleri*, endemic to the Brazilian Cerrado, to evaluate their capacity for biotechnological application.

Methods In this study we used bioinformatics tools to perform the phylogeny studies and to be able to identify conserved families and domains. In addition, the heterologous expression system with Escherichia coli BL21(DE3) was used for the production of Exo8574. Optimization of enzyme activity was assessed with the synthetic substrate pNPC, and secondary structure composition changes by Circular Dichroism. In addition to using AlphaFold 2 to predict the 3D structure of the protein.

Results Exo8574 showed 48.8% homology with a protein from an organism belonging to the phylum Firmicutes, which is abundant in the termite species studied here. The optimal activity was at a temperature of 55 °C and pH 6.0, with an influence of metal ions on activity, principal Fe^{2+}. The thermostability test showed a loss of half of the activity at 55°C after one hour of activity. The results of circular dichroism corroborated indicating the dependence of pH on the structure of Exo8574 through variations of the dichroic signal, with greater structuring in β-

antiparallel and α-helices at pH 6.0. Ramachandran plot results confirm the 3D structure as an almost ideal modelling.

Conclusion Biochemical and biophysical characterization results proved Exo8574 as a cellobiohydrolase as predicted by in silico analyses and showed its ability to act on substrates mimicking bonds present in lignocellulose and how this can be influenced according to different parameters applied.

Keywords: Family-9 Glycoside Hydrolase. *Syntermes wheeleri*. Metagenome. Cellobiohydrolase. Termite.

2.0 Introduction

The global market has increasingly engaged in developing products that add ecological value to Earth, from its industrial process to its use by society, and how this may affect the environment (Eswaran et al., 2021). Currently, much of the damage caused comes from products originated from non-renewable sources, such as Petroleum, used in the manufacture of biofuels (e.g., gasoline, diesel oil and kerosene) and derivatives such as plastics, which are related to the biggest problems of carbon gas emissions and pollution of oceans and soils, respectively (Yang et al., 2021; Landrigan et al, 2020). Replacing this type of fossil resources with a cleaner alternative, such as lignocellulose, could contribute to sustainability (Fei et al., 2022; Zhang et al., 2022).

Lignocellulose biomass is the most abundant sustainable carbon source on the planet reaching 20 billion tons per year, including wood, plants, agricultural

waste, forestry waste, and municipal solid waste, which can be converted into environmentally friendly products with high added value (López-Mondéjar et al., 2019; Liu et al., 2020). This material is mostly composed of cellulose (40-60% of dry weight), hemicellulose (20-35% of dry weight) and lignin (15-40% of dry weight), and the composition can vary according to the source: bacteria, marine organisms (*Tunicat*), gymnosperm or angiosperm plants, algae, etc. (Zoghlami & Päes, 2019; Heinze & Liebert, 2012; Zhao et al., 2017; Strnad & Zemljič, 2023). The challenge in the use of this material is overcome its complex structure that requires chemical, physical or biological action for the release of sugars and their respective utilization (Wu et al., 2022).

The use of pretreatments has been one of the solutions in order to facilitate biological processes of lignocellulose bioconversion, since they allow the access of enzymes to the substrate. Cellulases are one of the main enzymes responsible for this, acting in synergism to effectively release mono- and oligosaccharides (Contreras et al., 2020). They can be classified among the Glycosyl Hydrolases (GHs), which are enzymes capable of hydrolyzing glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate fraction (Chettri et al., 2020). GHs are organized into 183 families of which cellulases (endo-β-1,4-glucanase EC 3.2.1.4, cellobiohydrolases EC 3.2.1.91/EC 3.2.1.176 and β-glucosidases EC 3.2.1. 21) are in 24 of them (1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 16, 26, 30, 39, 44, 45, 48, 51, 116, 124, 131, 148, 175 and 180), and are arranged in CAZy (http://www.cazy.org/Glycoside-Hydrolases.html).

The global enzyme market size was valued at $10.6 billion by the end of 2020, $6.1 billion by 2022, and the revenue forecast in 2032 will be $9 billion. Global industrial enzyme sales are expected to increase at a remarkable CAGR (Compound annual growth rate) of 3.9% from 2022-2032 (https://www.grandviewresearch.com/industry-analysis/enzymes-industry). Cellulases are in high demand and cover approximately 20% of the global enzyme market and are very important in academic and industrial research due to their broad applications in pulp and paper, laundry, textile industry, detergent, food and feed, biofuel, brewing and wine, and others. The global cellulase market is estimated to grow at a CAGR of 6.30% during the forecast period (2023-2030). The market size was valued at $2.00 billion in 2022 and is expected to reach $3.10 billion by 2030 (Patil, 2023).

Since cellulases are a very extensive and attractive research topic, in this study we aimed to use microbial bioprospecting from the gut of the *Syntermes wheeleri*, a termite species endemic to the Brazilian Cerrado, to mine a cellobiohydrolase (Constantino, 1995; Santana et al., 2015). Microorganisms that live symbiotically in the gut of termites are responsible for doing most of the bioconversion of the lignocellulose ingested in the diet (Li & Greening, 2022). These organisms possess genes that can be mined by metagenomics, a tool that allows access to the genomic material of all species in the studied microenvironment, and thus discover a novel enzyme that presents ideal characteristics for industrial and biotechnological processes.

3.0 Material and Methods

3.1 Metagenomic sequence analyses

Syntermes wheeleri gut metagenome sequences were used to assemble genomes and make gene annotations using the platform ggKbase (University of California at Berkeley, California, USA - https://ggkbase.berkeley.edu/). To select the genes encoding for cellobiohydrolases, the genomic binning of *S. wheeleri* gut metagenome was searched using EC 3.2.1.91 to identify cellobiohydrolases. All sequences identified were exported as FASTA format proteins to be analyzed in Geneious Prime 2022 (https://www.geneious.com/download/previous-versions/). WG1, WG2, WG3 (WG: whole gut) are samples of the *S. wheeleri* gut in 3 biological replicates.

3.2 Sequence Alignment and Phylogenetic analysis

The *S. wheeleri* gut metagenome selected protein sequences were subjected to multiple alignment using the MAFFT alignment in Geneious Prime 2022. Sequences were edited to eliminate those with less than 200 amino acids that did not present start and stop codons, as well as those that produced a high number of gaps in the alignment. A new updated alignment was obtained each time a sequence was deleted. Afterwards, GH annotation were carried out using BLAST protein in Geneious Prime 2022, and then the metatranscriptome values (coverage per million reads) were verified, which was the indicator to choose the sequence to be synthesized and cloned into the pET24a (+) vector that carries a C-terminal His-tag (GenOne Biotechnologies, Rio de Janeiro, Brazil https://www.genone.com.br/). The phylogenetic tree was generated using FastTree in Geneious Prime 2022. All

bacterial enzyme sequences in Cazy correlated with the GHs of the EC 3.2.1.91 (http://www.cazy.org/Glycoside-Hydrolases.html) were used to align and generate the results.

3.3 Protein production and optimization

The recombinant proteins production using *E. coli* BL21(DE3) heterologous expression system occurs by the presence of a T7 RNA polymerase inducer, which is usually Isopropyl β-D-1-thiogalactopyranoside (IPTG) or its analog Lactose. For the use of IPTG, an inoculum of the clone EXO8574 (*E. coli* BL21 transformed with the plasmid carrying the cellobiohydrolases gene) was prepared in Luria-Bertani medium (LB Broth, L3022, Sigma-Aldrich) containing 100 µg/mL kanamycin (Kanamycin sulfate, 60615, Sigma-Aldrich). IPTG addition (IPTG, I6758, Sigma-Aldrich) was made at an optical density (OD) between 0.6 and 0.9 (SpectraMax® M3, Molecular Devices), and was tested with two different concentrations (0.1 and 1 mM) and temperatures (28 and 37 °C), for 3 h under agitation of 200 rpm (Incubator TE-424, TECNAL, Brazil).

The induction was then collected and cooled for 30 min in a cold storage at 4°C, before being centrifuged at 12,000 rpm, 4 °C for 30 min. The pellet formed was resuspended in lysis buffer (30 mM Tris-HCl, pH 8.0 and 300 mM NaCl) until homogeneous, and then sonicated for 15min (1s ON and 1s OFF), 40% amplitude (Sonics Vibra-Cell VCX 500 - VCX 750). The lysate was centrifuged on the same settings as before and then analyzed on SDS-PAGE 12% (Laemmli, 1970). Lactose was also analyzed as an inducer, however, unlike IPTG this method consists of using a self-inducing culture medium (Studier, 2014) that already contains lactose, without the need to measure the OD to insert it. The growth was overnight (≅18h) under

agitation of 200 rpm, both at 28°C and 37°C. Subsequently analyzed by SDS-PAGE 12%.

3.4 Solubilization of inclusion bodies and protein renaturation

The inclusion bodies of Exo8574 formed in cytoplasm of *E. coli* during the induction of 1 L- culture of clone EXO8574 were solubilized with urea, consisting of: lysis the pellet with 80 mL of 30 mM Tris-HCl, pH 8.0 and 300 mM NaCl, washing the lysate pellet with 50ml of distilled water, centrifuge at 12,000 rpm, 4°C for 30 min, resuspend in 2 mL of distilled water and then add 18 mL of 2 M urea to the washed pellet. Store at -20°C overnight. Thereafter, thaw quickly, centrifuge 12,000 rpm, 4 °C for 30 min and analyze the supernatant in SDS-PAGE 12% (Singhvi et al., 2021). Stock the supernatant for protein renaturation.

The protein renaturation process was performed testing three different buffers (Table 1 S1) in order to test the best one in recovery the structure of Exo8574 (Meyer et al., 2013; Singhvi et al., 2021). The solubilized proteins were dissolved at 50x ratio in renaturation buffer using a peristaltic pump (Peristaltic Pump P-1, Cytiva) with a flow rate of x1, at speed 1.5. Then left for 3 days in a cold storage at 4 °C and then centrifuged 12,000 rpm, 4 °C for 30 min in order to eliminate any precipitate. Then the samples using L-arginine were dialyzed by buffer exchange (30 mM Tris-HCl, pH 8.0) with the peristaltic pump at the same settings. They were then concentrated in an Amicon® 400 mL (Series 8000 Stirred Cells, Millipore).

3.5 Protein purification

Purification of the renatured Exo8574 protein was performed by two separate chromatography techniques on ÄKTA pure. First, a strong anion exchange column (HiPrep Q FF 16/10 20mL, Cytiva) with a column pressure of 0.5 mPa, a working flow rate of 5 mL/min and an elution gradient from 0 to 100% (30 mM Tris-HCl, pH 8.0 and 1 M NaCl) was tested. The second chromatography tested was molecular exclusion chromatography (Sephacryl S-300 120 mL, Cytiva) with a column pressure of 0.5 mPa, a working flow of 1 mL/min and 30 mM Tris-HCl pH 8.0 as elution buffer. The collected peaks were analyzed by 12% SDS-PAGE.

3.6 Molar extinction coefficient and protein concentration

The experimental molar extinction coefficient (or molar absorptivity) of the Exo8574 protein (ε280nm mg/ml) was determined by recording the absorbance at 280 nm in relation to the protein concentration (mg/mL) determined by the Lowry method (Lowry et al., 1995). A UV/Visible Spectrophotometer Jasco V-530 (Jasco Corporation, Tokyo, Japan) was used for (i) fixed point measurement, and (ii) absorbance correction considering light scattering at 350 nm (A350) (Eq. (1)).

$$A_{corr280} = A_{280} - A_{350} \qquad \text{Eq. (1)}$$

The Lowry method was calibrated using the known concentration of bovine serum albumin (BSA) protein. This protein was prepared in five different concentrations: 0.06, 0.125, 0.25, 0.5 and 1 mg/ml. From the linear graph of Exo8574 absorbance at 280 nm and the respective protein concentrations,

determined by the Lowry method, the angular slope was obtained as a measure of the extinction coefficient.

3.7 Enzymatic assays

3.7.1 p-Nitrophenyl standard curve (pNP)

For enzymatic activity calculations, a pNP standard curve was constructed using six different concentrations: 0.06, 0.12, 0.25, 0.5, 1.0 and 1.5 mM, in a volume of 75 µl. The solutions were heated to 45 °C for 30 min in the SpectraMax M3 (Molecular Devices, San Jose, CA, USA) plate spectrophotometer and then 150 µl of 1M sodium carbonate was added. Absorbances at 405 nm were read in the SpectraMax M3 and the values were exported and analyzed in OrigenPro 8 (https://www.originlab.com/origin; OriginLab, Northampton, MA, USA).

3.7.2 Effect of pH and temperature and thermostability

Assays to determine pH influence on Exo8574 enzymatic activity were performed in 96-well plates, using different pH values at 45 °C for 30 min. Buffers in a concentration of 100 mM were prepared in the pH values of 3, 4, 5 and 6 (citric acid/sodium citrate), 7, 8 and 9 (Tris base/ hydrochloric acid), 10 and 11 (sodium bicarbonate/ sodium hydroxide) and 12 (potassium chloride/sodium hydroxide). Enzyme reactions were performed with 32 µL of buffer, 38 µL of 4 mM pNPC (p-nithrophenyl-beta-D-cellobioside, Sigma-Aldrich) and 5 µL of enzyme (≈ 5,84 µM), at 45 °C for 30 min in a 96-well plate on the SpectraMax M3 (Molecular Devices). Then 150 µL of 1 M sodium carbonate (Na2CO3) was added to stop reaction and read at 405 nm. The experiment to determine the influence of temperature was

performed in 200 µl-tubes in a ThermoCycler (T100 thermal cyclers, BIORAD, CA, USA). The temperatures tested were 30 °C, 40 °C, 45 °C, 50 °C, 55 °C , 60 °C, 70 °C, 80 °C and 90 °C, and the reaction time was 30 min. Assays were performed using the buffer most appropriate for the optimum pH for the enzyme. Afterwards, the activity of exo8574 was tested for its thermostability by leaving the enzyme under the optimum temperature and pH and analyzing the activity every 1 hour for a total of 7h.

3.7.3 Influence of metal ions on Exo8574 activity

To determine the effects of metal ions on enzyme activity, Exo8574 was pre-incubated in the presence of Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+}, K^{+}, Mg^{2+}, Mn^{2+}, Na^{+} and Zn^{2+} for 20 min at 28°C at a final concentration of 1mM. After that, the enzymatic assay was performed under the optimum temperature and pH conditions.

3.7.4 Kinetic parameters

Exo8574 kinetic parameters Michaelis-Menten constant (Km), maximum rate (Vmax), catalytic constant (kcat), and kcat/Km were obtained from a Michaelis-Menten plot, which was constructed using pNPG substrate concentrations of 100, 200, 300, 400, 600, 800, 1000, 1200, 1400, 1600 and 1800 µM, and maintaining fixed the enzyme concentration. Assays were performed using the using the optimum pH and temperature. The V_{max} and K_m were calculated from double-reciprocal plots according to the method of Lineweaver-Burk (Eq. 2) (Lineweaver and Burk, 1934).

$$\frac{1}{V_0} = \frac{K_m}{V_{max}} x \frac{1}{[S]} + \frac{1}{V_{max}} \qquad \text{Eq. (2)}$$

In Equation above, V_0: enzyme reative velocity, K_m: Michaelis-Menten constant, V_{max}: Maximum velocity, [S]: Substrate concentration.

3.8 Secondary structure analysis by circular Dichroism

Far-UV spectra (190-260 nm) were recorded on a JascoJ-815 spectropolarimeter equipped with a Peltier-type temperature control system (Jasco Analytical Instruments, Tokyo, Japan) at 25 °C using quartz cuvettes with 0.1 cm optical path. The increment of the recordings was performed every 0.1 nm, with standard sensitivity, bandwidth of 1 nm, and digital integration time of 2 s for a scan rate of 50 nm/min in continuous mode. The protein was studied at different pH values with 5 mM sodium acetate buffer (pH 4.0 to 6.0) and 5 mM Tris-HCL (pH 7.0 to 9.0). The ellipticity for each wavelength (θλ) was converted to average molar ellipticity per residue (θ)λ using Eq. 3:

$$(\theta)_\lambda = \theta_\lambda / 10\ c\ n\ L \qquad \text{Eq. (3)}$$

where: θ_λ is the ellipticity (mdegrees°); n, the number of amino acid residues; c, the protein concentration (mol/L) and L, optical path length (cm). The value of $(\theta)_\lambda$ is expressed in mdegrees.cm^2/dmol (Adler et al., 1976; Woody et., 1996). The secondary structure content of the protein was estimated on the Bestsel web server (http://bestsel.elte.hu/index.php) (Micsonai et al., 2015; Micsonai et al., 2018). Five spectra accumulations were averaged and subtracted from the protein-free buffer spectra.

3.9 Alignment with structured sequences

The search for homologous sequences to Exo8574 was performed using BLASTp against the PDB and non-redundant sequence database as an option (Altschul et al., 1997; Berman et al., 2000). Alignment was then performed using T-COFFEE Multiple Sequence Alignment Server software, and visualized with ESPrint 3.0 software, which allows the alignment to be viewed along with the secondary structure representation (Notredame; Higgins and Heringa, 2000).

3.10 Exo8574 structure analysis

Due to the lack of experimental structure of the Exo8574 protein, the structure was predicted by AlphaFold 2 using the ColabFold software available on Google Collaboratory (Jumper et al., 2021; Mirdita et al., 2022). AlphaFold is a state-of-the-art Neural Network (NN) algorithm that predicts 3D protein structures reaching experimental resolution, as already demonstrated in the challenging 14th Critical Assessment of Protein Structure Prediction (CASP14) (Jumper et al., 2021). Highly accurate structures were verified by comparison with experimental structures, by the per-residue and overall predicted local distance difference test (pLDDT) score, and by the predicted average error (PAE). Molecular graphics and analysis were performed with PyMOL software (DeLano, 2008).

4.0 Results

4.1 Metagenomic sequence analysis

A total of 6 cellobiohydrolase (EC 3.2.1.91) protein sequences from the *S. wheeleri* gut were imported from ggKbase to Geneious Prime 2022 in FASTA protein format. During edition, 3 sequences were excluded, and the remaining 3 were classified as family-9 glycoside hydrolases according to the BLASPp. The gene *WG1_assembly_scaffold_8574_2* was selected to be cloned into pET24a (+) plasmid for protein production because it presented the highest coverage per million reads (0,005) among the remaining sequences. Fig.1 shows the phylogenetic tree of this sequence with other bacterial GHs (EC 3.2.1.91) contained in Cazy. It can be observed that it clusters among GH9.

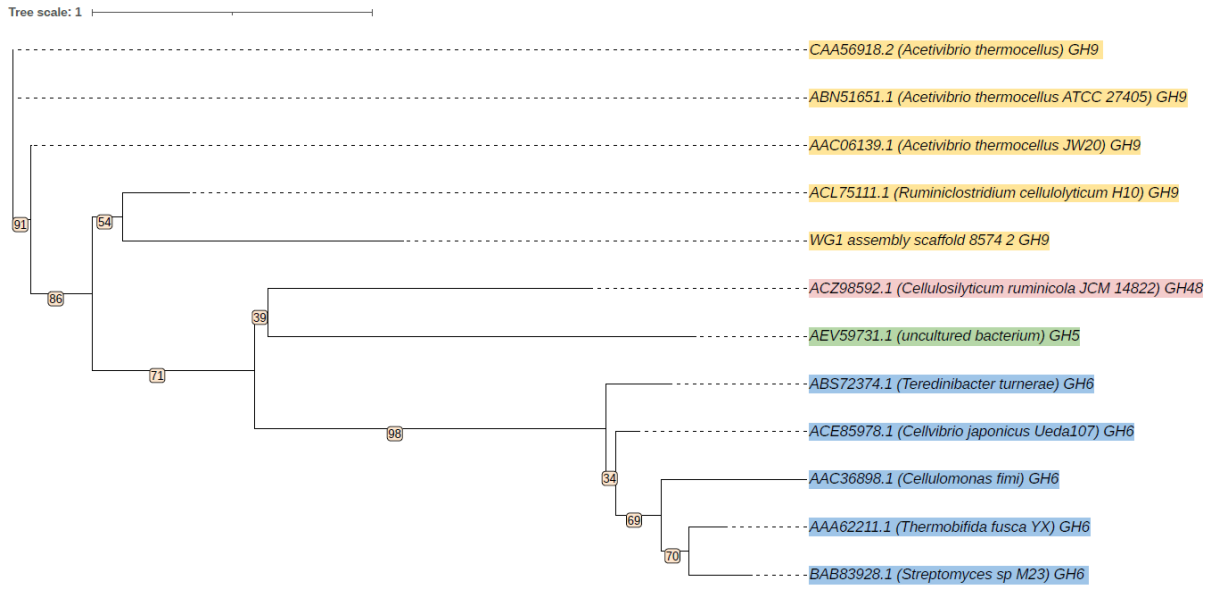


Fig. 1 Phylogenetic tree of all GHs bacterial sequences from EC 3.2.1.91 presented in Cazy. *Syntermes wheeleri* gut metagenomic GH9 cellobiohydrolase (*WG1_assembly_scaffold_8574_2)* and others GH9 are marked with yellow background, GH48 with red background, GH5 with green background and GH6 with blue background. All branches supported by bootstrap values above 34% are shown.

4.2 Exo8574 optimization and solubilization

E. coli BL21(DE3) heterologous expression can be influenced by factors such as inducer concentration and temperature. In the present work, these variations were applied in order to optimize the process. Regarding the variation of IPTG and temperature, Exo8574 showed the same resolution profile in SDS-PAGE 12%, for all combinations applied: 0.1 mM IPTG at 28 and 37 ºC and 1 mM for 28 and 37 ºC (Supplementary Fig. 1A). Exo 85-74 showed to be qualitatively better at 37 ºC, as it showed a larger protein band when compared to that at 28 ºC (Supplementary Fig. 1B). With these results, autoinducer medium with lactose was selected to induce the clone EXO8574 at 37 ºC. The protocol for solubilization of inclusion bodies using mild denaturation was effective in disaggregating the proteins (Supplementary Fig.2).

4.3 Exo8574 purification and molar extinction coefficient

Both purification methods applied showed satisfactory results, allowing to isolate Exo8574 from the *E.coli* proteins. Ion exchange purification generated an elution peak between 280 and 385 mL with ~30% Buffer B (300 mM NaCl) reaching a maximum of ~48.4% (530 mM NaCl) at the peak apex and UV_{280} of 350 mAU (Fig. 2A and B). Molecular exclusion purification generated a peak between 80 and 90 mL, with an apex of ~47 mAU (Fig. 2C and D).

The experimental molar extinction coefficient (ε280nm mg/mL) of Exo8574 was determined by combination of absorbance reading at 280 nm and the Lowry method. The Exo8574 absorption spectra (Fig. S4) in solutions of different concentrations (mg/mL) was determined by the Lowry method using BSA as the standard protein (Supplementary Fig.3). The protein extinction coefficient was

determined from the linear ratio of absorbance at 280 nm and protein concentration, ε280nm mg/mL, (Supplementary Fig.3Aii), as being 1.36 ml.cm−1.mg−1.

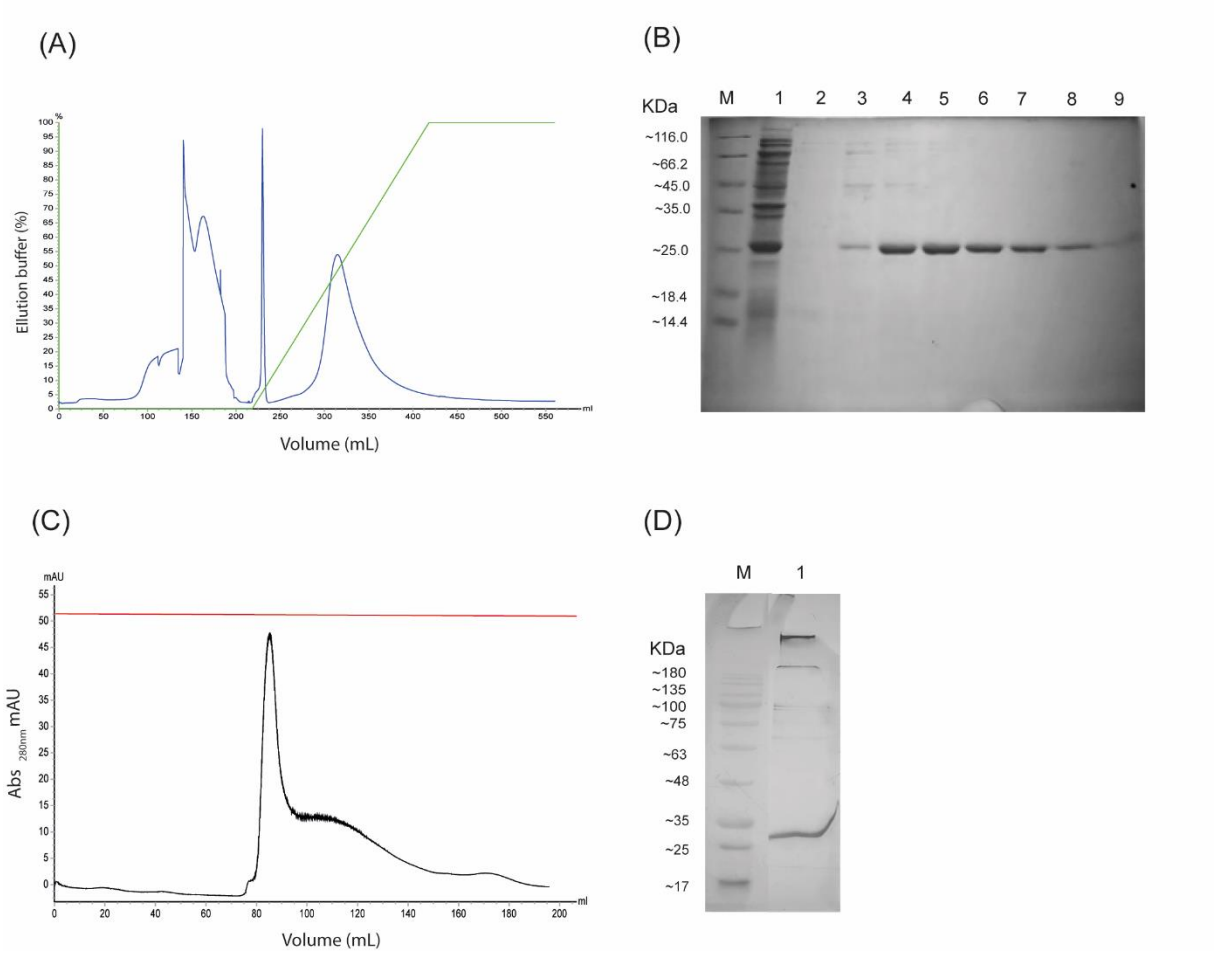


Fig. 2 Purification of Exo 85-74. **A** Chromatogram of the ion exchange purification. Blue line indicates the presence of protein represented in mAU, and green line indicates the elution buffer in 0-100 % gradient. Exo 85-74 showed a maximum of ~48.4 % at the apex of the peak. **B** SDS-PAGE 12% of the peak fractions from 280 to 385 mL. M: protein ladder; 1: injected sample; 2: 280-285 mL; 3: 285-300 mL; 4: 300-315 mL; 5: 315-325 mL; 6: 325-335 mL; 7: 335-345 mL; 8: 345-365 mL and 9: 365-385 mL. **C** Chromatogram of the molecular exclusion purification. The black line indicates the mAU, with a maximum of ~47 mAU, and the red line the conductivity of the elution buffer. **D** SDS-PAGE 12% of the 80-90 mL fraction. M: protein ladder and 1: 80-90 mL.

4.4 Effect of pH and temperature on Exo8574 activity and thermostability

Enzyme activity calculations were performed using a standard pNP concentration curve (Supplementary Fig.3). Exo8574 showed higher activity at

acidic pH (6.0), 3.03 x 10^{-3} IU/mL, and 55 °C, 4.7 x 10^{-3} IU/mL (Fig. 3). The international unit, IU/mL (μmol.min^{-1}.mL^{-1}), is defined as the amount of enzyme that catalyzes the conversion of one micromole of substrate per minute under the specified conditions of the assay method, in 1ml of enzyme. The thermostability assay showed that the enzyme loses more than half of its activity at one hour of heating, and decreases as the hours increase (Supplementary Fig. 4)

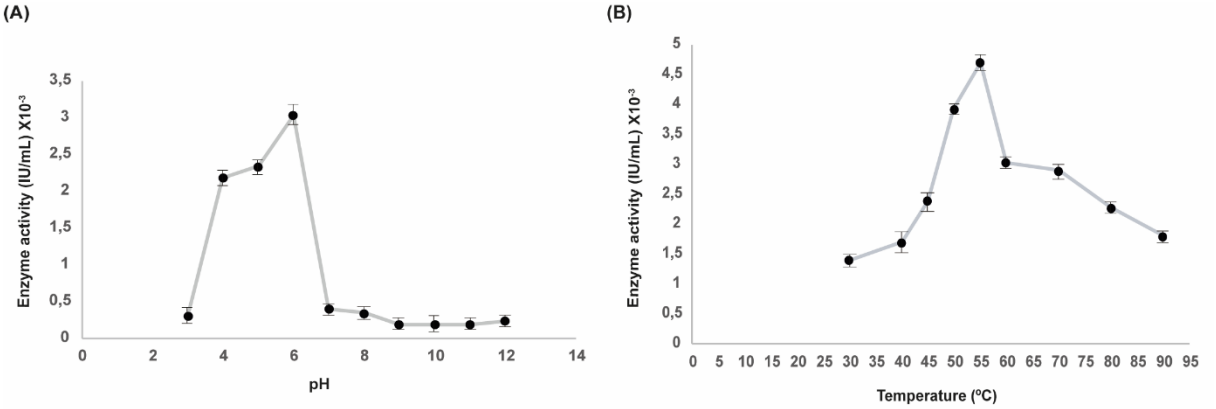


Fig. 3 Effect of pH and temperature on Exo8574 activity. **A** Exo8574 presented best activity at acidic pH (6.0) showing activity of 3.03 x 10^{-3} IU/ml. **B** Exo8574 presented best activity at temperature (55ºC) showing activity of 4.7 x 10^{-3} IU/ml.

4.5 Influence of metal ions on Exo8574 activity

The metal ions influenced the enzyme activity, both decreasing it ($Co^{2+,}$ K^{+}, Mn^{2+}, Na^{+} and Zn^{2+}) and increasing it (Ca^{2+}, Cu^{2+}, Fe^{2+} and Mg^{2+}). The cobalt ion was responsible for the lowest enzymatic activity of Exo8574 (4.17 x 10-3 IU/mL) while the ferrous ion influenced its best action (5.28 x 10-3 IU/mL), always compared to the activity without metal ions (WI) (Fig. 4).

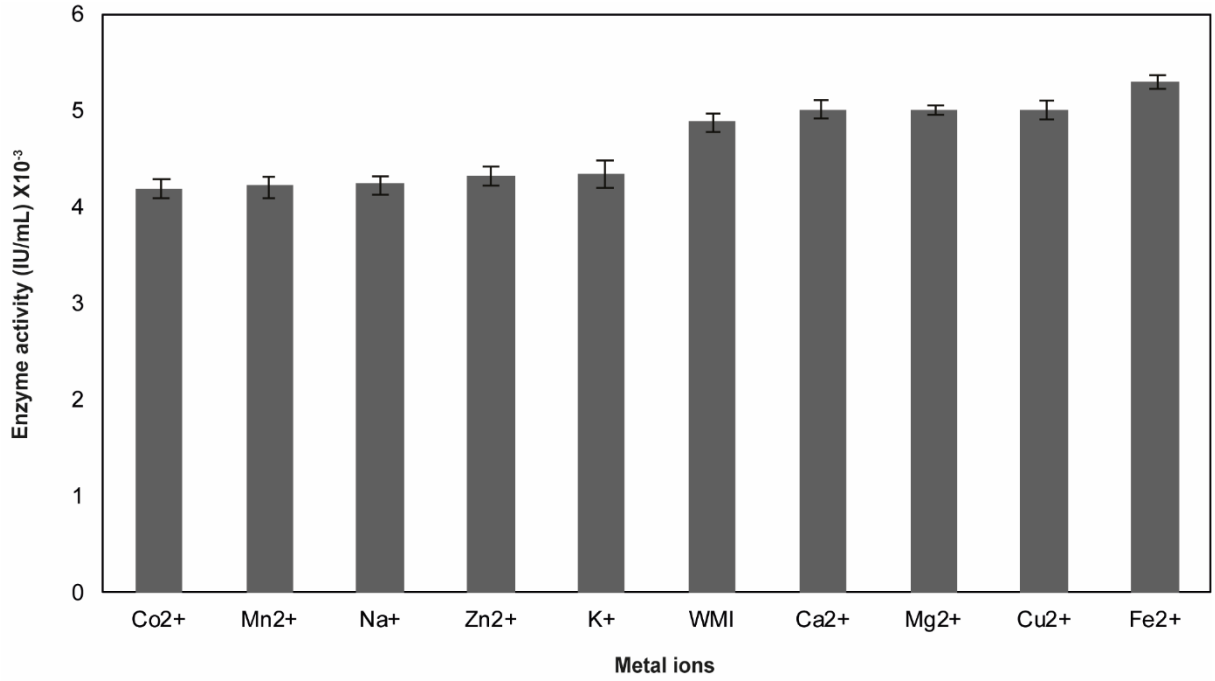


Fig. 4 Effect of metal ions on Exo8574 activity. For determining the effect of metal ions, enzyme assay was performed after pre-incubation of the enzyme with various metal ions each at a concentration of 1 mM at 28°C for 20 min.

4.6 Kinetic parameters

Optimized enzymatic assays were used to determine the effect of pNPC substrate concentrations on enzyme activity, using 5 µl of Exo8574 at 5.8 µM at pH 6.0 and 55°C. Results were ploted and used to construct the Michaelis-Menten and Lineweaver-Burk graphs (Fig. 5). The V_{max} reached was 5.07 ± 0,06 x 10^{-3} µmol /min, and the K_m, the affinity constant for the substrate (Michaellis constant) was 186,96 ± 4.6 µmol. Exo8574 k_{cat}, that represent the turnover, was 13 ± 0,7 min^{-1} and the k_{cat}/K_m, specificity constant (i.e., kinetic efficiency) was 0,07 min^{-1}. $µmol^{-1}$ (Table 1).

Table 1 Kinetic parameters of Exo8574

Protein	V_{max} ×10^{-3} (µmol /min)	K_m (µmol)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1}. $µmol^{-1}$)
Exo8574	5.07 ± 0.06	186.96 ± 4.6	13 ± 0,7	0,07

5 µl of Exo8574 at 5.8 µM were used in assay.

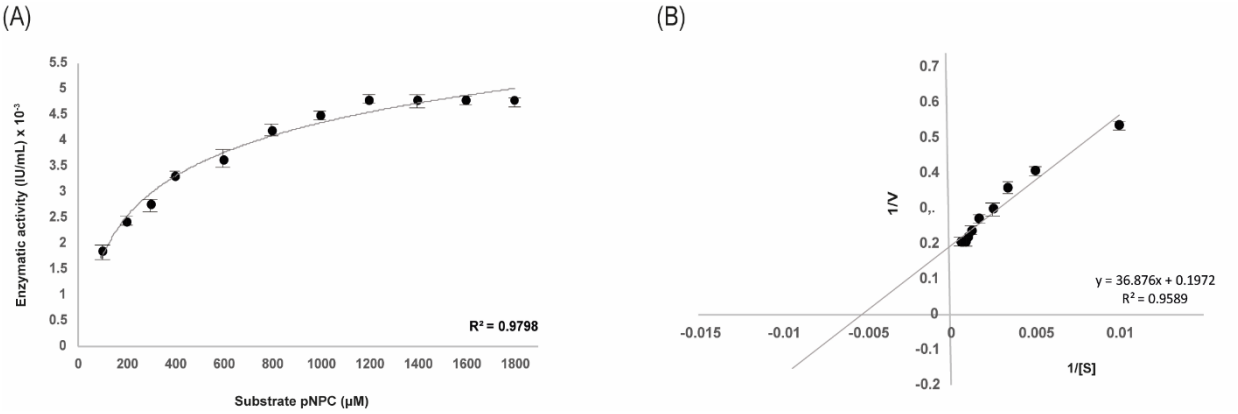


Fig. 5 Effect of pNPC concentrations on Exo8574 activity. **A** Michaelis-Menten curve at pH 6.0 and 55 °C. **B** Lineweaver-Burk curve at same conditions.

4.7 Secondary structure analysis by circular Dichroism

Far-UV CD studies were performed to determine the secondary structure composition of the enzyme at different pH values. At physiological pH (pH 7.0), the enzyme shows a dichroic spectrum with positive bands at 190-195 nm and negative bands between 205-220 nm (Fig. 6A). This feature is typical of proteins with a relatively high β-sheet content, as demonstrated by the deconvolution of the spectrum (Fig. 6B), where the enzyme shows 46% antiparallel β-sheets and 36% unordered structures (Kelly et al., 2005). In the physiological condition, the enzyme shows only 5% helical structures and 12% β-turn.

The Far-UV CD spectra show changes in its spectral features when the pH was modified. With the exception of pH 4.0, the spectra indicate that the enzyme largely retains its secondary structures in the pH range 5.0 to 9.0 (Fig. 6A). At pH 4, the dichroic signal virtually disappears indicating significant loss of protein structure. The estimates of the structural variations of the enzyme against the pH changes showed that the enzyme presents practically no parallel β-sheet structures in the pH

range studied. In addition, the contents of β-turns changed little during the study, while the contents of α-helix showed heights values at pHs 5.0 and 6.0, respectively (Fig. 6B).

The greatest variations in the content of secondary structures were found in antiparallel β-sheets and unordered structures. As the medium acidifies, the content of antiparallel β-sheets reaches 60% at pH 6.0, and then 20% at pH 5.0 (Fig. 6B). This behavior is inversely correlated with the decrease up to 9% at pH 6.0 of the unordered structures and their increase up to 39% at pH 5.0. The study indicates that the enzyme, when faced with pH variations, modifies its structure fundamentally in the content of antiparallel β-sheets to the detriment of its disordered structures or vice versa (Fig. 6B).

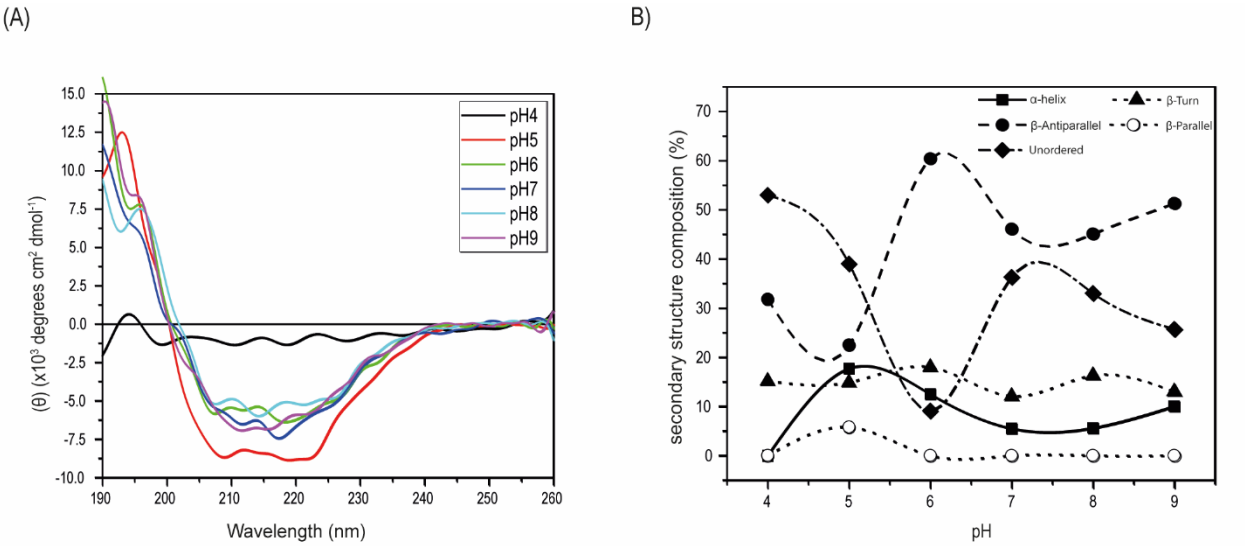


Fig. 6 Composition of secondary structures of Exo8574 by circular dichroism as a function of pH. Far-UV CD spectra (190-260 nm). A were recorded at 25 °C at different pH values using 5 mM of the corresponding pH solutions: sodium acetate (pH 4 to 6) and Tris-HCl (pH 8 and 9). The protein concentration was 2.0 µM. In each case, baselines were subtracted using control buffer without protein. **B** The changes in secondary structure composition as a function of pH. In the graph the square symbol represents the composition of helical structures, the solid and open circles the antiparallel and parallel beta-sheet structures, respectively, the triangle the beta-spin structures, and the diamonds for unordered structures. The Bestsel algorithm was used on the Internet (http://bestsel.elte.hu/index.php). The parameters of protein molecular weight 27726.93 Da, 251 amino acids, 2 µM protein concentration, and optical path length of 0.1 cm were used for the deconvolutions.

4.8 Alignment with structured sequences and 3D prediction

Exo 85-74 belonging to the GH9 family is composed of 245 amino acid residues, and according to the alignment performed with the best hits resulting from BLASTp with PDB sequences, it presents conserved active sites at positions 174 H (histidine), 218 D (aspartic acid) and 227 E (glutamic acid) (Fig. 7).

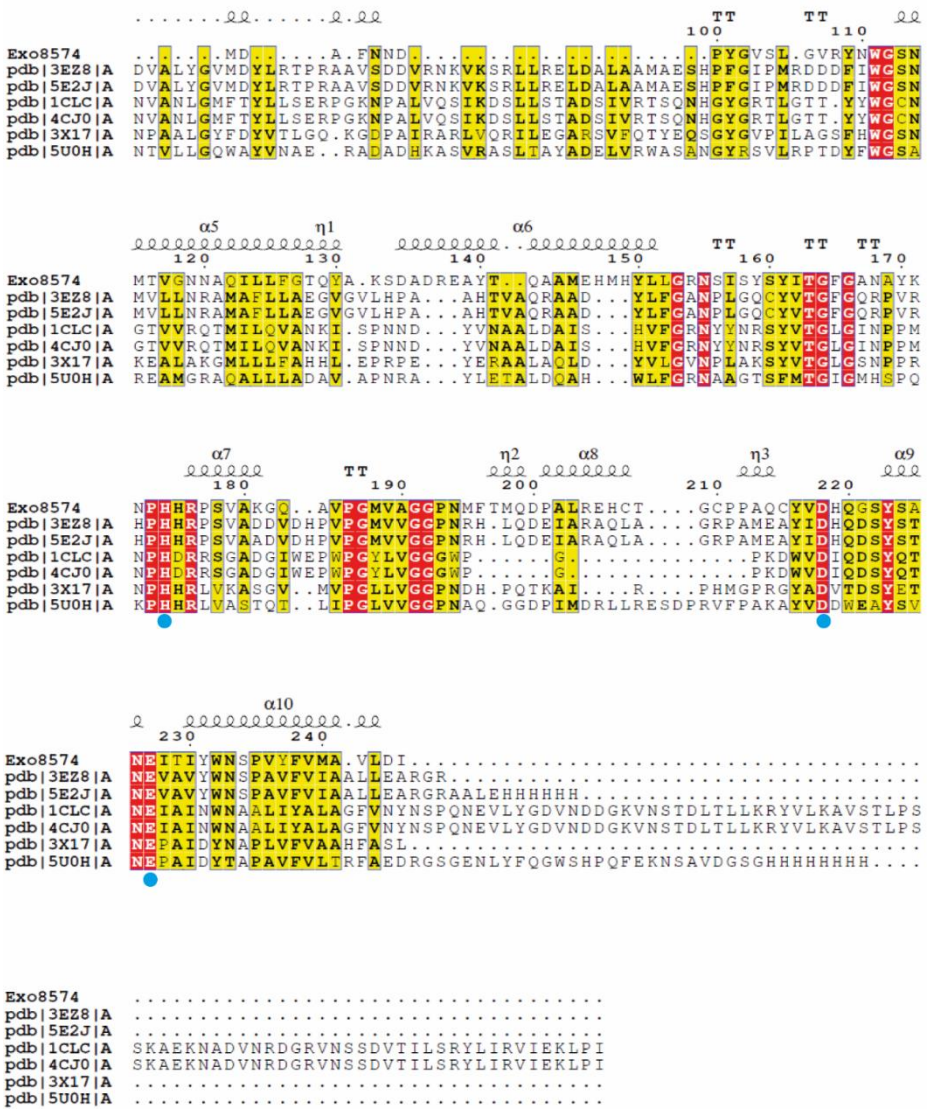


Fig. 7. Alignment of Exo8574 with Blast best hit sequences with PDB sequences. Exo8574 enzyme sequence shows 225 amino acids with conserved domain for GH9. The active sites are shown in blue circles: 174 H (histidine), 218 D (aspartic acid) and 227 E (glutamic acid). Red highlighting is for conserved residues in all sequences and yellow for almost all. Spiral structures are α-helix, η:3_{10}-helix and TT (β-turn).

The predicted three-dimensional structure of Exo8574 is shown in Fig. 8A and B, which had a good accuracy according to the Ramachandran Plot (Fig. 8C) that gave 88.3% of the amino acid residues in favorable regions. In addition, it was possible to see the spatial location of the active sites in the modeling.

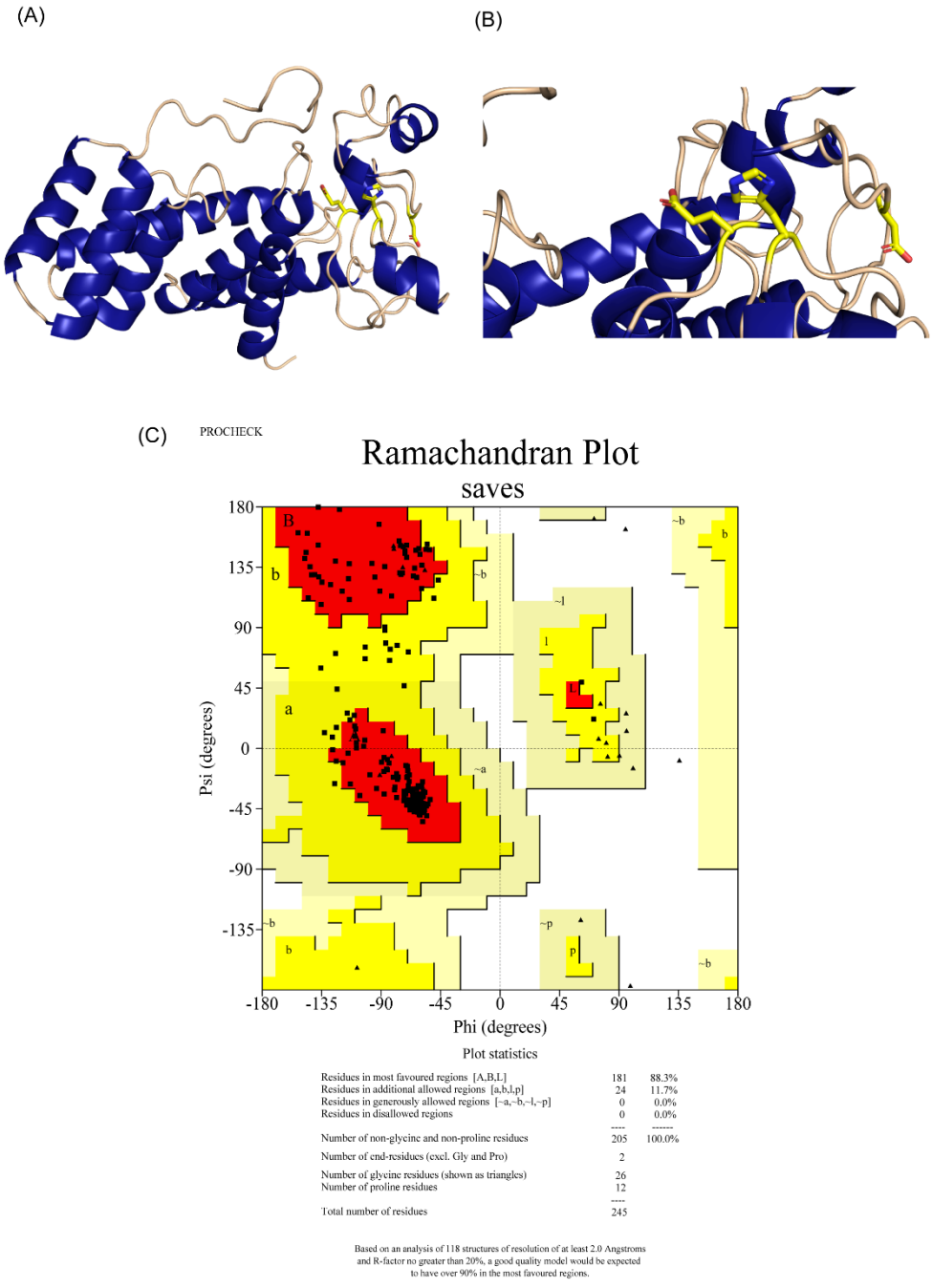


Fig 8. Structural modeling of Exo8574. A and B Theoretical three-dimensional structure of Exo8574 with labeling of conserved active sites. **C** Ramachandran Plot for Exo8574. The plot shows 88.3% of amino acid residues in the most favored region and 11.7% are in the additional favored region.

5.0 Discussion

The host-associated microbiome is known to influence many characteristics, and has been used frequently in the last 10 years for the understanding of metabolic processes and their consequences in the microhabitat where they are present (Ross and Hayes, 2022). Advances in sequencing technologies have enabled rapid sequencing of genetic material from samples such as water, air, soil, animals and even humans allowing to directly study the genetic composition of communities of species (from viruses to megafauna), a field known as metagenomics (Czech et al., 2022).

The main idea of the work is to bioprospect enzymes capable of converting lignocellulosic material into high added value chemicals. With this, a set of gene data annotated from the bacterial metagenome of *Syntermes wheeleri* was used to mine these enzymes (Santana et al., 2015). Termites are considered one of the main degraders of lignocellulosic material in the ecosystem, and they play a crucial role in forming terrestrial ecosystems. They rely on a community of microorganisms in their digestive tract to obtain some essential products for their survival, such as sugars, hydrogen and acetate, which are further metabolized by these symbiotic microorganisms as sources of energy and nutrition (Zhou et al., 2019).

The bacterial community present in the *S. wheeleri* gut is mostly composed of bacteria from the phyla Firmicutes and Spirochaetes (Santana et al., 2015). These two phyla have already been annotated as the most abundant for four other termite species, both lower and higher, and the percentage of one or the other may vary depending on the phylogeny and diet to which these termites are subjected (Su et

al., 2016). The cellobiohydrolases genes imported from ggKbase were mostly annotated as phylum Firmicutes, with representative of the phylum Proteobacteria as well, in addition to uncultured bacteria. Unculturable microorganisms have been widely used as sources of study for new enzymes since their access is only through sequenced genetic material. They are present in the most diverse ecosystems and especially in extreme environments (Iquebal et al., 2022).

The protein screened for characterization, Exo8574, was annotated as belonging to the phyla Firmicutes and showed 48.8% homology with the protein from *Paenibacillus phyllosphaerae*, a species already related to the production of lignocellulosic converting enzymes (Rivas et al., 2005). BLASTp showed that Exo8574 present similarity to GH9 sequences and clusters with other bacterial GH9s from EC 3.2.1.91. This family encode genes widespread among cellulolytic microorganisms (except aerobic fungi) and plants, but genes encoding this family of enzymes are particularly abundant in anaerobic cellulosome-producing bacteria. These are large, heterogeneous extracellular complexes that efficiently degrade cellulose and related polysaccharides from the plant cell wall (Davison and Blexter, 2005; Ravachol et al., 2014; Zheng et al., 2023).

After the in silico analysis processes, the sequences were added in pET gene expression vectors, more specifically in pET24a (+) and transformed into *E.coli* BL21(DE3). Among the prokaryotes modified for the production of carbohydrate-active enzymes (CAZymes), *E. coli* is one of the most engineered and used, as it presents easy-to-manipulate genetics and a fast growth rate (Lopes, Martins and Goldbeck, 2021). Usually, part of the recombinant proteins in *E. coli* remain

intracellularly or as inclusion bodies (Baneyx and Mujacic, 2004). With this in mind, production optimization processes were carried out and the ideal parameters were selected among those tested.

The production of Exo 85-74 was most efficient by autoinduction with lactose at 28°C. Inclusion bodies were solubilized using the freeze-thaw method with 2 M urea, which has been successful for other proteins (Qi, Sun and Xiong, 2015; Singhvi et al., 2021). Enzymes solubilized with urea were denatured and therefore needed to be renatured, i.e. get their native form back. The simplest buffer allowed the renaturation of Exo 85-74 (Meyer et al., 2013). There is no single solution for protein refolding. The identification of optimal conditions represents a bottleneck for process development. The best conditions are hardly predictable, and are usually identified empirically through trial and error approaches (Buscajoni et al., 2022).

After renaturation the proteins were purified by molecular exclusion and used at a concentration of 5.8 µM for enzymatic characterization. It showed best activity at pH 6.0 and temperature 55 °C, with 4.7 x 10^{-3} IU/mL. The detection of activity at different pH values has been reported for termite cellulases, and the pH at which they have maximum activity varies between different species (Lima et al., 2014; Lima et al., 2020). Temperature has also been analyzed for other cellulase enzymes from termites and others, and they generally show values between 40 to 60 °C (Lima et al., 2014, Cairns and Esen, 2010; Florindo et al., 2018; Suwan et al., 2017; Ghosh et al., 2018). The K_m obtained by Lineweaver-Burk showed that Exo8574 requires 186.96 uM of pNPC to reach half the maximum reaction rate, and has a turnover of

13min^{-1}, i.e. the maximum number of substrate molecules converted to product per unit time, which is influenced by the concentration of enzyme used.

Activities of enzyme were stimulated in the presence of Ca^{2+}, Mg^{2+}, Cu^{2+} and Fe^{2+} with ferrous ion been responsible for an increasing if 1,08-fold. On the other hand, a week activity was observed in the presence of Zn^{2+}, Na^{+}, Mn^{2+} and Co^{2+} with cobalt ion decreasing it 1.16-folf. Metal ions are known to influence the activity of cellulase enzymes, and can interact directly with the active site or also with the substrate, and in both cases result in better activity or even inhibit the reaction to happen. The Fe2+ and Cu2+ ions, responsible for participating in electron transfer reactions, have already been reported as inhibitors of this class of enzymes, however, in this work it was shown to have the opposite effect. This knowledge is important for industrial processes that always need a large variation of enzymes to use in variable processes and conditions (Lippard and Berg, 1994; Tejirian and Xu, 2010; de Cassia Pereira et al., 2017).

To complement the biochemical analyses, some biophysical studies were performed. Circular Dichroism (CD), for example, complemented the biochemical pH results, showing that the protein does indeed present structural dependence on pH, and that at pH 6.0, of better enzymatic activity, it is more structured in β-antiparallel (~60%), β-turns (~18%) and α-helix (~14%). β-antiparallel together with the unordered regions (~10%), were the ones that suffered the most changes with the pH change. Due to their fluctuating structures, unordered proteins offer many advantages for cellular function. The flexibility of a disordered protein means that the protein can be easily accessed (Stollar and Smith, 2020). The alignment with

structured PDB sequences suggested the conservation of the active sites of Exo8574 at amino acids 174 H (histidine), 218 D (aspartic acid) and 227 E (glutamic acid) which can be observed in the other aligned GH9 sequences. In the three-dimensional structure it can be seen that the active sites are in the regions of disorder, which may allow greater malleability of access to the substrate as previously stated.

6.0 Conclusion

In this study we characterized a cellobiohydrolase from the bacterial metagenome of the gut of *Syntermes wheeleri*, a termite species endemic to the Brazilian Cerrado. It showed similarity with an enzyme from *Paenibacillus phyllosphaerae*, indicating a possible participation of this microorganism in the conversion of lignocellulose in the gut of *S. wheeleri*, as previously reported for another termite species. Exo8574 is the first cellobiohydrolase GH9 (EC 3.2.1.91) enzyme characterized in detail that originated from this higher termite species. Activity assays with the synthetic substrate pNPC allowed to classify it indeed as a cellobiohydrolase. It shows better activity at acid pH 6.0 and temperature 55°C, which is in the range of other bacterial cellulases from other termite species, and can be used in industrial processes that require more acidic environments. The secondary structure composition of Exo8574 is influenced by pH change and shows improvement in its activity when in the presence of Fe^{2+}, showing the influence of metal ions in cellulase activity. The active sites were conserved with other GH9 for both endoglucanases and cellobiohydrolase, and may share the same hydrolytic attack despite being enzymes that act on different substrates. Assays with natural

substrates will be explored in the future together with other cellulases to analyze the synergistic effect on lignocellulose conversion and its potential as a biotechnological tool.

7.0 References

Baneyx, F., Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol,* **22**, 1399–1408. https://doi.org/10.1038/nbt1029

Buscajoni, L., Martinetz, M. C., Berkemeyer, M., & Brocard, C. (2022). Refolding in the modern biopharmaceutical industry. *Biotechnology advances*, *61*, 108050. https://doi.org/10.1016/j.biotechadv.2022.108050

Cairns, J. R. K., & Esen, A. (2010). [beta]-Glucosidases. *Cellular and Molecular Life Sciences*, *67*(20), 3389. https://doi.org/10.1007/s00018-010-0399-2

Chettri, D., Verma, A. K., & Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology reports (Amsterdam, Netherlands)*, *28*, e00525. https://doi.org/10.1016/j.btre.2020.e00525

Constantino, R. (1995). Revision of the Neotropical termite genus *Syntermes Holmgren* (Isoptera: Termitidae). *The University of Kansas Science Bulletin*, *55*, 455–518. https://www.biodiversitylibrary.org/part/781

Contreras, F., Pramanik, S., M. Rozhkova, A., N. Zorov, I., Korotkova, O., P. Sinitsyn, A., Schwaneberg, U., & D. Davari, M. (2020). Engineering Robust Cellulases for Tailored Lignocellulosic Degradation Cocktails. *International Journal of Molecular Sciences*, *21*(5), 1589. https://doi.org/10.3390/ijms21051589

Czech, L., Stamatakis, A., Dunthorn, M., & Barbera, P. (2022). Metagenomic Analysis Using Phylogenetic Placement - A Review of the First Decade. *Frontiers in bioinformatics*, *2*, 871393. https://doi.org/10.3389/fbinf.2022.871393

Davison, A., & Blaxter, M. (2005). Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Molecular biology and evolution*, *22*(5), 1273–1284. https://doi.org/10.1093/molbev/msi107

de Cassia Pereira, J., Giese, E. C., de Souza Moretti, M. M., dos Santos Gomes, A. C., Perrone, O. M., Boscolo, M., ... & Martins, D. A. B. (2017). Effect of metal ions, chemical agents and organic compounds on lignocellulolytic enzymes activities. *Enzyme inhibitors and activators*, *29*, 139-164.

Eswaran, N., Parameswaran, S., & Johnson, T. S. (2021). Biofuels and Sustainability. *Methods in molecular biology (Clifton, N.J.)*, *2290*, 317–342. https://doi.org/10.1007/978-1-0716-1323-8_20

Fei, X., Wang, J., Zhang, X., Jia, Z., Jiang, Y., & Liu, X. (2022). Recent Progress on Bio-Based Polyesters Derived from 2,5-Furandicarbonxylic Acid (FDCA). *Polymers*, *14*(3), 625. https://doi.org/10.3390/polym14030625

Florindo, R. N., Souza, V. P., Manzine, L. R., Camilo, C. M., Marana, S. R., Polikarpov, I., & Nascimento, A. S. (2018). Structural and biochemical characterization of a GH3 β-glucosidase from the probiotic bacteria Bifidobacterium adolescentis. *Biochimie*, *148*, 107–115. https://doi.org/10.1016/j.biochi.2018.03.007

Ghosh, D., Jana, B. B., Lahiri, S., Bhakta, J., & Bhattacharjee, A. (2018). Assessing the cellulase enzyme heterogeneity of bacterial strains and their feedback to cattle manure degradation in a greenhouse model of in vivo pond ecosystem. *Environmental monitoring and assessment*, *190*(8), 452. https://doi.org/10.1007/s10661-018-6821-1

Heinze T., Liebert T. (2012). Celluloses and Polyoses/Hemicelluloses. *Polymer Science: A Comprehensive Reference*. Elsevier. 83–152.

Iquebal, M. A., Jagannadham, J., Jaiswal, S., Prabha, R., Rai, A., & Kumar, D. (2022). Potential Use of Microbial Community Genomes in Various Dimensions of Agriculture Productivity and Its Management: A Review. *Frontiers in microbiology*, *13*, 708335. https://doi.org/10.3389/fmicb.2022.708335

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica et biophysica acta*, *1751*(2), 119–139. https://doi.org/10.1016/j.bbapap.2005.06.005

Laemmli U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685. https://doi.org/10.1038/227680a0

Landrigan, P. J., Stegeman, J. J., Fleming, L. E., Allemand, D., Anderson, D. M., Backer, L. C., Brucker-Davis, F., Chevalier, N., Corra, L., Czerucka, D., Bottein, M. D., Demeneix, B., Depledge, M., Deheyn, D. D., Dorman, C. J., Fénichel, P., Fisher, S., Gaill, F., Galgani, F., Gaze, W. H., … Rampal, P. (2020). Human Health and Ocean Pollution. *Annals of global health*, *86*(1), 151. https://doi.org/10.5334/aogh.2831

Li, H., & Greening, C. (2022). Termite-engineered microbial communities of termite nest structures: a new dimension to the extended phenotype. *FEMS microbiology reviews*, *46*(6), fuac034. https://doi.org/10.1093/femsre/fuac034

Lima, R. A. T., De Oliveira, G., Souza, A. A., Lopes, F. A. C., Santana, R. H., Istvan, P., Quirino, B. F., Barbosa, J., De Freitas, S., Garay, A. V., & Krüger, R. H. (2020). Functional and structural characterization of a novel GH3 β-glucosidase from the gut metagenome of the Brazilian Cerrado termite Syntermes wheeleri. *International journal of biological macromolecules*, *165*(Pt A), 822–834. https://doi.org/10.1016/j.ijbiomac.2020.09.236

Lima, T.de A., Pontual, E. V., Dornelles, L. P., Amorim, P. K., Sá, R. A., Coelho, L. C., Napoleão, T. H., & Paiva, P. M. (2014). Digestive enzymes from workers and soldiers of termite Nasutitermes corniger. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, *176*, 1–8. https://doi.org/10.1016/j.cbpb.2014.07.001

Lineweaver, H. and Burk, D. (1934) The determination of enzy me dissociation constant. Journal of the American Chemical Society, 56, 658-666. doi:10.1021/ja01318a036

Lippard, S. J., & Berg, J. M. (1994). *Principles of bioinorganic chemistry*. University Science Books.

Liu, Y. J., Li, B., Feng, Y., & Cui, Q. (2020). Consolidated bio-saccharification: Leading lignocellulose bioconversion into the real world. *Biotechnology advances*, *40*, 107535. https://doi.org/10.1016/j.biotechadv.2020.107535

Lopes, A. M. M., Martins, M., & Goldbeck, R. (2021). Heterologous Expression of Lignocellulose-Modifying Enzymes in Microorganisms: Current Status. *Molecular biotechnology*, *63*(3), 184–199. https://doi.org/10.1007/s12033-020-00288-2

López-Mondéjar, R., Algora, C., & Baldrian, P. (2019). Lignocellulolytic systems of soil bacteria: A vast and diverse toolbox for biotechnological conversion processes. *Biotechnology advances*, *37*(6), 107374. https://doi.org/10.1016/j.biotechadv.2019.03.013

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *The Journal of biological chemistry*, *193*(1), 265–275. https://doi.org/10.1016/S0021-9258(19)52451-6

Meyer, D., Sielaff, F., Hammami, M., Böttcher-Friebertshäuser, E., Garten, W., & Steinmetzer, T. (2013). Identification of the first synthetic inhibitors of the type II transmembrane serine protease TMPRSS2 suitable for inhibition of influenza virus activation. *The Biochemical journal*, *452*(2), 331–343. https://doi.org/10.1042/BJ20130101

Meyer, D., Sielaff, F., Hammami, M., Böttcher-Friebertshäuser, E., Garten, W., & Steinmetzer, T. (2013). Identification of the first synthetic inhibitors of the type II transmembrane serine protease TMPRSS2 suitable for inhibition of influenza virus activation. *The Biochemical journal*, *452*(2), 331–343. https://doi.org/10.1042/BJ20130101

Patil, M. (2023, june 29). *Global Market Analysis on Cellulase market, Chlorogenic Acid market, Spandex Fibermarket forecasted till 2030*. Recuperado de

https://www.einnews.com/pr_news/642111523/global-market-analysis-on-cellulase-market-chlorogenic-acid-market-spandex-fibermarket-forecasted-till-2030.

Qi, X., Sun, Y., & Xiong, S. (2015). A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. *Microbial cell factories*, *14*, 1-12. https://doi.org/10.1186/s12934-015-0208-6

Ravachol, J., Borne, R., Tardif, C., de Philip, P., & Fierobe, H. P. (2014). Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium Clostridium cellulolyticum. *The Journal of biological chemistry*, *289*(11), 7335–7348. https://doi.org/10.1074/jbc.M113.545046

Rivas, R., Mateos, P. F., Martínez-Molina, E., & Velázquez, E. (2005). Paenibacillus phyllosphaerae sp. nov., a xylanolytic bacterium isolated from the phyllosphere of Phoenix dactylifera. *International journal of systematic and evolutionary microbiology*, *55*(Pt 2), 743–746. https://doi.org/10.1099/ijs.0.63323-0

Ross, E. M., & Hayes, B. J. (2022). Metagenomic Predictions: A Review 10 years on. *Frontiers in genetics*, *13*, 865765. https://doi.org/10.3389/fgene.2022.865765

Santana, R. H., Catão, E. C., Lopes, F. A., Constantino, R., Barreto, C. C., & Krüger, R. H. (2015). The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. *Microbial ecology*, *70*(2), 545–556. https://doi.org/10.1007/s00248-015-0581-z

Singhvi, P., Verma, J., Panwar, N., Wani, T. Q., Singh, A., Qudratullah, M., Chakraborty, A., Saneja, A., Sarkar, D. P., & Panda, A. K. (2021). Molecular Attributes Associated With Refolding of Inclusion Body Proteins Using the Freeze-Thaw Method. *Frontiers in microbiology*, *12*, 618559. https://doi.org/10.3389/fmicb.2021.618559

Stollar, E. J., & Smith, D. P. (2020). Uncovering protein structure. *Essays in biochemistry*, *64*(4), 649–680. https://doi.org/10.1042/EBC20190042

Strnad, S., & Zemljič, L. F. (2023). Cellulose-Chitosan Functional Biocomposites. *Polymers*, *15*(2), 425. https://doi.org/10.3390/polym15020425

Studier F. W. (2014). Stable expression clones and auto-induction for protein production in E. coli. *Methods in molecular biology (Clifton, N.J.)*, *1091*, 17–32. https://doi.org/10.1007/978-1-62703-691-7_2

Su, L., Yang, L., Huang, S., Su, X., Li, Y., Wang, F., Wang, E., Kang, N., Xu, J., & Song, A. (2016). Comparative Gut Microbiomes of Four Species Representing the Higher and the Lower Termites. *Journal of insect science (Online)*, *16*(1), 97. https://doi.org/10.1093/jisesa/iew081

Suwan, E., Arthornthurasuk, S., & Kongsaeree, P. T. (2017). A metagenomic approach to discover a novel β-glucosidase from bovine rumens. *Pure and Applied Chemistry*, *89*(7), 941-950. https://doi.org/10.1515/pac-2016-0924

Tejirian, A., & Xu, F. (2010). Inhibition of cellulase-catalyzed lignocellulosic hydrolysis by iron and oxidative metal ions and complexes. *Applied and environmental microbiology*, *76*(23), 7673–7682. https://doi.org/10.1128/AEM.01376-10

Wu, Z., Peng, K., Zhang, Y., Wang, M., Yong, C., Chen, L., Qu, P., Huang, H., Sun, E., & Pan, M. (2022). Lignocellulose dissociation with biological pretreatment towards the biochemical platform: A review. *Materials today. Bio*, *16*, 100445. https://doi.org/10.1016/j.mtbio.2022.100445

Yang, Y., Zhang, Z. W., Liu, R. X., Ju, H. Y., Bian, X. K., Zhang, W. Z., Zhang, C. B., Yang, T., Guo, B., Xiao, C. L., Bai, H., & Lu, W. Y. (2021). Research progress in bioremediation of petroleum pollution. *Environmental science and pollution research international*, *28*(34), 46877–46893. https://doi.org/10.1007/s11356-021-15310-6

Zhang, W., Zhang, P., Wang, H., Li, J., & Dai, S. Y. (2022). Design of biomass-based renewable materials for environmental remediation. *Trends in biotechnology*, *40*(12), 1519–1534. https://doi.org/10.1016/j.tibtech.2022.09.011

Zhao, Y., Moser, C., Lindström, M. E., Henriksson, G., & Li, J. (2017). Cellulose Nanofibers from Softwood, Hardwood, and Tunicate: Preparation-Structure-Film Performance Interrelation. *ACS applied materials & interfaces*, *9*(15), 13508–13519. https://doi.org/10.1021/acsami.7b01738

Zheng, W., Wang, Y., Xu, Y., Liao, F., Li, T., Li, X., ... & Zang, Y. (2023). Genome-wide identification of GH9 gene family and the assessment of its role during fruit abscission zone formation in Vaccinium ashei. 17 April 2023, PREPRINT (Version 1) available at Research Square. https://doi.org/10.21203/rs.3.rs-2783726/v1

Zhou, J., Duan, J., Gao, M., Wang, Y., Wang, X., & Zhao, K. (2019). Diversity, Roles, and Biotechnological Applications of Symbiotic Microorganisms in the Gut of

Termite. *Current microbiology*, *76*(6), 755–761. https://doi.org/10.1007/s00284-018-1502-4

Zoghlami, A., & Paës, G. (2019). Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. *Frontiers in chemistry*, *7*, 874. https://doi.org/10.3389/fchem.2019.00874

8.0 Appendix

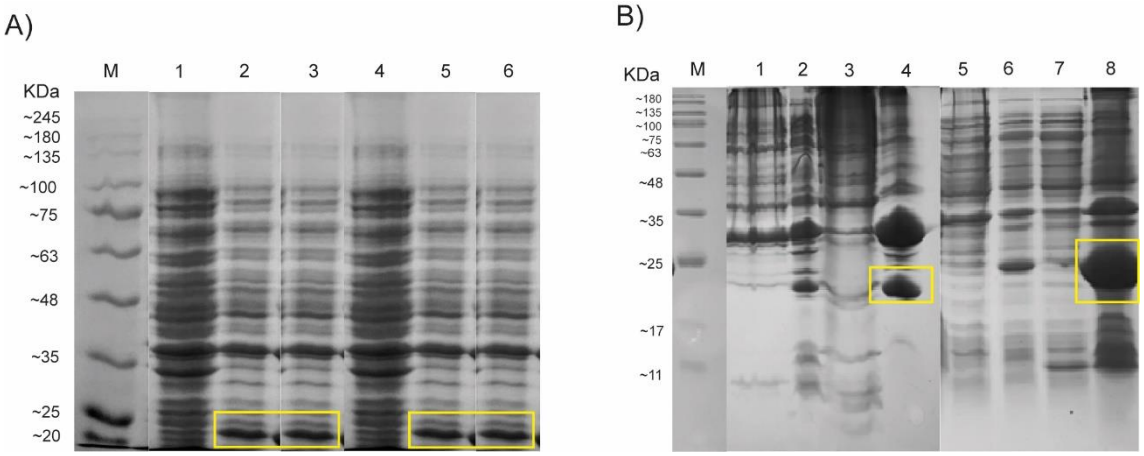


Fig. S1 Optimization of Exo8574 gene expression by variation of inducer and temperature. **A** Exo8574 induced with IPTG. M: marker; 1: pre-inoculum; 2: 0.1 mM IPTG and 28 °C pellet (insoluble); 3: 1mM IPTG and 28 °C pellet (insoluble); 4: pre-inoculum; 5: 0.1 mM IPTG and 37 °C pellet (insoluble); 6: 1mM IPTG and 37 °C pellet (insoluble). **B** Exo8574 with lactose autoinduction. M: marker; 1: pre-inoculum; 2: Induction (28 °C); 3: soluble; 4: pellet (insoluble); 5: pre-inoculum; 6: Induction (37 °C); 7: soluble; 8: pellet (insoluble). The yellow square was used to highlight the induced proteins.

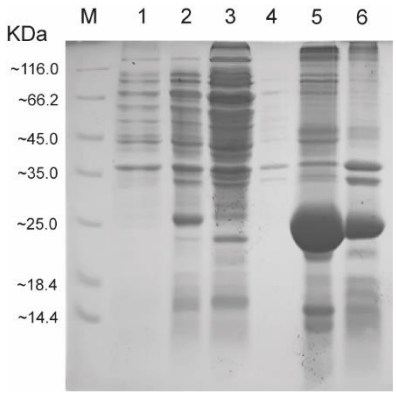


Fig. S2 Solubilization of Exo8574 inclusion bodies. M: marker; 1: pre-inoculum; 2: induction; 3: soluble; 4: water wash; 5: 2 M urea and 6: pellet (insoluble).

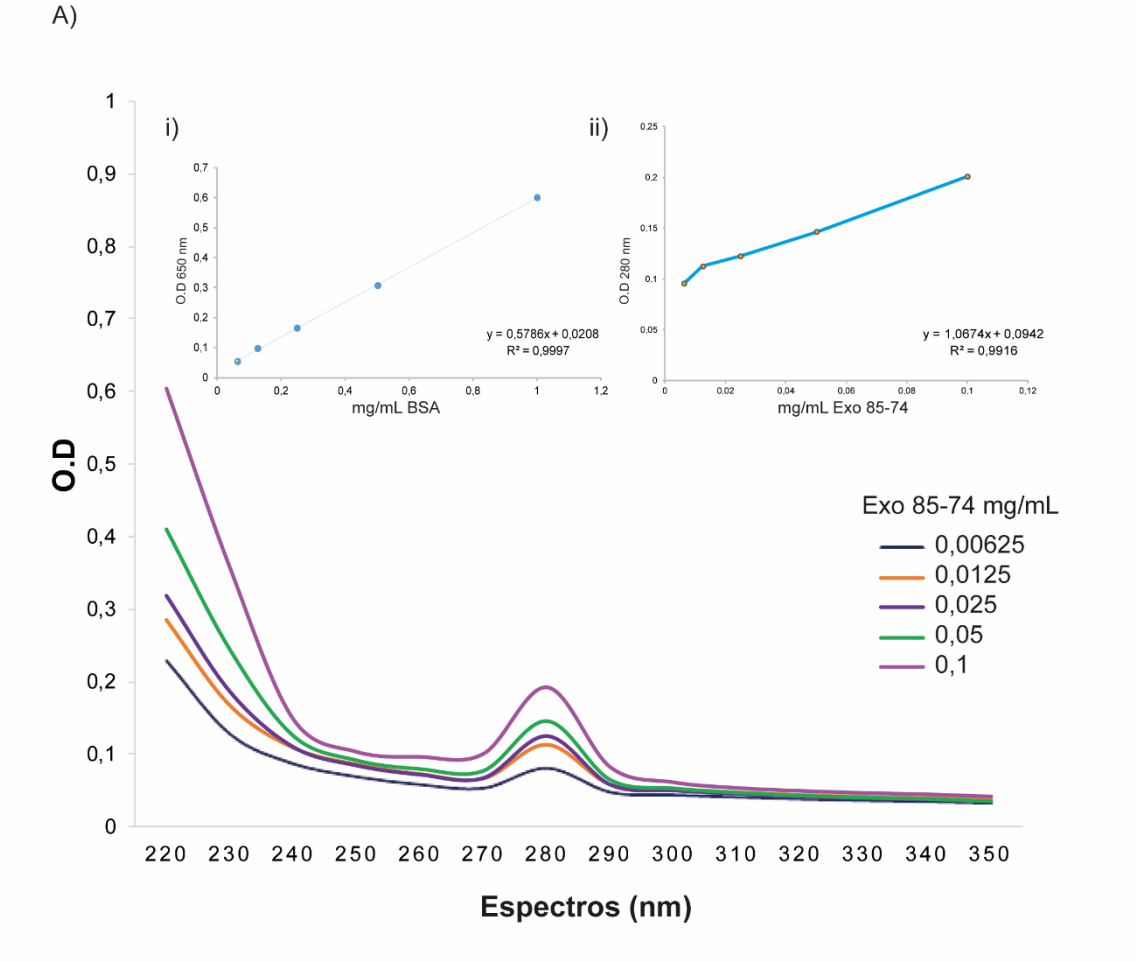


Fig. S3 Experimental molar extinction coefficient of Exo8574. A The molar extinction coefficient of Exo8574 was 1.067 ml.cm-1.mg-1. **i)** graph and line equation of BSA, relating $O.D_{650nm}$ x mg/mL; **ii)** graph and line equation of Exo8574, relating $O.D_{280nm}$ x mg/mL.

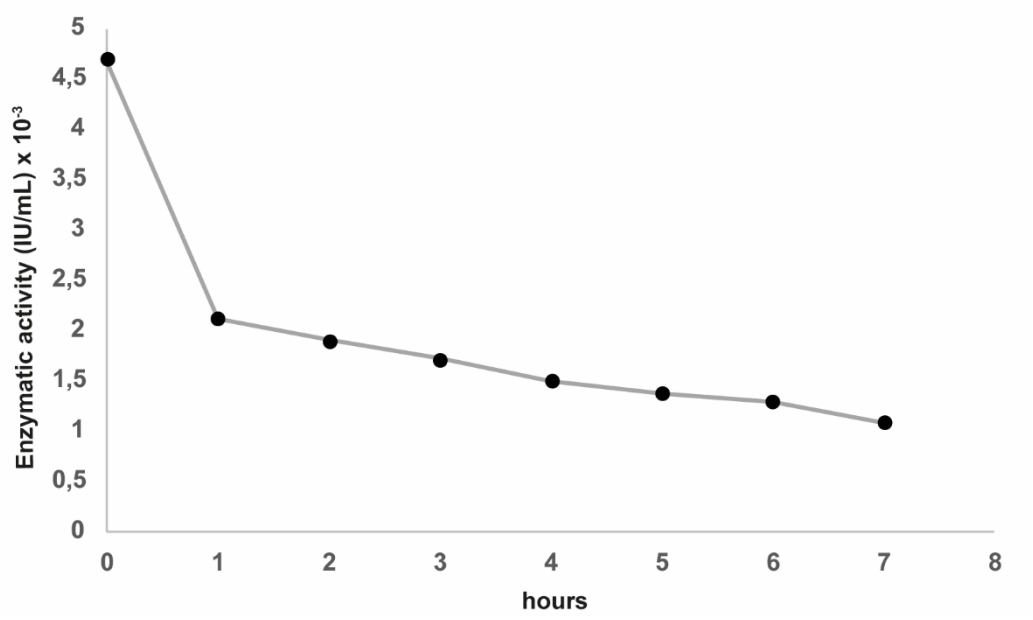


Fig. S4 Thermostability assay of Exo8574 at 55°C, pH 6.0. Activity measured every 1h for 7h in total.

Capítulo 4

Functional and structural characterization of a novel GH3 β-glucosidase from the gut metagenome of the Brazilian Cerrado termite *Syntermes wheeleri*[1]

[1] Artigo publicado no periódico *International Journal of Biological Macromolecules* https://doi.org/10.1016/j.ijbiomac.2020.09.236

1.0 Abstract

In this study, a GH3 family β-glucosidase (Bgl7226) from the metagenomic sequences of *Syntermes wheeleri* gut, a Brazilian Cerrado termite, was expressed, purified and characterized. The enzyme showed two optimum pHs (pH 7 and pH 10), and a maximum optimum temperature of about 40 ºC using 4-Nitrophenyl β-D-glucopyranoside (pNPG) as substrate. Bgl7226 showed higher enzymatic activity at basic pH, but higher affinity (K_m) at neutral pH. However, at neutral pH the Bgl7226 enzyme showed higher catalytic efficiency (k_{cat}/K_m) for pNPG substrate. Predictive analysis about the enzyme structure-function relationship by sequence alignment suggested the presence of multi-domains and conserved catalytic sites. Circular dichroism results showed that the secondary structure composition of the enzyme is pH-dependent. Small conformational changes occurred close to the optimum temperature of 40 °C, and seem important for the highest activity of Bgl7226 observed at pH 7 and 10. In addition, the small transition in the unfolding curves close to 40° C is typical of intermediates associated with proteins structured in several domains. Bgl7226 has significant β-glucosidase activity which could be attractive for biotechnological applications, such as plant roots detoxification, specifically our group is interested in cassava roots (*Manihot esculenta*) detoxification.

Keywords: GH3-β-glucosidase, *Syntermes wheeleri*, metagenome

2.0 Introduction

Bioprospecting is defined as the exploration of biodiversity for commercially valuable biochemical and genetic resources to achieve economic and conservation objectives. To achieve these purposes, the existing diversity of microorganisms can act as a reservoir of resources from which individual species with special characteristics can be explored [1]. Microorganisms comprise the majority of soil biomass and diversity and are crucial for its function. Microbial diversity and the pathways used for the production of enzymes, metabolites, bioactive compounds, together with more advanced biotechnological tools such as functional genomics, metagenomics, green chemistry, systems biology, synthetic biology, bioinformatics and metabolic engineering may become an ideal platform for the production of value-added products [2, 3].

The study of microorganisms is traditionally performed by laboratory cultivation techniques, however, only 0.1 to 1% can be considered cultivable in these conditions [4]. Over the years, many techniques have emerged to improve the analysis of "non-cultivable" majority, or rather, those of difficult cultivation. Metagenomics works, which use total DNA of a microbial community, have been widely used in bioprospecting to assess the potential of species present in natural ecosystems, such as thermal sources and termite guts for obtaining thermoenzymes and lignocellulose-enzymes, respectively, for example [4-6]. This approach has enabled the discovery of new genes originated from "non-cultivable" microorganisms that encode enzymes with biochemical and biophysical characteristics compatible with industrial applications [7].

Biorefineries have being using lignocellulose biomass to obtain co-products [8]. Lignocellulose is the most abundant biopolymer in terrestrial environments [9]. It is composed of cellulose (30-60%), hemicellulose (14-40%) and lignin (7-25%), in terms of dry matter. Cellulose is a linear biopolymer made up of *β*-D glucopyranose units covalently linked with 1→4 glycosidic bonds composing the cell walls of plants [10]. Carbohydrate structural diversity is controlled by a set of enzymes involved in their assembly (Glycosyl transferases-GTs), breakdown (Glycosyl hydrolases-GHs, polysaccharide lyases – PLs and carbohydrates esterases-CEs), and also part of the enzymes for the auxiliary activities – AAs, and for the recognition of carbohydrates (carbohydrate-binding module - CBM), collectively designated as carbohydrate-active enzymes (CAZymes). CAZymes have been classified into families based on sequence alignments for more than 27 years, and this classification has become standard analysis [11]. Cellulases (i.e., β-1.4-endoglucanases and β-1.4-celobiohidrolases), as well as β-glucosidases are glycosyl hydrolases or GHs. The current classification of GHs is based on the similarity of amino acids [12]. An updated version of this classification is available at the CAZy website (*http://www.cazy.org*) [13]. Currently, GHs are grouped into 167 families [14].

Complete hydrolysis of cellulose into glucose requires synergistic action of at least three types of enzymes: β-1.4-endoglucanases (EC 3.2.1.4), β-1.4-celobiohidrolases (EC 3.2.1.91 and EC 3.2.1.176) and β-glucosidases (EC 3.2.1.21). β-glucosidases hydrolyze cellobiose, a homo-disaccharide β-1.4 of glucose, into two molecules of glucose, but some of them can also release glucose from non-reducing extremities of soluble cellobio-oligosaccharides [15]. β-glucosidases have been used in industrial applications including aroma improvement of juices and for

releasing aromatic compounds in wine, hydrolysis of soybean to produce isoflavone glycosides, toxicity reduction of animal feed, production of anti-cancer compounds, and cellulose degradation for lignocellulosic biomass conversion in biofuel production [16]. Due to their varied roles in nature, these versatile enzymes can be useful in various synthetic reactions, as reviewed by [17].

In this work, we aimed to bioprospect the gut metagenome of an endemic "higher termite", *Syntermes wheeleri,* from Brazil's Cerrado biome (a savanna-like vegetation), to find β-glucosidases. During evolution, "higher termites" acquired symbiotic protists and lost the cellulolytic flagellates, and these protists are responsible for a significant cellulolytic activity in the gut [18, 19]. Termites consume 3 to 7 billion tons of lignocellulosic materials per year, being one of the most prolific and efficient decomposers of lignocellulose on Earth [20]. *S. wheeleri* is a neotropical termite species part of the Termitidae (Syntermitinae) family which feeds on litter (dry leaves). It stands out for its size with the *S. wheeleri* soldier measuring approximately 2 cm, while the majority of species measure between 0.3 and 1 cm. In Brazil, *S. wheeleri* mainly occurs in the Cerrado [21].

Although several studies have focused on endogenous endoglucanases of termites and cellulases from termite gut flagellates [22, 23], CAZymes from the higher termite gut bacteria have also been receiving scientific attention since the first report of the metagenomic analysis in the hindgut [24]. Recent studies have started to unveil a wide array of bacterial CAZymes genes encoding lignocellulose-degrading enzymes from termite gut [25-28]. Our previous work on 16S rRNA gene phylogeny profile of higher termite gut samples revealed the presence of highly diverse families of bacteria, including several species considered "non-cultivable"

[29]. "Non-cultivable" bacteria in the termite's gut are a potential source for the discovery of new genes that encode enzymes that hydrolyze lignocellulose [30]. Here, we describe for the first time the characterization of a new bacterial GH3 β-glucosidase from the metagenome of the gut microbiota of a Termitidae species, *Syntermes wheeleri*.

3.0 Material and Methods

3.1 Metagenomic analyses

Syntermes wheeleri gut metagenome sequences were used to assemble genomes and make gene annotations using the platform ggKbase (University of California at Berkeley, California, USA - https://ggkbase.berkeley.edu/, Kruger et al, manuscript in preparation). To select the genes encoding for β-glucosidases the genomic binning of *S. wheeleri* gut metagenome was searched using EC 3.2.1.21 to identify β-glucosidases. All sequences identified were exported as FASTA format proteins to be analyzed in Geneious 10.2 (https://www.geneious.com). P11, P12, P13, P31, P32, P33, WG1, WG2, WG3 (P1: first proctodeal segment, P3: third proctodeal segment and WG: whole gut) are samples from different sections of the *S. wheeleri* gut in 3 biological replicates.

3.2 Alignments, family annotations and phylogenetic analysis

The *S. wheeleri* gut metagenome selected protein sequences were subjected to multiple alignment using the MAFFT alignment in Geneious 10.2

(https://www.geneious.com). Sequences were edited to eliminate those with less than 200 amino acids that did not present start and stop codons, as well as those that produced a high number of gaps in the alignment. A new updated alignment was obtained each time a sequence was deleted. To assign sequences to Glycosyl Hydrolase (GH) families, they were subjected to BLAST protein-protein NCBI analysis using Geneious 10.2. After that, sequences within the same GH family were identified by searching the curated databanks CAZy (http://www.cazy.org/) and UniProt (https://www.uniprot.org/) with the purpose of aligning them with the *S. wheeleri* gut metagenomic sequences. This alignment was performed using the MAAFT alignment in Geneious 10.2, and the phylogenetic trees were calculated and displayed using FastTree with the same software.

3.3 Plasmid construction

The gene chosen to be characterized was selected according to the *S. wheeleri* metatranscriptome [29]. The table S1 shows information about the genes as family, phylum, organism, identity, genome integrity and gene expression values. From this table, the β-glucosidase that presented the highest metatranscriptome value was selected, as it implies high mRNA expression. The selected gene sequence was synthesized (GenOne Biotechnologies, Rio de Janeiro, Brazil - http://www.genone.com.br/) and cloned into the pET24a (+) vector that carries a C-terminal His-tag (Fig. S1). The synthesized gene sequence was codon optimized for expression in *Escherichia coli* and unwanted restriction sites were removed.

3.4 Protein expression

Escherichia coli BL21 (DE3) was transformed with the plasmid by electroporation (Gene Pulser Xcell™ System, BIO-RAD). Transformed colonies were selected in agar LB (LB Broth, Sigma-Aldrich) supplemented with 100 µg/mL kanamycin. To verify the production of protein, the clone BGL7226 (*E. coli* BL21 transformed with the plasmid carrying the β-glucosidase gene) was grown in 5 mL LB broth until the $O.D_{600}$ reached 0.6 to 0.9. At this point, 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and protein expression was induced for 3 h at 37 °C and 200 rpm (Incubator TE-424, TECNAL, Brazil). Aliquots of cells were collected to evaluate protein expression profiles by 12% SDS-PAGE conducted under denaturing conditions as previously described [31].

3.5 Bgl7226 expression optimization and solubility test

Clone BGL7226 was inoculated into 25 mL of LB media and grown until $O.D_{600}$ was between 0.6 and 0.9. At this point, two different conditions (i.e., 0.1 mM IPTG and overnight growth at 28 °C and 200 rpm, and 1 mM IPTG and growth at 37 °C for 3 h at 200 rpm) were used test which was the best condition for protein production. Aliquots of 1 mL of pre-inoculum cells and cells after induction were harvested and stocked at 4 °C for further analysis. After induction, 24 mL of cells were centrifuged at 16 000 x g for 30 min at 4 °C. The supernatant was discarded. The resulting pellet was resuspended in 3.5 mL of buffer A (30 mM Tris-HCl, pH 8.0 and 300 mM NaCl), and then sonicated using a Branson Ultrasonics™ Sonifier S-250A sonicator using the following parameters: 15 min, duty cycle (30%) and

output control level 4. The lysate was centrifuged at 16 000 x g for 30 min at 4 °C, and the resulting pellet was resuspended in 3.5 mL of buffer A, 1 mL of which was saved for analysis, and the supernatant stocked at 4 °C. A 12% SDS-PAGE was used to analyze the protein profile.

3.6 Protein purification

A 500 mL-culture of clone BGL7226 was grown under the following inductive conditions: 0.1 mM IPTG, at 28 °C overnight (16 h) and 200 rpm. The pellet was lysed as previously described, and subsequently centrifuged. The new pellet was resuspended in 40 mL of buffer B (30 mM Tris-HCl - pH 8.0, 300 mM NaCl and 1M urea) and left overnight under agitation of 100 rpm at 25 °C to solubilize inclusion bodies by action of the chaotropic agent urea. After this, the content was centrifuged at 16000 x g for 30 min at 4 °C. The supernatant was concentrated using Amicon™ (MerkMillipore) and then applied to a molecular exclusion column (Superdex S200, GE) for simultaneous purification and on-column renaturation. The sample was injected with a flow of 1.6 mL/min in the pre-balanced column with buffer A.

The protein was collected from an expected column volume (~72 mL), since the calibration of the column had already been performed with other proteins of known size. Protein peaks were collected and analyzed by SDS-PAGE and, in addition, the peak fractions of the Bgl7226 protein were analyzed separately and stained with silver nitrate [32]. In addition, an activity test was performed for each of the fractions of this peak, adding 25 µl of 4 mM p-Nitrophenyl β-D-Glucopyranoside

(pNPG, Sigma-Aldrich); 25 µL of 100 mM sodium acetate pH 5.0 (Sigma-Aldrich) + 25 µl of the enzyme Bgl7226 [33]. The blank was the same without the addition of enzymes. This assay was performed on a 96-well plate in a SpectraMax® M3 (Molecular Devices), reactions were incubated for 30 min at 45 °C. To stop the reactions, 150 µl 1 M sodium carbonate (Sigma-Aldrich) was added, and absorbance was measured at 405 nm. The graphics were obtained using OrigenPro 8 (OriginLab, Northampton, MA, USA).

3.7 Molar extinction coefficient and protein concentration

The experimental molar extinction coefficient (or molar absorptivity) of the Bgl7226 protein ($\varepsilon 280^{nm}_{mg/ml}$) was determined by recording the absorbance at 280 nm in relation to the protein concentration (mg/mL) determined by the Lowry method [34]. A UV/Visible Spectrophotometer Jasco V-530 (Jasco Corporation, Tokyo, Japan) was used for (i) fixed point measurement, and (ii) corrected absorbance considering light scattering at 350 nm (A350) (Eq. 1).

$$Acorr280 = A280 - A350 \quad \text{Eq. (1)}$$

The Lowry method was calibrated using the known concentration of bovine serum albumin (BSA) protein. This protein was prepared in five different concentrations: 0.06, 0.125, 0.25, 0.5 and 1 mg/mL. From the linear graph of absorbance at 280 nm for Bgl7226 and the respective protein concentrations, determined by the Lowry method, the angular slope was obtained as a measure of the extinction coefficient.

3.8 Enzymatic assays.

3.8.1 p-Nitrophenyl standard curve (pNP)

The construction of the standard pNP curve, used for enzymatic activity calculations, was constructed using six different concentrations of this product, 0.06, 0.12, 0.25, 0.5, 1.0 and 1.5 mM, in a volume of 75 µl. The solutions were heated to 45 °C for 30 min in the SpectraMax M3 (Molecular Devices, San Jose, CA, USA) plate spectrophotometer and then 150 µl of 1M Sodium carbonate was added. Absorbances at 405 nm were read in the SpectraMax M3 and the values were exported and analyzed in OrigenPro 8 (OriginLab, Northampton, MA, USA).

3.8.2 Effect of Bgl7226 concentration on enzymatic reaction

Assays to determine the effect of Bgl7226 enzyme concentration on enzymatic hydrolysis of pNPG were performed in 96-wells plates in triplicate. The protein concentrations analyzed were 200; 150; 100; 75; 50 and 31 nM. The discontinuous enzymatic assay was performed to obtain the initial velocity (V_0) of the reactions with 12 absorbance measurements: time 0 s and approximate intervals of the following 8 readings were made at each 40 s, and the last 3 readings in the time of 7, 15 and 30 min. The temperature of assay was 50 °C. Reactions were interrupted by addition of 1 M sodium carbonate stop solution. Data were exported to OrigenPro 8 (OriginLab, Northampton, MA, USA) for analysis.

3.8.3 Effect of the pH and temperature

Assays to determine pH influence on Bgl7226 enzymatic activity were performed in 96 wells plates, using different pHs values at 50 °C for 30 min. Buffers in a concentration of 100 mM, were prepared in the pH values of 3, 4 and 5 (citric acid/sodium citrate), 6 and 7 (monobasic sodium phosphate/dibasic sodium phosphate), 8 and 9 (Tris base/hydrochloric acid), 10 (glycine/sodium hydroxide), 11 (sodium bicarbonate/sodium hydroxide) and 12 (sodium chloride/sodium hydroxide). Enzymatic reactions were performed using the enzyme concentration previously selected (25 µL of enzyme at 0.1 µmol/L ≈ 0.25 $x10^{-11}$ mol in assays), and readings were performed in the SpectraMax M3 (Molecular Devices, San Jose, CA, USA), at intervals similar to those described, in order to obtain the V_0 of the reaction. Data were analyzed in OrigenPro 8 (OriginLab, Northampton, MA, USA).

The experiment to determine the influence of temperature was performed in 200µl-tubes in a ThermoCycler (T100 thermal cyclers, BIO-RAD, CA, USA) with temperatures of 20 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 60 °C, 70 °C, 80 °C and 90 °C, and time of 30 min. The assays were performed using buffer for the optimum pH for the enzyme. In this experiment, besides the blank and negative control (C-), a positive control (C+) was used (i.e., pNP product rather than the pNPG substrate). The values of C+ were considered to be 100%, and were used to calculate relative activity. Absorbance readings at 405 nm were determined using a SpectraMax M3 (Molecular Devices, San Jose, CA, USA).

3.8.4 Kinetic parameters - Michaelis-Menten curve

Bgl7226 kinetic parameters Michaelis-Menten constant (K_m), maximum rate (V_{max}), catalytic constant (k_{cat}), and k_{cat}/K_m were obtained from a Michaelis-Menten plot, which was constructed using pNPG substrate concentrations of 100; 200; 300; 400; 600; 800; 1000; 1200 and 1400 μM, and maintaining fixed the enzyme concentration at 0.1 μmol/L (25 μL of enzyme ≈0.25 $x10^{-11}$ mol). The discontinuous enzymatic assays were performed using the optimum pH and temperature. Enzymatic activity (EA) was calculated from absorbance at 405 nm by following equation and using OriginPro 8 (OriginLab, Northampton, MA, USA):

$$EA = \left(\frac{V_0}{\varepsilon_{pNP\ 405nm}}\right) * V_{assay} \qquad \text{Eq. (2)}$$

where EA is enzyme activity (kat or mol(pNP)/s), V_0 is initial velocity (absorbance arbitrary units/s), $\varepsilon_{pNP\ 405nm}$ is the molar extinction coefficient of the product p-nitrophenol (pNP) (L/mol cm), and V_{assay} is the assay volume (L). According to the international system of units, the unit of enzyme activity is the "*katal*" (kat = 1 mol/s), that converts one mole of substrate per second under specified reaction conditions [35].

3.9 Secondary structure and structural stability analysis by circular dichroism

The Bgl7226 protein secondary structure profile was characterized by circular dichroism (CD) using a spectropolarometer Jasco J-815 (Jasco Corporation, Tokyo, Japan) equipped with a Peltier temperature control system (Analytical Instruments, Japan). The dichroic spectra were obtained in a 0.1 cm quartz cuvette at 25°C. The

pH dependence tests were performed with the protein at a concentration of 0.066 mg/mL. Buffers used in the tests were 5 mM sodium acetate, pH 4.0; 5 mM Tris-HCl, pH 7.0 and pH 8.5 and 5 mM glycine buffer/sodium hydroxide, pH 10.0. Five dichroic spectra were obtained at the far UV region (190-260 nm) with intervals of 0.2 nm, and a scan rate of 100 nm/min, response time of 4 s, and band width of 1.0; and the mean values were corrected by subtracting the contribution of buffers.

The elipticities (mdg) obtained were converted into molar ellipticity ([θ]) (deg.cm^2. dmol^{-1}) based on the mean residue molecular mass of 115 Da [36]. The dependence of the secondary structure on the pH was estimated from the spectra adjusted using the software CD Spectra Deconvolution Vs 2.1 CDNN [37].

Thermostability curves were obtained for pH 4.0, 7.0 and 10.0 at 208 nm with the temperature varying from 25 to 95 °C, at a rate of 0.2 °C /min. In parallel, dichroic spectra were recorded in the far UV (190-260 nm) region, with intervals of 10 °C. Thermal denaturation curves were obtained considering the values of molar ellipticity ([θ]) versus temperature [36].

3.10 Bgl7226 structure analysis

The homology model of Bgl7226 was generated by Modeller 9v23 using as template the enzyme with the highest score found by a BLASTp search in the Protein Data Bank (PDB) [38, 39]: BaBgl3 from *Bifidobacterium adolescentis* (PDB entry: 5WAB). Structural adjustment was made in PyMol [40]. Energy minimization of the modified model was initially performed in 3D refiner and then in UCSF Chimera with default parameters [41, 42]. The quality of the refined structure was evaluated by the

Ramachandran plot generated by RAMPAGE, by Z-score generated by QMEAN and by average 3D-1D score calculated by Verify3D [43-45]. The structure depictions were generated using PyMol (40). The multiple alignment was then made using the T-COFFEE Multiple Sequence Alignment Server software and visualized with the ESPrint 3.0 software, which allows displaying the alignment along with the representation of the secondary structure [46, 47].

4.0 Results

4.1 Metagenomic sequences analysis

A total of 222 β-glucosidase protein sequences from *S. wheeleri* gut were imported from ggKbase to Geneious in FASTA protein format. During edition, 174 sequences were excluded, and the remaining 48 were classified as glycoside hydrolases belonging to the GH3 family. High bootstrap values (>70%) for clades of termite and cured database sequences supported them as belonging to family GH3 [48] (Fig. 1). The sequences are listed on Table S1 – supplemental material, and metatranscriptome results were the main factor to select the sequence to be studied. The chosen gene, named *Bgl7226*, was taken from the scaffold P33_assembly_scaffold_72_26 because it presented the highest coverage per million (0.611) among the *S. wheeleri* gut metatranscriptome β-glucosidases, in other words, it presented more reads mapping to the metagenome, which could indicate a higher level of mRNA expression. The *Bgl7226* gene was cloned into pET24a (+) plasmid for protein production (Fig. S1).

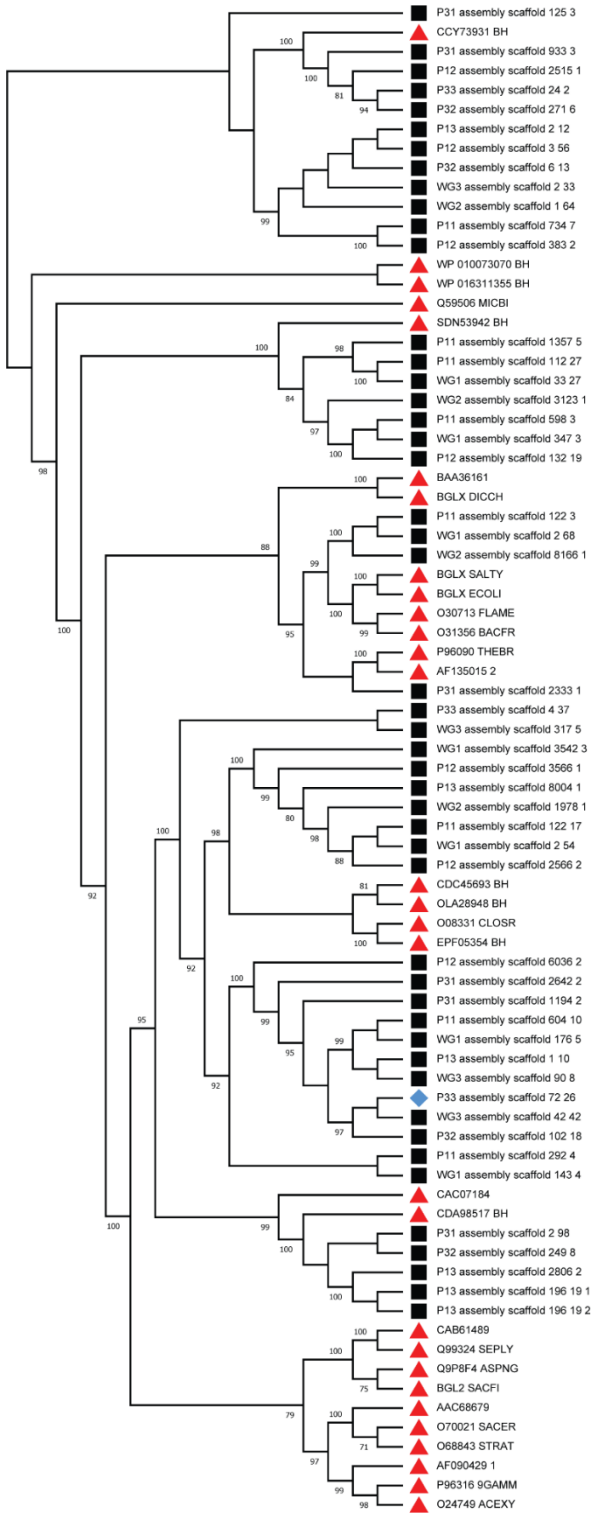


Fig1. Phylogenetic tree of glycosyl hydrolase family 3 (GH3) β-glucosidases from 9 metagenomes assemblies. *Syntermes wheeleri* gut metagenomic GH3 β-glucosidases are marked with black squares (■), β-glucosidases from a cured database (CAZy) are marked with red triangles (▲), and Bgl7226 is marked with a blue diamond (◆). All branches supported by bootstrap values above 70% are shown. P11, P12, P13, P31, P32, P33, WG1, WG2, WG3 are samples from different sections of the *Syntermes wheeleri* gut in 3 biological replicates.

4.2 Gene expression, solubility test and optimization

After induction with IPTG, SDS-PAGE revealed a specific band with a molecular mass of approximately 90 kDa (Fig. S2A), similar to the theoretical molecular mass of 89.9 kDa, based on the Bgl7226 amino acid sequence. Analysis of protein expression in the cell culture supernatant and pellet was performed to determine whether the protein produced was soluble or insoluble. A protein of 89.9 kDa appears almost exclusively in the cell culture pellet, probably in inclusion bodies (Fig. S2B). Temperature for cell growth and the IPTG concentration used for induction of protein expression were changed, however, Bgl7226 protein continued to be produced mostly in the insoluble fraction (Fig. S3).

4.3 Protein purification and molar extinction coefficient

Bgl7226 purification was performed by molecular exclusion technique, yielding several peaks (Fig. 2A). SDS-PAGE of peak 3 (P3) lead to a single band, both in Coomassie blue and silver nitrate stained gels (Fig. 2Aii and 2Bi). Rechromatography of peak 3 and enzyme activity assay using pNPG as substrate showed that peaks b, c and d present higher activity agreeing with highest Abs280 (Fig. 2B).

The experimental molar extinction coefficient ($\varepsilon^{280nm}_{mg/mL}$) of Bgl7226 was determined by combination of absorbance reading at 280 nm and the Lowry method. The Bgl7226 absorption spectra (Fig. S4) in solutions of different concentrations (mg/mL) was determined by the Lowry method using BSA as the standard protein (Fig. S4Ai). The protein extinction coefficient was determined from

the linear ratio of absorbance at 280 nm and protein concentration, $\varepsilon^{280nm}_{mg/mL}$ (Fig. S4Aii), as being 1.36 $ml.cm^{-1}.mg^{-1}$.

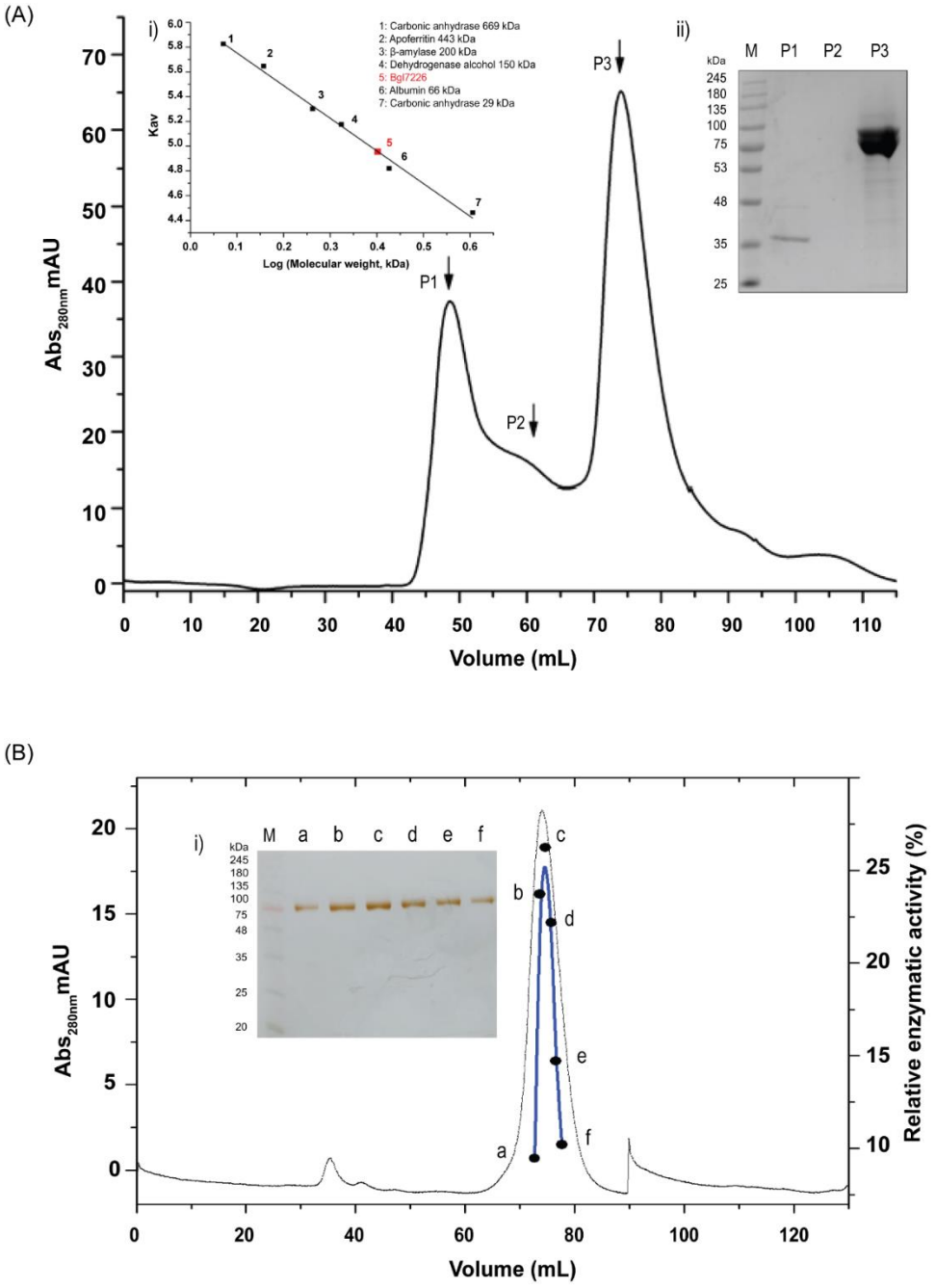


Fig 2. Purification of Bgl7226. (A) Chromatographic profile for Bgl7226 purification by molecular exclusion chromatography in denaturing conditions, using the Superdex S200 column (GE). Bgl7226 protein was eluted in the expected volume (72 mL); i) Superdex S200 calibration curve; ii) SDS-PAGE gel from chromatography peaks (P1, P2 and P3), P3 contains Bgl7226 (90 kDa). (B) Re-chromatogram of P3 showing a major peak collected in 6 fractions, named 'a' to 'f'. The relative enzymatic activity is represented as a blue line with black dots at the end points of each corresponding fraction; i) SDS-PAGE gel from the P3 fractions (a, b, c, d, e and f) stained with silver nitrate. M is the BLUeye pre-stained protein ladder (Merck).

4.4 Biochemical characterization of Bgl7226

Enzyme activity calculations were performed using a standard pNP concentration curve (Fig. S5). Before enzymatic characterization, a discontinuous enzymatic assay was performed to determine the influence of enzyme concentration on the enzyme activity using pNPG as substrate. This study allowed selection of the appropriate enzyme concentration for enzyme characterization and determination of kinetic parameters. In this experiment, initial velocities (V_0) were determined up to approximately 400 seconds, using different concentrations of enzyme. A fixed non-limiting concentration of substrate, approximately 10-times the maximum enzyme concentration analyzed (1.3 mM), was used. As expected, $O.D_{405nm}$ increases with enzyme concentration, and consequently higher enzymatic activities were determined from initial velocities (V_0) (Fig. S6A). Linearity was observed between enzymatic activities, calculated from the initial rates (V_0) (Fig. S6A), as a function of the total (mg) (Fig. S6B). In addition, specific activity of 2.92 ± 0.22 $x10^{-9}$ mol of pNP/s/mg of enzyme (or kat/mg of enzyme) was estimated (Fig. S6B). According to this linear plot, we selected 0.1 μM of enzyme to use in subsequent kinetics assays for enzyme characterization.

The effect of pH and temperature on the enzyme activity was evaluated by discontinuous enzymatic assays using pNPG as substrate. A fixed non-limiting concentration of substrate (1.3 mM) and 25 μL of enzyme at 0.1 μM (0.25 $x10^{-11}$ mol) were used. From the V_0 determined up to approximately 400 seconds, at different pHs (Fig. 3A), a graph of enzymatic activity versus pH (Fig. 3B) was obtained. The enzymatic activities were calculated from V_0 using the Eq (2) and expressed as kat or mol/s (Fig. 3B). The optimal pHs for Bgl7226 activity were pH 7.0 and pH 10.0.

Enzymatic activity at pH 10.0 was 1.24-fold greater than at pH 7.0, with values of enzymatic activity of 7.24 $x10^{-11}$ ± 0.01 mol of pNP/s and 5.82 ± 0.01 mol of pNP/s, respectively (Fig. 3B). The effect of temperature on the relative enzyme activity was determined for each optimal pH using fixed non-limiting concentration of substrate (1.3 mM) and 100 nM of enzyme. In both pHs Bgl7226 relative enzymatic activity was highest at 40 °C (Fig. 3C).

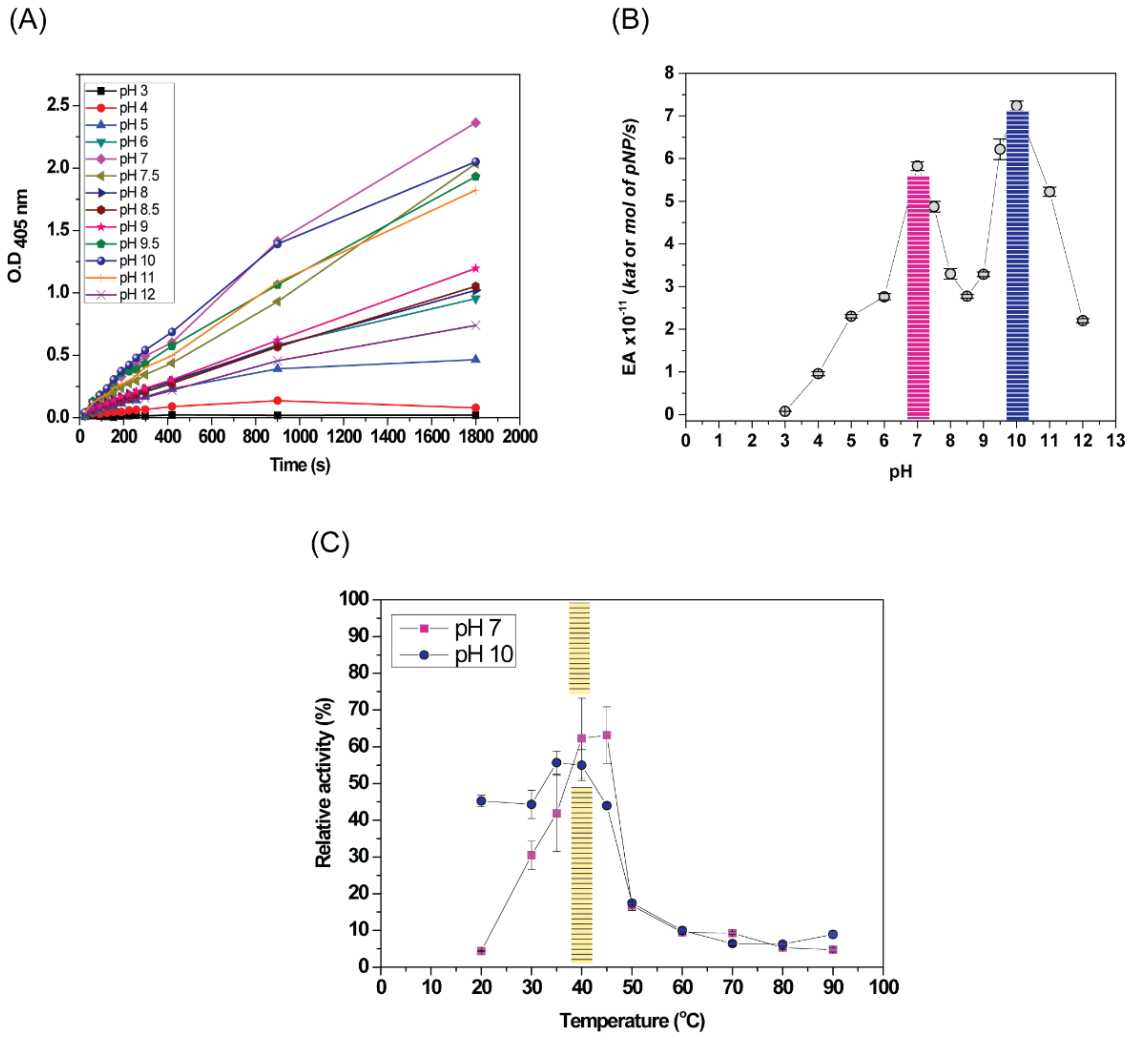


Fig3. Effect of pH and temperature on Bgl7226 activity. (A-B) The optimal pH for activity was determined by incubating the enzyme in buffers of various pH values at 50 °C for 20 min. The highest activities were observed at pH 7.0 and 10.0. (C) Effect of temperature on Bgl7226 activity. Enzyme activity was assayed in 100 mM monobasic sodium phosphate/dibasic sodium phosphate buffer (pH 7.0) and 100 mM glycine/sodium hydroxide buffer (pH 10.0) for 30 min at the indicated temperatures.

4.5 Bgl7226 kinetic parameters

Optimized discontinuous enzymatic assays were used to determine the effect of pNPG substrate concentrations on enzyme activity, using 25 µL of enzyme at 0.1 µM (0.25 x10^{-11} mol) at pH 7.0, pH 10.0 and 40 °C. The initial velocities (V_0) for each optimal pH (pH 7.0, Fig. 4A and pH 10.0, Fig. 4C) were determined up to approximately 300 seconds for different substrate concentrations. The Michaelis-Menten plots (at pH 7.0 and pH10.0) were obtained from the enzymatic activities calculated from V_0 using the Eq (2) and expressed as kat or mol/s (Fig. 4B and Fig. 4D, respectively). The V_{max} in pH 10.0 was 1.13-fold higher than at pH 7.0 (Table 1), indicating that Bgl7226 is more active at basic pH. However, as indicated by K_m, affinity for the pNPG substrate was 9.4-fold higher at pH 7.0 (Table 1). Bgl7226 k_{cat} for the pNPG substrate at pH 10.0 was higher than at pH 7.0, but the k_{cat}/K_m ratio or specificity constant (i.e., kinetic efficiency) was higher at pH 7.0. This result indicates that the kinetic efficiency of Bgl7226 is 8.17-fold higher at pH 7.0 than at pH 10.0 (Table 1).

Table 1. Kinetic parameters of Bgl7226 at pH 7.0 and pH 10.0.

Protein	pH	V_{max} x10^{-11} (mol/s)	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}. $µM^{-1}$)
Bgl7226	7.0	1.95 ± 0.09	53.10 ± 16.28	7.80 ± 0.52	0.147
	10.0	2.21 ± 0.40	498.93 ± 22.91	8.84 ± 0.20	0.018

25 uL of enzyme concentration at 0.1 µmol/L (0.25 x10^{-11} mol) was used in assays.

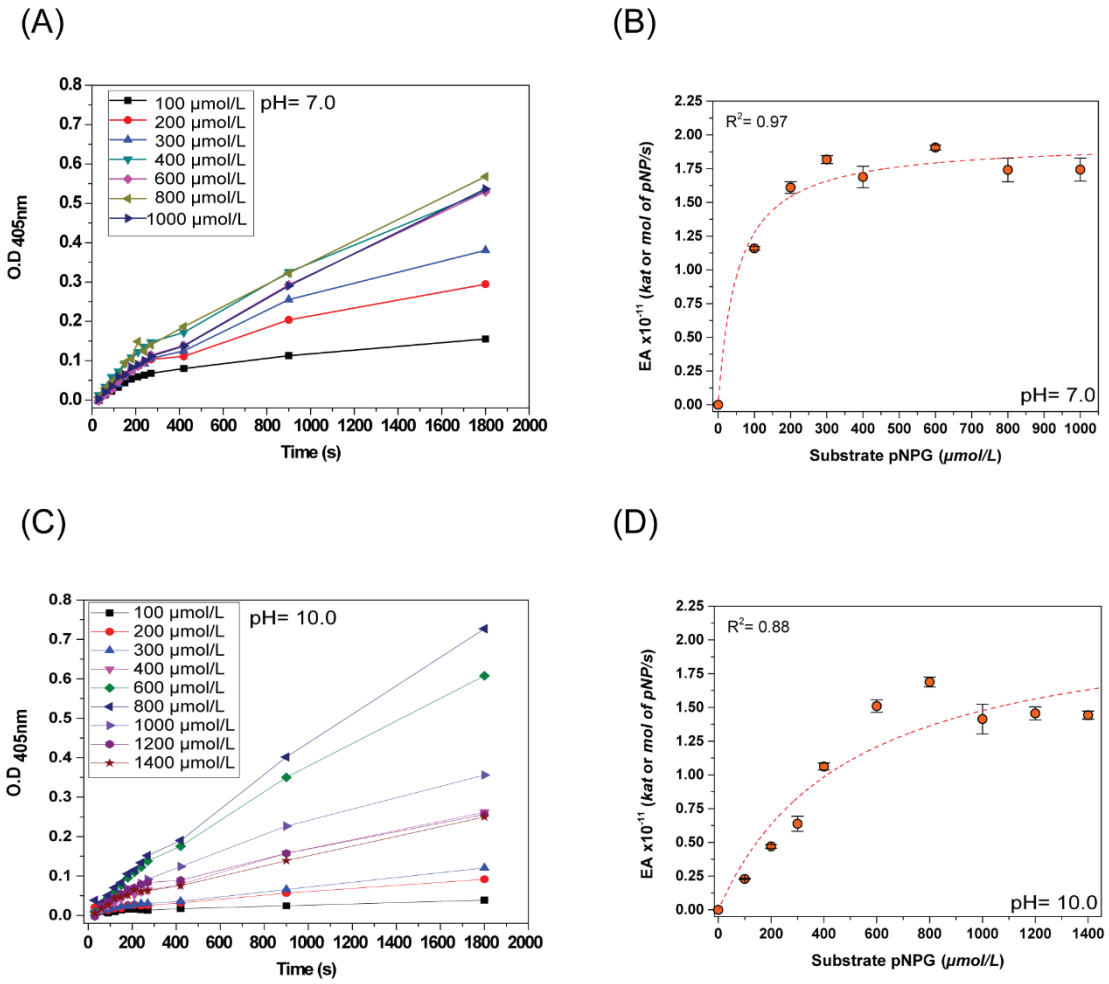


Fig4. Bgl7226 Michaelis-Menten curve at pH 7.0 and 10.0. (A-B) Michaelis-Menten plot of the kinetics of Bgl7226 using different pNPG concentrations in 100 mM monobasic sodium phosphate/dibasic sodium phosphate buffer (pH 7.0) at 40 °C. (C-D) Michaelis-Menten plot of Bgl7226 kinetics using different pNPG concentrations in 100 mM glycine/sodium hydroxide buffer (pH 10.0) at 40 °C.

4.6 Analysis of secondary structure and structural stability by Circular Dichroism (CD)

CD assays for Bgl7226 showed that the protein has secondary structure at 25.0 °C independent of pH; with similar dichroic spectra at pH 4.0, 7.0 and 8.5, showing little difference at pH 10, corresponding to the structured protein at all tested pHs (Fig. 5A and Table 2). The protein spectra have a pronounced negative band corresponding to the alpha-helix structure (208 nm and 222 nm) and beta-

sheet (218 nm). This result was observed at all pHs, however, at pH 10.0 the secondary structure content is higher for alpha-helix (27.3%), on the other hand, there was a reduction in the parallel and antiparallel beta-sheet content (15.6%), when compared to acid pH (22.1%). These results show that Bgl7226 is structured at all pHs, with no differences in the secondary structure content at 25.0 °C, except at pH 10.0, where the protein appears to be more structured than at pH 4.0, 7.0 and 8.5.

Thermostability assays show that at pH 4.0, the protein did not denature even at high temperatures, presenting similar spectra for temperatures ranging from 25.0 to 95.0 °C (Fig. 5B and 5C). At pH 7.0, the thermal unfolding curve showed a gradual reduction of the signal from approximately -8,500 deg.cm^2.dmol^{-1} to zero, reaching the unfolded state as the temperature increased to 95.0 °C (Fig. 5B). It is noteworthy that the unfolded curve shows a small transition close to 40.0 °C, and a larger transition close to melting temperature (Tm) of 65.0 °C, compatible with low conformational changes and a total unfolding process, respectively. This result is in agreement with those presented in Figure 5D, in which the Bgl7226 spectra drastically decrease the dichroic signal around 65.0-75.0 °C, close to Tm of 64.5 °C, and lose the dichroic signal around 85.0-95.0 °C, typical of the protein denaturing process. In addition, Bgl7226 did not show refolding capability, judging by a pattern of protein aggregation visualized at 95 °C. At pH 10.0, the unfolding curve shows a small transition close to 40.0 °C and a reduction in the dichroic signal of approximately -10,500 to -8,200 deg.cm^2.dmol^{-1} (from 25.0-95.0 °C) compatible with low conformational changes and a partial unfolding process (Fig. 5B). In addition, a marked reduction was observed in the dichroic bands at 208, 218 and 222 nm, with

displacement to the nearest region at 200 nm around 35.0-45.0 °C, which is characteristic of conformational changes and increase in the unordered structure (Fig. 5C-E). The results of thermostability show that Bgl7226 is more stable in the acidic range.

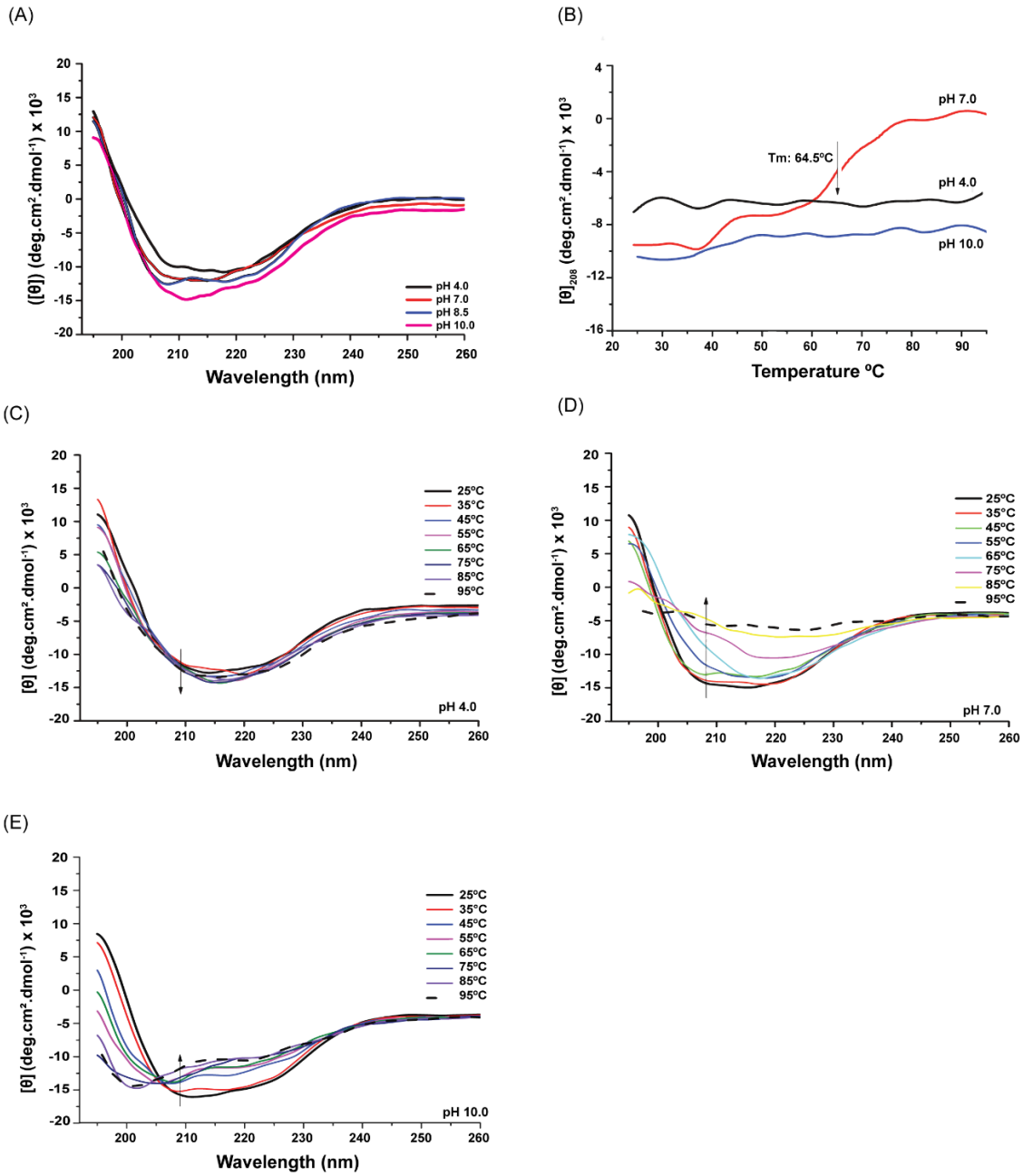


Fig5. Analysis of Bgl7226 secondary structure, and the effect of temperature on structural stability. (A) Bgl7226 dichroic spectra obtained by circular dichroism at different pHs at 25 °C. (B) Bgl7226 thermal denaturation curves monitored at 208 nm for pH 4.0, 7.0 and 10.0 with temperatures ranging from 25-95 °C. (C, D and E) Bgl7226 dichroic spectra collected between 25 and 95 °C at pH 4.0, 7.0 and 10.0. Tm (Tmelting) for the protein at pH 7.0 is indicated by arrow.

Table 2. Secondary structure content of the Bgl7226 enzyme in different pHs obtained by deconvolution of dichroic spectra using the CDNN program and from the homology model.

Secondary structure (%)	pH 4.0	pH 7.0	pH 8.5	pH 10.0	Bgl7226 Model
α-Helix	25.5	24.0	25.9	27.3	30.5
β-Anti-parallel	14.0	14.9	11.4	10.1	13.1
β-Parallel	6.2	5.8	5.7	5.4	4.5
β -Turn	16.8	17.5	16.6	19.4	12.4
Unordered	35.4	32.2	32.9	34.8	39.5

4.7 Bgl7226 structure analysis

Structural features of Bgl7226 were investigated by homology modeling and compared to closely related sequences and structures. Homology model validation shows a good quality of the model (Table 2 - supplemental). Additionally, the secondary structure of 30.5% α-helix, 17.5% β-strand, and 51.9% other obtained from the Bgl7226 model is in agreement with the same data collected from the deconvolution of circular dichroism spectra (Table 2). Based on the sequence and the three-dimensional structure alignment with BaBgl3 from *Bifidobacterium adolescentis* (PDB entry: 5WAB), the Bgl7226 is arranged in three domains connected by two linkers (residues 291-302 and 487-547) (Fig. 6A). Domain 1 (residues 1- 290) is folded as a TIM barrel. Domain 2 (residues 301- 486) has a α/β sandwich fold. Domain 3 has a fibronectin type III fold (FnIII) with β-strands. The active site is widely solvent-exposed and located at the C-terminal portion of the barrel, specifically, at the interface of domains 1 and 2 (Fig. 6B). Based on the sequence alignment, residue D233 should be the catalytic nucleophile located in domain 1, while E417 from domain 2 shall function as the general acid/base

residue. D46, R121, K154, H155, R165, and Y201 were identified to be positioned in hydrogen bonding distance to tentative substrate polar groups, thus forming the substrate-binding site. All of these residues are conserved, as shown in Figure 7.

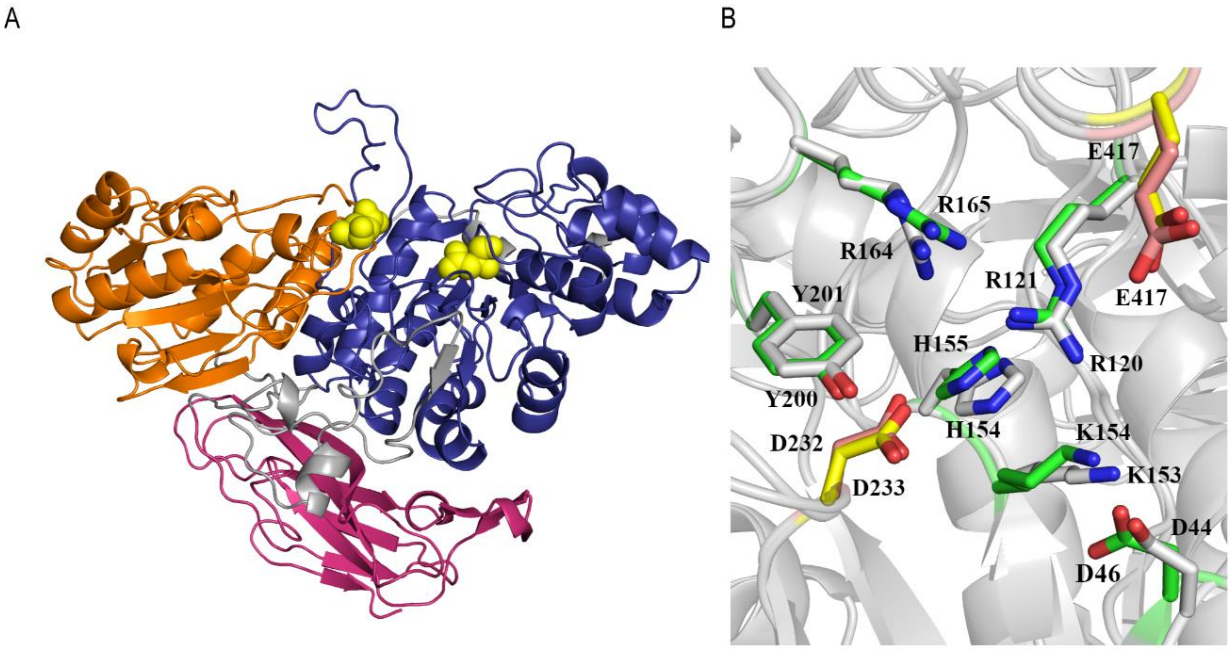


Fig6. Homology-modelling of Bgl7226 protein structure. (A) Cartoon illustration of the Bgl7226 overall structure colored by domain: domain 1 in deep blue, domain 2 in orange, and domain 3 in pink. D233 and E417, which are represented by yellow spheres, indicate the location of the active site. (B) Structure superposition of the Bgl7226 active site with BaBgl3 from *B. adolescentis* (PDB entry 5WAB) with an r.m.s.d of 0.34 Å. The amino acids directly involved in the catalytic reaction are shown in sticks where carbon atoms are wheat (BaBgl3) and yellow (Bgl7226) while the substrate/product binding residues, white (BaBgl3, lower numbers) and green (Bgl7226, higher numbers).

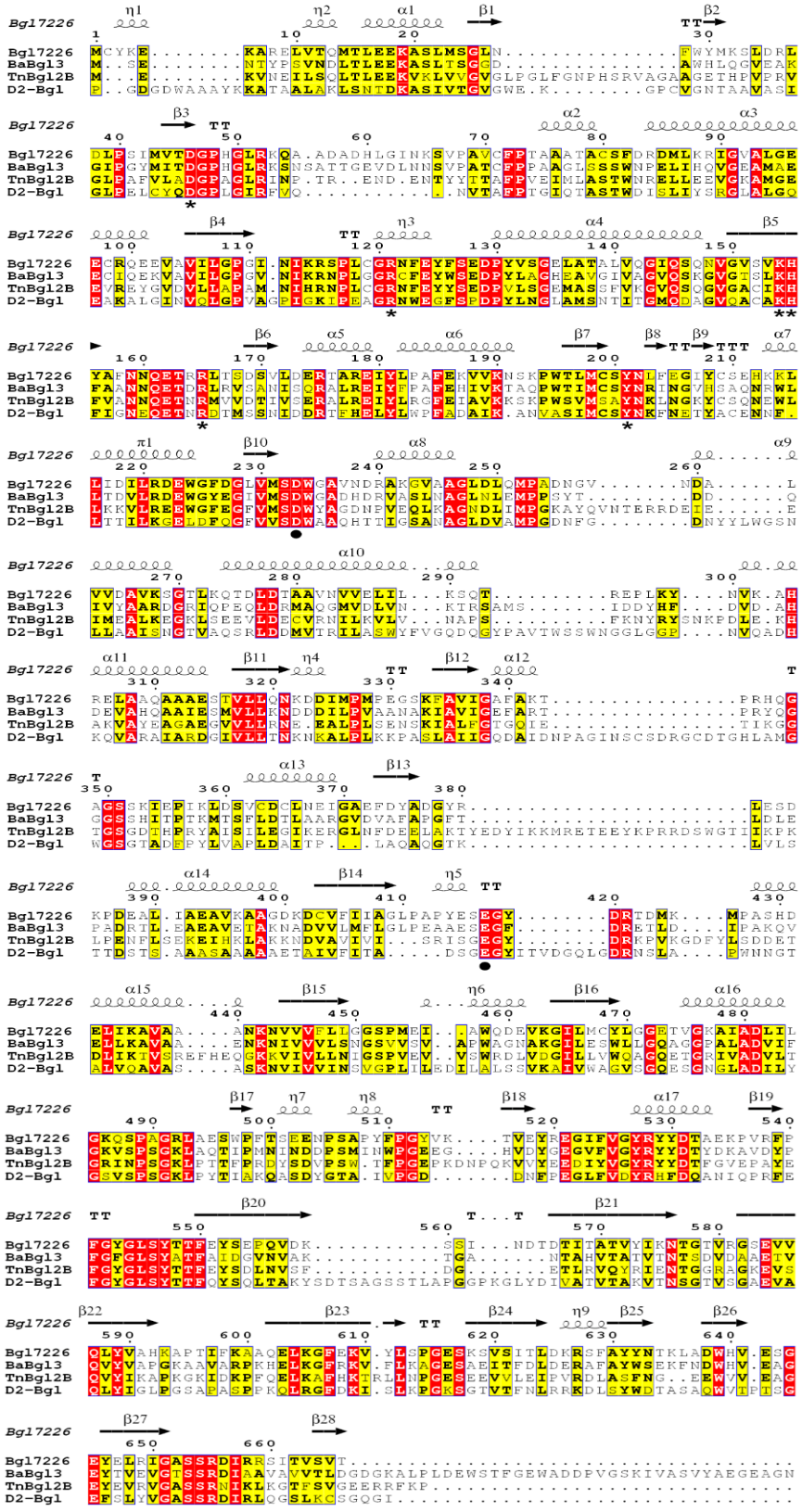


Fig7. Multiple sequence alignment of β-glucosidases from *S. wheeleri* (Bgl7226), *B. adolescentis* (BaBgl3, PDB entry: 5WAB), *T. neapolitana* (TnBgl3B PDB entry: 2X40) and *C. raphigera* (D2-BGL, PDB entry: 6JXG). Solid circles below residues indicate the catalytic pair. Asterisks below residues indicate the amino acids involved in substrate binding. Squiggles indicate the α-helices, and arrows indicate β-sheets.

5.0 Discussion

A previous study from our group described the gut microbial community of an abundant Brazilian Cerrado termite, *Syntermes wheeleri* [29]. Termites are considered one of the main lignocellulosic material degraders of the ecosystem, and play a crucial role in the formation of terrestrial ecosystems. Due to their high capacity to degrade cellulose, resulting from the action of symbiotic microorganisms, they are able to obtain the energy necessary to survive. This metabolic collaboration established between termites and symbionts makes the digestive tract of these insects an efficient bioreactor that produces cellulases, hemicellulases, β-glucosidases, and auxiliary enzymes, capable of releasing sugars from the main polysaccharides of lignocellulose, cellulose and hemicellulose. [49]. Given the ability termites have to degrade lignocellulosic materials*, S. wheeleri* was regarded as a promising source of β-glucosidases with potential for industrial application.

After edition, the remaining 49 β-glucosidases sequences from the *S. wheeleri* gut metagenome were classified as glycosyl hydrolase 3 (GH3). This is one of the largest families of carbohydrate-active enzymes, and includes members with distinct enzymatic activities, such as β-D-glucosidase (EC 3.2.1.21), β-D-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55) and N-acetyl-β-D-glucosaminidase (EC 3.2.1.52) [50]. GH3 enzymes play a key role in maintenance functions of organisms, as well as in their interactions with the environment, including pathogen-host interactions. Initially, they were described as bacterial and fungal β-glucosidases involved in the biodegradation of plant cellulose [51-54]. In termites,

GH3 enzymes were related to the final metabolism of oligosaccharide conversion into simple sugars [24].

Taxonomic analysis of sequences showed that the prevalent phyla in the *S. wheeleri* gut β-glucosidases were Firmicutes and Spirochaetes. A recent study that compared the gut microbiome of four species representing higher and lower termites, demonstrated the prevalence of Firmicutes and Spirochaetes, whose abundance can change depending on phylogeny and termite diet [55]. *S. wheeleri* gut bacterial community presents a higher percentage of Firmicutes (76.65%) followed by Spirochaetes (4.97%), Bacteroidetes (3.67%), Acidobacteria (2.56%), and Proteobacteria (2.28%) [29]. Although Spirochaetes are the second most abundant bacterial phylum in the gut of *S. wheeleri*, abundance is low in this species compared to that in other higher termites. This difference may be explained by the fact that *S. wheeleri* does not feed on wood, but litter [16, 19].

The Bgl7226 sequence, selected for biochemical and biophysical characterization, was classified according to bioinformatics analyses as belonging to the Spirochaetes phylum presenting a 53.0% protein identity with a β-glucosidase from *Treponema azotonnutricium* genome. *T. azotonnutricium* was first described in termites as being associated with the nitrogen fixation process and hydrogen production, which is a product of the fermentation of lignocellulose polysaccharides [56, 57]. In addition, metagenomic studies suggest they are responsible for encoding putative CAZymes in the hindgut of *Nasutitermes* species, and other works have also shown that the prevalent *Treponema* lineages in the higher wood-feeding termite gut are highly diverse and mostly as-yet uncultured [58,59]. Bgl7226

being annotated as a protein from this genus, suggests the participation of *Treponema sp*. in the final metabolism of lignocellulose in ''soil feeding'' termites, and the participation of a GH3 family member in this process [24].

Bgl7226 was produced under low temperature and low inducer concentration as an attempt to avoid the formation of inclusion bodies [60]. Nevertheless, Bgl7226 was still present in inclusion bodies, which accumulate in the cytoplasm or in the periplasmic space if the protein has a signal peptide [61, 62]. To bypass this problem, Bgl7226 was solubilized using 1 M urea, a mild denaturing condition to prevent large structural changes [60]. During the Bgl7226 purification process urea was removed and protein structure was reestablished.

After Bgl7226 purification, the effects of pH and temperature, two fundamental factors that affect enzymatic activity, were analyzed [63, 64]. Bgl7226 presented two optimal pHs for its activity, pH 7.0 and pH 10.0, which are similar to the high termite gut pH (i.e., ranges between 6.0 and 10.5) [65]. This result may suggest that, during the enzyme renaturation-purification process, at least two different populations of enzymes with respect to conformation and the catalytically active center could be obtained. There are other published studies which reported other β-glucosidases with two optimal pHs. Two lower termite β-glucosidases (CfGlu1B and CfGlu1C) from *Cototermes formosanus* showed optimal activity at pH 5 and CfGlu1C still retained 50% of its maximum activity in pH 8 [66], an alkaline pH similar to Bgl7226.

Experiments to determine the influence of pH on Bgl7226 secondary structure showed that this protein is structured in all pHs, with only small changes in

the percentages of alpha-helix and beta-sheet structures, which characterizes the secondary structure of this protein as pH independent at 25.0 °C. At pH 7.0, 24.4% of Bgl7226 was in alpha-helix structure and 20.0% in beta-sheet structure, the values for pH 10.0 are 27.3% and 15.6%, respectively. This result indicates that there is only a 3-5% variation of these structures in these two pHs, and that in pH 10.0 the protein is more structured in alpha-helix, so it seems that Bgl7226 can transit between two conformations that occur at pH 7.0 and pH 10.0. The percentage of secondary structure obtained for Bgl7226 agrees with those for other GH3 β-glucosidases reported in the literature. For example, 5WAB and 5K6O deposited in PDB (Protein Data Bank), present 27% and 25% alpha-helix structure and 18% and 19% beta-sheet structure, respectively [67, 68].

Bgl7226 presented different structural variations at different temperatures, varying their thermostability in different pHs. The results obtained for pHs 7.0 and 10.0 show that Bgl7226 is more stable in the basic range when compared to the neutral range. Altogether, these results on thermostability show that Bgl7226 is more stable at acidic pH, and, interestingly, that the small conformational changes that occur close to 40.0 °C (Fig. 5B, 5D and 5E), characterized as optimum temperature, seem to be very important for the highest activity of Bgl7226 observed at pH 7.0 and 10.0 (Fig. 3). Indeed, the mentioned small transition in the unfolding curves close to 40.0 °C, mainly at pH7.0, is typical of intermediates normally found in proteins structured in several domains, as predicted for Bgl7226.

The influence of temperature on Bgl7226 enzymatic activity was evaluated for the two optimal pHs, and results showed that in both pHs, the higher activity

occurred approximately at 40 °C. Mesophilic β-glucosidases, such as Bgl7226, may present high activity at temperatures between 30-65 °C, but they generally become inactivated above 55 °C [66]. Bgl7226 enzymatic activity decreased dramatically starting at approximately 40 °C at pH 10.0, and from approximately 45 °C at pH 7.0 (Fig. 3) as a result of protein denaturation. At 50 °C, Bgl7226 only shows 15% activity in both pHs (Fig. 7). Recent reports revealed a similar optimum temperature of 45 ºC for another GH3 β-glucosidase from the probiotic bacterium *Bifidobacterium adolescents*, and of 40 °C for a GH1 β-glucosidase from the cow rumen bacterium *Cellulosilyticum ruminicola* [67, 69]. Studies on β-glucosidases from higher termites have reported approximate optimum temperatures of 50 ºC and 40 ºC, for the enzyme from *Neotermes koshunensis* and *Coptotermes formosanus,* respectively [70, 65].

Kinetic parameters were determined using the Michaelis-Menten curve. K_m indicates that the enzyme has 9-fold more affinity for pNPG in the neutral than basic pH. When comparing the Bgl7226 K_m's with those of other termite β-glucosidases, which presented K_m of 1.7 mM, 2.2 mM and 1.47 mM of pNPG [70, 65], it can be concluded that Bgl7226 presented higher affinity for the substrate in both pHs. The Bgl7226 k_{cat}'s indicate that the enzyme in pH 10.0 is catalytically 1.13-fold faster with pNPG substrate compared to pH 7.0. Regarding the constant that gives the kinetic specificity or efficiency of the enzyme (k_{cat}/K_m), Bgl7226 at pH 7.0 is 8-fold more catalytically efficient than at pH 10.0. Furthermore, these constants for Bgl7226 are higher than those of other termite endogenous β-glucosidases (0.002

and 0.006 s^{-1}. μM^{-1}) [65, 71], suggesting an important participation of bacteria in the lignocellulose metabolism in termites.

Given the conserved structure and sequence of the Bgl7226, it is assumed that domains 1 and 2 have a direct role in the architecture of the active site. Even though the function of domain 3 (FnIII) remains unknown, it has been associated with different enzyme activities such as chitinase, cellobiohydrolase, α-amylase, and pullulanase [72, 73]. Additionally, the presence of the FnIII domain in different types of enzymes supports the hypothesis that it may have less specific roles such as to act as a stable and/or flexible linker [73]. This is in agreement with the molecular dynamics simulation of BaBgl3 that showed domain 3 as being highly flexible in comparison with the catalytic domains [67]. Residue D233 is highly conserved throughout the GH3 family, but E417 is not well conserved. Indeed, the GH3 BsNagZ from *Bacillus subtilis* has a histidine at this position [74]. Bgl7226 has a shorter loop close to the active site entry region than the corresponding loop in TnBgl3B from *Thermotoga neapolitana* (10 amino acids from residues 377 to 387 vs 25 amino acids from residue 409 to 434) [75]. The same feature was observed for BaBgl3 and D2-BGL from *Bifidobacterium adolescentis* and *Chaetomella raphigera*, respectively [67, 76]. Other authors have suggested that loop size [72] leads to differences in the accommodation of the substrate resulting in changes in K_m and in the efficiency of enzyme catalysis.

6.0 Conclusion

To summarize, this work reports on the biochemical and biophysical characteristics of Bgl7226. This is the first bacterial GH3 β-glucosidase from the gut metagenome of an abundant and large Brazilian Cerrado "higher termite", *Syntermes wheeleri,* to be characterized in such detail. This enzyme most likely plays a role in ingested plant biomass degradation in the termite gut. We observed that the bacterium *Treponema azotonnutricium* (Spirochaetes) could be participating in this process in the hindgut of *S. wheeleri*, as already observed in other termites. GH3 family enzymes can have different activities. pNPG assays suggest that Bgl7226 acts as a β-glucosidase as predicted by bioinformatics analysis. We identified a GH3 enzyme that may be associated with the final process of converting oligosaccharides to simple sugars in "higher termites", as reported in other works. Nevertheless, in future studies Bgl7226 activity with different natural substrates will be explored. Bgl7226 has interesting biochemical properties, such as two optimum pHs, higher affinity (K_m) and efficiency (k_{cat}/K_m) when compared to other endogenous termite β-glucosidases highlighting the importance of prokaryotes in lignocellulose degradation in "higher termites". Furthermore, both thermostability assays and thermal structural changes analysis demonstrated that the optimal enzymatic activity was at 40 °C. Bgl7226 was structurally similar to other GH3 proteins in PDB regarding secondary structure and conserved domains. The omics approach and enzyme characterization were important to identify Bgl7226 and to better understand the characteristics of this GH3 "higher termite" gut enzyme. Since ß-glucosidases have diverse biotechnological applications in industries such as biofuel production, food flavor enhancement, and detoxification plant roots for

human an animal consumption, further studies will be needed to determine the ideal industrial application for this novel enzyme. Specifically, our group would be interested in cassava roots (*Manihot esculenta)* detoxification for animal feed.

7.0 References

1. B.K. Singh, Exploring microbial diversity for biotechnology: the way forward, Trends Biotechnol. 28 (2010) 111-116. https://doi.org/10.1016/j.tibtech.2009.11.006
2. M. Sharma, N. Gautam, Impact of microbial Divesity on Environmental Satability, in: V. Parmar, P. Malhotra, D. Mathur (Eds.), Green Chemistry in Environmental Sustainability and Chemical Education, Springer., Singapore, 2018, pp. 81-91.
3. V. Kumar, A. Kumar, D. Chhabra, P. Shukla, Improved biobleaching of mixed hardwood pulp and process optimization using novel GA-ANN and GA-ANFIS hybrid statistical tools, Bioresour Technol. 271 (2019) 274-282. https://doi.org/10.1016/j.biortech.2018.09.115.
4. N.V. Ravin, A.V. Mardanova, K.G. Skryabin, Metagenomics as a Tool for the Investigation of Uncultured Microorganisms, Genetika. 51 (2015) 519–528. https://doi.org/10.1134/S1022795415050063
5. S. Kumar, A.K. Dangi, P. Shukla, D. Baishya, S.K. Khare, Thermozymes: Adaptive strategies and tools for their biotechnological applications, Bioresour Technol. 278 (2019) 372-382. https://doi.org/10.1016/j.biortech.2019.01.088.

6. T. Nimchua, T. Thongaram, T. Uengwetwanit, S. Pongpattanakitshote, L. Eurwilaichitr, Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes, J Microbiol Biotechnol. 22 (2012) 462–469. https://doi.org/10.4014/jmb.1108.08037.
7. D. Cowan, Q. Meyer, W. Stafford, S. Muyanga, R. Cameron, P. Wittwer, Metagenomic gene discovery: past, present and future, Trends Biotechnol. 23 (2005) 321–329. https://doi.org/10.1016/j.tibtech.2005.04.001.
8. Y.P. Zhang, Reviving the carbohydrate economy via multi-product lignocellulose biorefineries, J Ind Microbiol Biotechnol. 35 (2008) 367–375. https://doi.org/10.1007/s10295-007-0293-6.
9. W.K. Cornwell, J.H.C. Cornelissen, S.D. Allison, J. Bauhus, P. Eggleton, C.M. Preston, F. Scarff, J.T. Weedon, C. Wirth, A.E. Zanne**,** Plant traits and wood fates across the globe: rotted, burned, or consumed?, Global Change Biol. 15 (2009) 2431-2449. https://doi.org/10.1111/j.1365-2486.2009.01916.x.
10. K. Kucharska, P. Rybarczyk, I. Hołowacz, R. Łukajtis, M. Glinka, M. Kamiński, Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes, Molecules. 23 (2018) 2937. https://doi.org/10.3390/molecules23112937.
11. L. Huang, H. Zhang, P. Wu, S. Entwistle, X. Li, T. Yohe, H. Yi, Z. Yang, Y. Yin, dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation, Nucleic Acids Res. 46 (2018) 516–521. https://doi.org/10.1093/nar/gkx894.

12. B. Henrissat, A. Bairoch, Updating the sequence-based classification of glycosyl hydrolases, Biochem J. 316 (1996) 695–696. https://doi.org/10.1042/bj3160695.

13. 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) 233–238. https://doi.org/10.1093/nar/gkn663.

14. Carbohydrate Active Enzymes database, Glycoside Hydrolase Family classification. http://www.cazy.org/Glycoside-Hydrolases.html, 2020 (accessed 01 July 2020).

15. B.C. Behera, B.K. Sethi, R.R. Mishra, S.K. Dutta, H.N. Thatoi, Microbial cellulases - Diversity & biotechnology with reference to mangrove environment: A review, J Genet Eng Biotechnol. 15 (2017) 197–210. https://doi.org/10.1016/j.jgeb.2016.12.001.

16. Y. Bhatia, S. Mishra, V.S. Bisaria. 2002. Microbial beta-glucosidases: cloning, properties, and applications, Crit Rev Biotechnol. 22 (2002) 375–407. https://doi.org/10.1080/07388550290789568.

17. R.R. Singhania, A.K. Patel, R.K. Sukumaran, C. Larroche, A. Pandey. 2013. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production, Bioresour Technol. 127 (2013) 500–507. https://doi.org/10.1016/j.biortech.2012.09.012.

18. A. Brune, Symbiotic digestion of lignocellulose in termite guts, Nat Rev Microbiol. 12 (2014), 168–180. https://doi.org/10.1038/nrmicro3182.

19. A. Mikaelyan, J.F.H. Strassert, G. Tokuda, A. Brune, The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes spp*.), Environ Microbiol. 16 (2014) 2711–2722. https://doi.org/10.1073/pnas.1810550115.

20. G. Tokuda, Y. Tsuboi, K. Kihara, S. Saitou, S. Moriya,N. Lo, J. Kikuchi, Metabolomic profiling of 13C-labelled cellulose digestion in a lower termite: insights into gut symbiont function, Proc Biol Sci. 281 (2014) 20140990. https://doi.org/10.1098/rspb.2014.0990.

21. R. Constantino, 1995. Revision of the Neotropical Termite Genus *Syntermes* Holmgren (Isoptera: Termitidae), Kans. Univ. sci. bull. 55 (1995) 455-518.

22. 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. 2 (2009) 25. https://doi.org/10.1186/1754-6834-2-25.

23. G. Tokuda, Plant cell wall degradation in insects: Recent progress on endogenous enzymes revealed by multi-omics technologies, In: Advances in Insect Physiology, Academic Press., Cambridge, 2019, pp. 97-136.

24. F. Warnecke, P. Luginbühl, N. Ivanova, M. Ghassemian, T.H. Richardson, J.T. Stege, M. Cayouette, A.C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H.G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, … J.R. Leadbetter, Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450 (2007) 560–565. https://doi.org/10.1038/nature06269.

25. M. Calusinska, M. Marynowska, M. Bertucci, B. Untereiner, D. Klimek, X. Goux, D. Sillam-Dussès, P. Gawron, R. Halder, P. Wilmes, P. Ferrer, P. Gerin, Y. Roisin, P. Delfosse, Integrative omics analysis of the termite gut system adaptation to *Miscanthus* diet identifies lignocellulose degradation enzymes, Commun Biol. 3 (2020) 275. https://doi.org/10.1038/s42003-020-1004-3.

26. N. Liu, H. Li, M.G. Chevrette, L. Zhang, L. Cao, H. Zhou, X. Zhou, Z. Zhou, P.B. Pope, C.R. Currie, Y. Huang, Q. Wang, Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite, ISME J. 13 (2019) 104-117. https://doi.org/10.1038/s41396-018-0255-1.

27. M. Marynowska, X. Goux, D. Sillam-Dussès, C. Rouland-Lefèvre, R. Halder, P. Wilmes, P. Gawron, Y. Roisin, P. Delfosse, M. Calusinska, Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites, Microbiome. 8 (2020) 96. https://doi.org/10.1186/s40168-020-00872-3.

28. G. Tokuda , A. Mikaelyan, C. Fukui, Y. Matsuura, H. Watanabe, M. Fujishima, A. Brune, Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites, Proc Natl Acad Sci. 115 (2018) 11996-12004. https://doi.org/10.1073/pnas.1810550115.

29. R.H. Santana, E.C. Catão, F.A. Lopes, R. Constantino, C.C. Barreto, R.H. Krüger, The Gut Microbiota of Workers of the Litter-Feeding Termite *Syntermes wheeleri* (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities, Microb Ecol. 70 (2015) 545–556. https://doi.org/10.1007/s00248-015-0581-z.

30. T. Nimchua, T. Thongaram, T. Uengwetwanit, S. Pongpattanakitshote, L. Eurwilaichitr, Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes, J Microbiol Biotechnol. 22 (2012) 462–469. https://doi.org/10.4014/jmb.1108.08037. U.K.
31. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature. 227 (1970) 680–685. https://doi.org/10.1038/227680a0.
32. B.R. Oakley, D.R. Kirsch, N.R. Morris, A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels, Anal Biochem. 105 (1980) 361–363. https://doi.org/10.1016/0003-2697(80)90470-4.
33. W.A. Wood, T.K. Scott, Biomass. Part A. Cellulose and hemicellulose, first ed., Academic Press, 1988.
34. O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent, J Biol Chem. 193 (1951) 265–275.
35. J. Labuda, R.P. Bowater, M. Fojta, G. Gauglitz, Z. Glatz, I. Hapala, J. Havliš, F. Kilar, A. Kilar, L. Malinovská, H.M.M. Sirén, P. Skládal, F. Torta, M. Valachovič, M. Wimmerová, Z. Zdráhal, D.B. Hibbert, Terminology of bioanalytical methods (IUPAC Recommendations 2018), *Pure and Applied Chemistry*, *90* (2018), 1121-1198. https://doi.org/10.1515/pac-2016-1120.
36. C.N. Pace, J.M. Scholtz, Measuring the conformational stability of a protein, Protein structure: A practical approach. 2 (1997) 299-321.
37. G. Böhm, CDNN: CD spectra deconvolution software version 2.1. University of Halle-Wittenberg, Halle, Germany, 1997.

38. S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST. A new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

39. B. Webb, A. Sali, Protein structure modeling with MODELLER, Methods Mol Biol. 1137 (2014) 1–15. https://doi.org/10.1007/978-1-4939-0366-5_1.

40. W.L. DeLano, The PyMOL Molecular Graphics System. https://pymol.org/2/.

41. D. Bhattacharya, J. Cheng, 3Drefine: consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization, Proteins. 81 (2013) 119-131. https://doi.org/10.1002/prot.24167.

42. E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem. 25 (2004) 1605-12. https://doi.org/10.1002/jcc.20084.

43. S.C. Lovell, I.W. Davis, W.B. 3rd. Arendall, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: phi,psi and Cbeta deviation, Proteins. 50 (2003) 437-50. https://doi.org/10.1002/prot.10286.

44. P. Benkert, M. Künzli, T. Schwede, QMEAN server for protein model quality estimation, Nucleic Acids Res. 37 (2009) 510-514. https://doi.org/10.1093/nar/gkp322.

45. D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396-404. https://doi.org/10.1016/s0076-6879(97)77022-8.

46. C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: a novel method for fast and accurate multiple sequence alignment, J.Mol.Biol. 302 (2000) 205–217. https://doi.org/10.1006/jmbi.2000.4042.

47. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) 320-324. https://doi.org/10.1093/nar/gku316.

48. D.M. Hillis, J.J. Bull, An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis, *Systematic Biology*. 42 (1993) 182–192. https://doi.org/10.1093/sysbio/42.2.182.

49. J. Zhou, J. Duan, M. Gao, Y. Wang, X. Wang, K. Zhao, Diversity, Roles, and Biotechnological Applications of Symbiotic Microorganisms in the Gut of Termite, Curr Microbiol. 76 (2019) 755–761. https://doi.org/10.1007/s00284-018-1502-4.

50. D. Dodd, S. Kiyonari, R.I. Mackie, I.K. Cann, Functional diversity of four glycoside hydrolase family 3 enzymes from the rumen bacterium *Prevotella bryantii* B14, J Bacteriol. 192 (2010) 2335–2345. https://doi.org/10.1128/JB.01654-09.

51. D. Faure, The family-3 glycoside hydrolases: From housekeeping functions to host-microbe interactions, Appl. Environ. Microbiol. 68 (2002) 1485–1490. https://doi.org/10.1128/aem.68.4.1485-1490.2002.

52. H.T. Tsujibo, N. Hatano, T. Mikami, A. Hirasawa, K. Miyamoto, Y. Inamori, A novel β-N-acetylglucosaminidase from *Streptomyces thermoviolaceus* OPC-520:gene cloning, expression and assigment to family 3 of the glycosyl hydrolases, Appl. Environ. Microbiol. 64 (1998) 2920–2924. https://doi.org/10.1128/AEM.64.8.2920-2924.1998.

53. N.N. van Peij, J. Brinkmann, M. Vrsanska, J. Visser, L.H. de Graaff, β-Xylosidase activity, encoded by *xln*D, is essential for complete hydrolysis of xylan by *Aspergillus niger* but not for induction of the xylanolytic enzyme spectrum, Eur. J. Biochem. 245 (1997) 164–173. https://doi.org/10.1111/j.1432-1033.1997.00164.x.

54. B. Cournoyer, D. Faure, Radiation and functional specialization of the family-3 glycoside hydrolases, J Mol Microbiol Biotechnol. 5 (2003) 190–198. https://doi.org/10.1159/000070269.

55. L. Su, L. Yang, S. Huang, X. Su, Y. Li, F. Wang, E. Wang, N. Kang, J. Xu, A. Song, Comparative Gut Microbiomes of Four Species Representing the Higher and the Lower Termites, Journal of insect science. 16 (2016) 97. https://doi.org/10.1093/jisesa/iew081.

56. J.R. Graber, J.R. Leadbetter, J.A. Breznak, Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts, Appl Environ Microbiol. 70 (2004) 1315–1320. https://doi.org/10.1128/aem.70.3.1315-1320.2004.

57. A.Z. Rosenthal, E.G. Matson, A. Eldar, J.R. Leadbetter, RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete

species in co-culture, ISME J. 5 (2011) 1133–1142. https://doi.org/10.1038/ismej.2011.3.

58. S. He, N. Ivanova, E. Kirton, M. Allgaier, C. Bergin, R.H. Scheffrahn, N.C. Kyrpides, F. Warnecke, S.G. Tringe, P. Hugenholtz. Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites, PLoS One. 8 (2013) 61126. https://doi.org/10.1371/journal.pone.0061126.

59. M. Yuki , H. Kuwahara, M. Shintani, K. Izawa, T. Sato, D. Starns, Y. Hongoh, M. Ohkuma, Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose, Environ Microbiol. 17 (2015) 4942-4953. doi: 10.1111/1462-2920.12945. https://doi.org/10.1111/1462-2920.12945.

60. S.M. Singh, A.K. Panda, Solubilization and refolding of bacterial inclusion body proteins, J Biosci Bioeng. 99 (2005) 303–310. https://doi.org/10.1263/jbb.99.303.

61. J.P. Arié, M. Miot, N. Sassoon, J.M. Betton , Formation of active inclusion bodies in the periplasm of *Escherichia coli*, Mol Microbiol. 62 (2006) 427–437. https://doi.org/10.1111/j.1365-2958.2006.05394.x.

62. A.S. Coquel, J.P. Jacob, M. Primet, A. Demarez, M. Dimiccoli, T. Julou, L. Moisan, A.B. Lindner, H. Berry, Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect, PLoS computational biology. 9 (2013) e1003038. https://doi.org/10.1371/journal.pcbi.1003038.

63. K. Talley, E. Alexov, On the pH-optimum of activity and stability of proteins, Proteins. 78 (2010) 2699–2706. https://doi.org/10.1002/prot.22786.

64. G.N. Somero, Proteins and temperature, Annu Rev Physiol. 57 (1995) 43–68. https://doi.org/10.1146/annurev.ph.57.030195.000355.

65. T. Thongaram, Y. Hongoh, S. Kosono, M. Ohkuma, S. Trakulnaleamsai, N. Noparatnaraporn, T. Kudo, Comparison of bacterial communities in the alkaline gut segment among various species of higher termites, Extremophiles: life under extreme conditions. 9 (2005) 229–238. https://doi.org/10.1007/s00792-005-0440-9.

66. T. Feng, H. Liu, Q. Xu, J. Sun, H. Shi, Identification and Characterization of Two Endogenous β-Glucosidases from the Termite *Coptotermes formosanus*, Appl Biochem Biotechnol. 176 (2015) 2039–2052. https://doi.org/10.1007/s12010-015-1699-7.

67. R.N. Florindo, V.P. Souza, L.R. Manzine, C.M. Camilo, S.R. Marana, I. Polikarpov, A.S. Nascimento, Structural and biochemical characterization of a GH3 β-glucosidase from the probiotic bacteria *Bifidobacterium adolescentis*, Biochimie. 48 (2018) 107–115. https://doi.org/10.1016/j.biochi.2018.03.007.

68. Ramírez-Escudero, M., Del Pozo, M. V., Marín-Navarro, J., González, B., Golyshin, P. N., Polaina, J., Ferrer, M., & Sanz-Aparicio, J. (2016). Structural and Functional Characterization of a Ruminal β-Glycosidase Defines a Novel Subfamily of Glycoside Hydrolase Family 3 with Permuted Domain

Topology. *The Journal of biological chemistry*, *291*(46), 24200–24214. https://doi.org/10.1074/jbc.M116.747527.

69. E. Suwan, S. Arthornthurasuk, P.T. Kongsaeree, A metagenomic approach to discover a novel β -glucosidase from bovine rumens, Pure Appl Chem. 89 (2017) 941-950. https://doi.org/10.1515/pac-2016-0924.

70. C.A. Uchima, G. Tokuda, H. Watanabe, K. Kitamoto, M. Arioka, Heterologous expression and characterization of a glucose-stimulated β-glucosidase from the termite *Neotermes koshunensis* in *Aspergillus oryzae*, Appl Microbiol Biotechnol. 89 (2011) 1761–1771. https://doi.org/10.1007/s00253-010-2963-y.

71. S. Arthornthurasuk, W. Jenkhetkan, E. Suwan, D. Chokchaichamnankit, C. Srisomsap, P. Wattana-Amorn, J. Svasti, P.T. Kongsaeree, Molecular Characterization and Potential Synthetic Applications of GH1 β-Glucosidase from Higher Termite *Microcerotermes annandalei*, Applied biochemistry and biotechnology. 186 (2018) 877–894. https://doi.org/10.1007/s12010-018-2781-8.

72. I.A. Kataeva, R.D. 3rd. Seidel, A. Shah, L.T. West, X.L. Li, L.G. Ljungdahl, The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface, Appl Environ Microbiol. 68 (2002) 4292-300. https://doi.org/10.1128/aem.68.9.4292-4300.2002.

73. V. Valk, R.M. Kaaij, L. Dijkhuizen, The evolutionary origin and possible functional roles of FNIII domains in two *Microbacterium aurum* B8. A granular

starch degrading enzymes, and in other carbohydrate acting enzymes, Amylase. 1 (2017) 1–11. https://doi.org/10.1515/amylase-2017-0001.

74. S. Litzinger, S. Fischer, P. Polzer, K. Diederichs, W. Welte, C. Mayer, Structural and kinetic analysis of *Bacillus subtilis* N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism, J Biol Chem. 285 (2010) 35675-84. https://doi.org/10.1074/jbc.M110.131037.

75. T. Pozzo, J.L. Pasten, E.N. Karlsson, D.T. Logan, Structural and Functional Analyses of β-Glucosidase 3B from *Thermotoga neapolitana*: A Thermostable Three-Domain Representative of Glycoside Hydrolase 3, J. Mol. Biol. 397 (2010) 724–739. https://doi.org/10.1016/j.jmb.2010.01.072.

76. M.R. Kao, H.W. Kuo, C.C. Lee, K.Y. Huang, T.Y. Huang, C.W. Li, C.W. Chen, A.H. Wang, S.M. Yu, T.D. Ho, *Chaetomella raphigera* β-glucosidase D2-BGL has intriguing structural features and a high substrate affinity that renders it an efficient cellulase supplement for lignocellulosic biomass hydrolysis. Biotechnol Biofuels 12 (2019) 258. https://doi.org/10.1186/s13068-019-1599-0.

8.0 Appendix

Table S1. *Syntermes wheeleri* gut metagenomic β-glucosidases after screening in Geneious software 10.2. P11, P12, P13, P31, P32, P33, WG1, WG2, WG3 (P1: first proctodeal segment, P3: third proctodeal segment and WG: whole gut) are samples from different sections of the *Syntermes wheeleri* gut in 3 biological replicates. The transcriptome coverage is present on the last column.

Sequence annotation	GH Family	Protein Identity (%)	Phylum	Organism	Genome completeness	Metatranscriptome coverage
P11_assembly_scaffold_112_27	GH3	63.3	?	unclassified sequences	near complete	x
P11_assembly_scaffold_122_17	GH3	57.3	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,044229556
P11_assembly_scaffold_122_3	GH3	40.1	Actinobacteria	*Bifidobacterium actinocoloniiforme*	partial	0,046429765
P11_assembly_scaffold_1357_5	GH3	66.4	?	unclassified sequences	megabin	x
P11_assembly_scaffold_292_4	GH3	47.5	Spirochaetes	*Treponema azotonutricium*	megabin	0,011825911
P11_assembly_scaffold_598_3	GH3	66.0	?	unclassified sequences	megabin	x
P11_assembly_scaffold_604_10	GH3	52.8	Spirochaetes	*Treponema azotonutricium*	partial	0,030298608
P11_assembly_scaffold_734_7	GH3	40.6	Firmicutes	*Clostridium cellulovorans*	megabin	0,019956384
P12_assembly_scaffold_132_19	GH3	66.5	?	unclassified sequences	partial	x
P12_assembly_scaffold_2515_1	GH3	39.5	Firmicutes	*Paenibacillus barengoltzii*	megabin	0,001001497
P12_assembly_scaffold_2566_2	GH3	57.3	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,012383254
P12_assembly_scaffold_3_56	GH3	42.5	Firmicutes	*Clostridium cellulovorans*	near complete	0,0665746
P12_assembly_scaffold_3566_1	GH3	56.9	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,002791969
P12_assembly_scaffold_383_2	GH3	40.9	Firmicutes	*Clostridium cellulovorans*	megabin	0,011936956
P12_assembly_scaffold_6036_2	GH3	45.9	Firmicutes	BJP_IG2157_Clostridiales_52_13	megabin	0,003669033
P13_assembly_scaffold_1_10	GH3	53.0	Spirochaetes	*Treponema azotonutricium*	megabin	0,015058046
P13_assembly_scaffold_196_19	GH3	50.2	Firmicutes	Lachnospiraceae bacterium CAG:215	megabin	0,002722825
P13_assembly_scaffold_2_12	GH3	42.5	Firmicutes	*Clostridium cellulovorans*	megabin	0,014836519
P13_assembly_scaffold_2806_2	GH3	44.8	Firmicutes	*Dorea longicatena* CAG:42	megabin	x
P13_assembly_scaffold_8004_1	GH3	57.6	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,011288004
P31_assembly_scaffold_1194_2	GH3	53.0	Spirochaetes	*Treponema azotonutricium*	megabin	x
P31_assembly_scaffold_125_3	GH3	40.3	Firmicutes	*Clostridium cellulovorans*	megabin	x
P31_assembly_scaffold_2_98	GH3	45.0	Firmicutes	Lachnospiraceae bacterium CAG:215	near complete	x
P31_assembly_scaffold_2333_1	GH3	49.4	Firmicutes	Firmicutes bacterium CAG:24	megabin	x
P31_assembly_scaffold_2642_2	GH3	52.7	Spirochaetes	*Treponema azotonutricium*	megabin	x
P31_assembly_scaffold_933_3	GH3	38.7	Firmicutes	*Paenibacillus barengoltzii*	megabin	x
P32_assembly_scaffold_102_18	GH3	52.5	Spirochaetes	*Treponema azotonutricium*	megabin	x
P32_assembly_scaffold_249_8	GH3	45.0	Firmicutes	Lachnospiraceae bacterium CAG:215	megabin	x
P32_assembly_scaffold_271_6	GH3	39.8	Firmicutes	*Ruminococcus callidus*	megabin	x
P32_assembly_scaffold_6_13	GH3	42.5	Firmicutes	*Clostridium cellulovorans*	megabin	x
P33_assembly_scaffold_24_2	GH3	37.6	Firmicutes	*Bacillus alcalophilus*	megabin	0,053120928
P33_assembly_scaffold_4_37	GH3	60.8	Firmicutes	*Pseudoflavonifractor capillosus*	near complete	0,019161303
P33_assembly_scaffold_72_26	GH3	53.0	Spirochaetes	*Treponema azotonutricium*	partial	0,611764523
WG1_assembly_scaffold_143_4	GH3	47.5	Spirochaetes	*Treponema azotonutricium*	megabin	x
WG1_assembly_scaffold_176_5	GH3	52.8	Spirochaetes	*Treponema azotonutricium*	megabin	0,016618001
WG1_assembly_scaffold_2_54	GH3	57.3	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,002627318
WG1_assembly_scaffold_2_68	GH3	40.1	Actinobacteria	*Bifidobacterium actinocoloniiforme*	megabin	0,028361543
WG1_assembly_scaffold_33_27	GH3	64.3	?	unclassified sequences	megabin	x
WG1_assembly_scaffold_347_3	GH3	66.0	?	unclassified sequences	megabin	x
WG1_assembly_scaffold_3542_3	GH3	54.7	Firmicutes	Firmicutes bacterium CAG:194	megabin	x
WG2_assembly_scaffold_1_64	GH3	42.5	Firmicutes	*Clostridium cellulovorans*	near complete	0,030974363
WG2_assembly_scaffold_1978_1	GH3	57.5	Firmicutes	Firmicutes bacterium CAG:424	megabin	0,00724238
WG2_assembly_scaffold_3123_1	GH3	71.4	?	unclassified sequences	megabin	x
WG2_assembly_scaffold_8166_1	GH3	43.1	Proteobacteria	*Agarivorans albus*	megabin	x
WG3_assembly_scaffold_2_33	GH3	42.5	Firmicutes	*Clostridium cellulovorans*	megabin	0,025886827
WG3_assembly_scaffold_317_5	GH3	60.8	Firmicutes	*Pseudoflavonifractor capillosus*	megabin	0,003357396
WG3_assembly_scaffold_42_42	GH3	53.0	Spirochaetes	*Treponema azotonutricium*	megabin	0,042351099
WG3_assembly_scaffold_90_8	GH3	53.0	Spirochaetes	*Treponema azotonutricium*	megabin	0,017060567

Yellow highlight: Informations of the β-glucosidase characterized in this work. (Bgl7226).

Table S2. Summary of Bgl7226 homology model and statistical validation.

Main template			RAMPAGE	Verify 3D	Normalized QMEAN4
PDB-entry	**Resolution (Å)**	**Sequence identity (%)**	**Ramachandran plot (%) [a]**	**Average 3D-1D score ≥0.2 (%) [b]**	**Z-score [c]**
5WAB	2.45	44.08	94.9, 4.7 and 1.4	91.9	0.73

[a]Numbers correspond to favored region, allowed region and outlier region, respectively. [b] 80% of the residues in a structure is required to be considered a good model. [c]Z-scores close to zero indicates good agreement between the model structure and high-resolution X-ray structures of similar size.

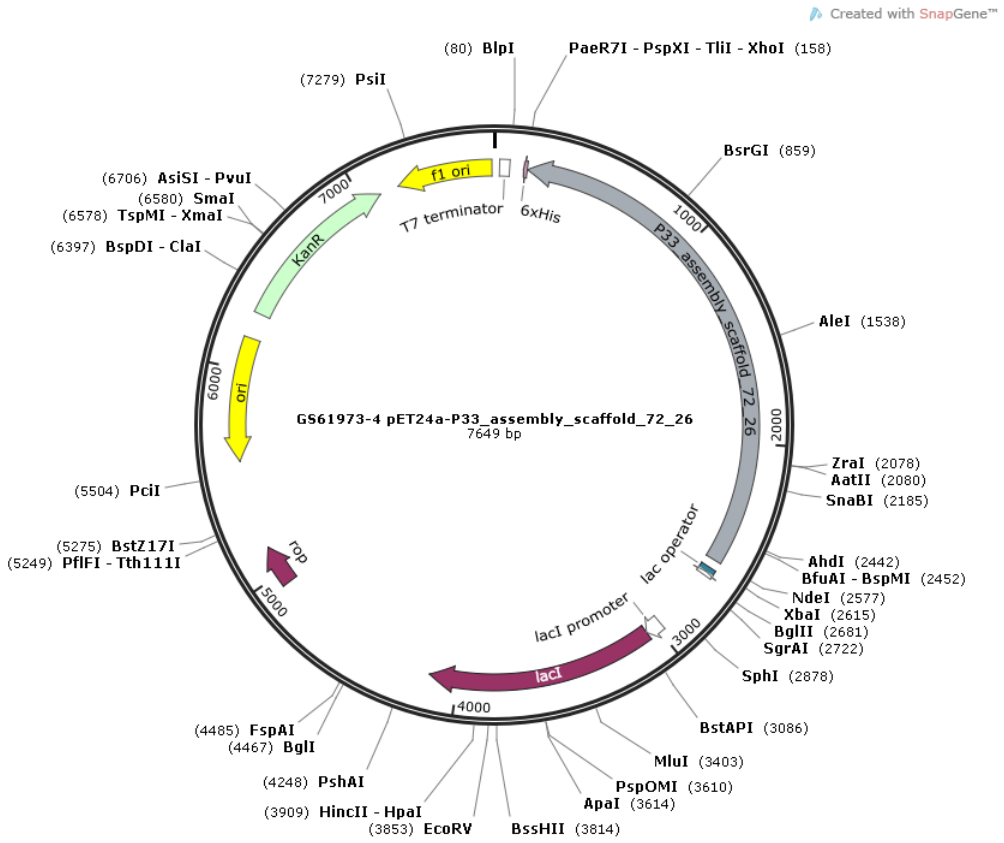


Fig S1. Expression plasmid with the gene for Bgl7226. The *Bgl7226* synthetic gene from the P33_assembly_Scaffold_72_26 was cloned into NdeI/XhoI digested pET24a(+). The gray arrow represents the *Bgl7226* gene under control of the T7*lac* promoter.

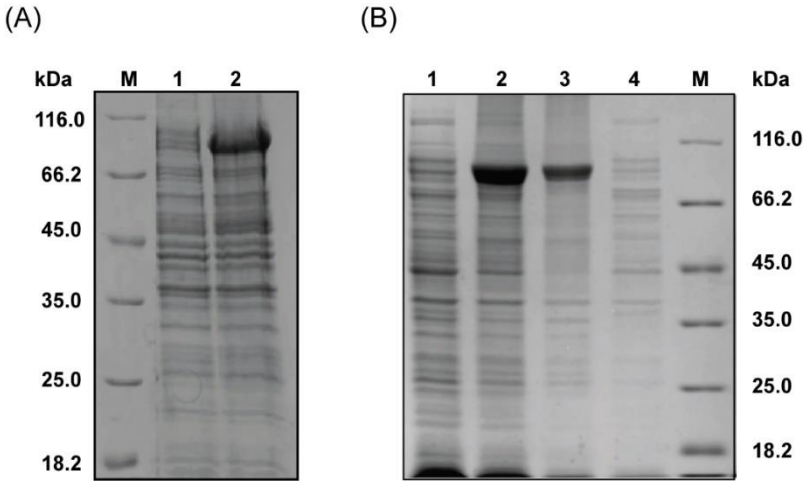
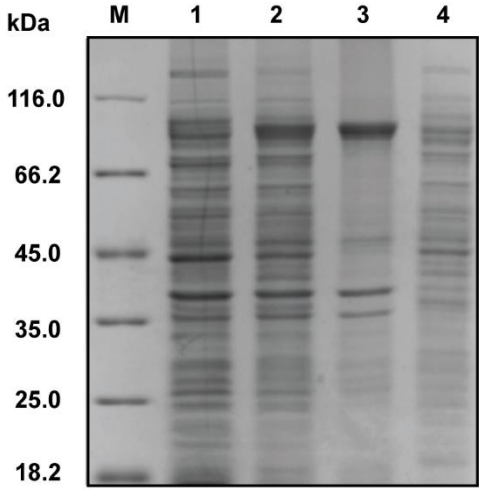


Fig S2. SDS-PAGE analysis to evaluate the expression and solubility of Bgl7226. (A) *Bgl7226* gene expression was induced with 1 mM IPTG at 37 °C for 3h, and 12% SDS-PAGE analusis reveled insoluble Bgl7226 protein (pellet). The approximate size of Bgl7226 is 90 kDa. M (Unstained Protein MW Marker, Thermo Fisher Scientific), 1 (inoculum) and 2 (induction). (B) Bgl7226 solubility test showed that the protein was being produced in inclusion bodies (insoluble). 1 (inoculum), 2 (induction), 3 (insoluble fraction, 4 (soluble fraction and M (Unstained Protein MW Marker, Thermo Fisher Scientific).



FigS3. **Optimization of Bgl7226 production.** Bgl7226 protein continued to be produced mostly in insoluble form (inclusion bodies). M (Unstained Protein MW Marker, Thermo Fisher Scientific), 1 (pre-inoculum), 2 (induction), 3 (insoluble part) and 4 (soluble part).

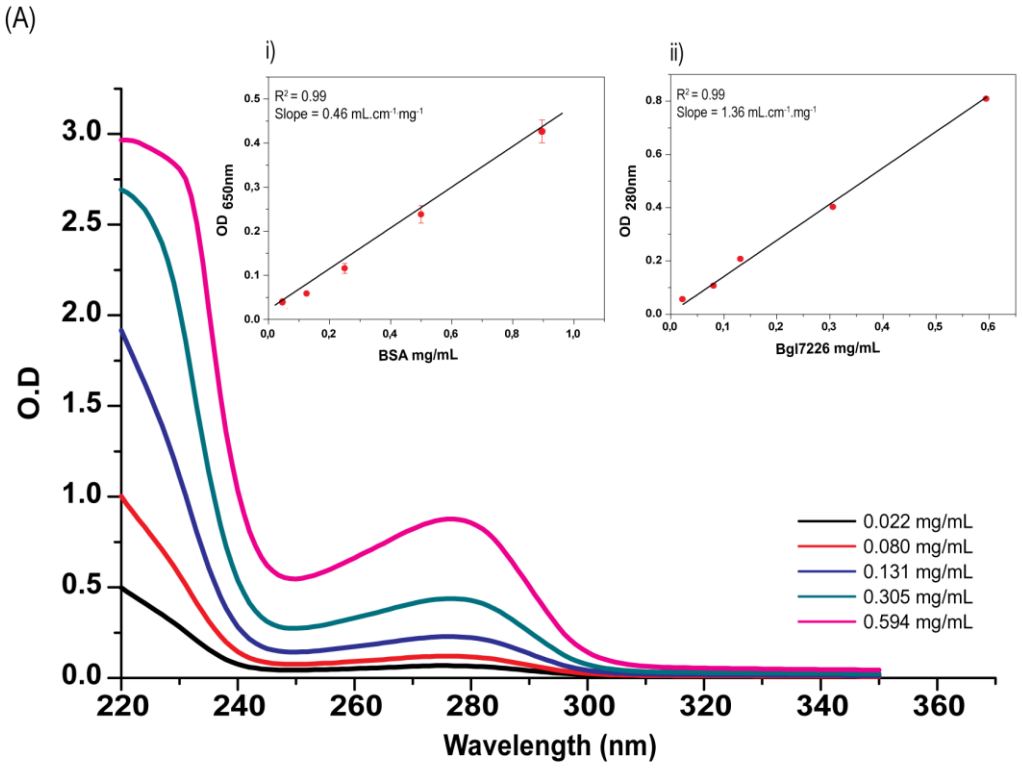


Fig S4. Bgl7226 experimental molar extinction coefficient. The molar extinction coefficient of Bgl7226 was 1.36 $mL.cm^{-1}.mg^{-1}$. i) graph and equation for BSA, relating $O.D_{650nm}$ x mg/mL; ii) graph and equation for Bgl7226, relating $O.D_{280nm}$ x mg/mL.

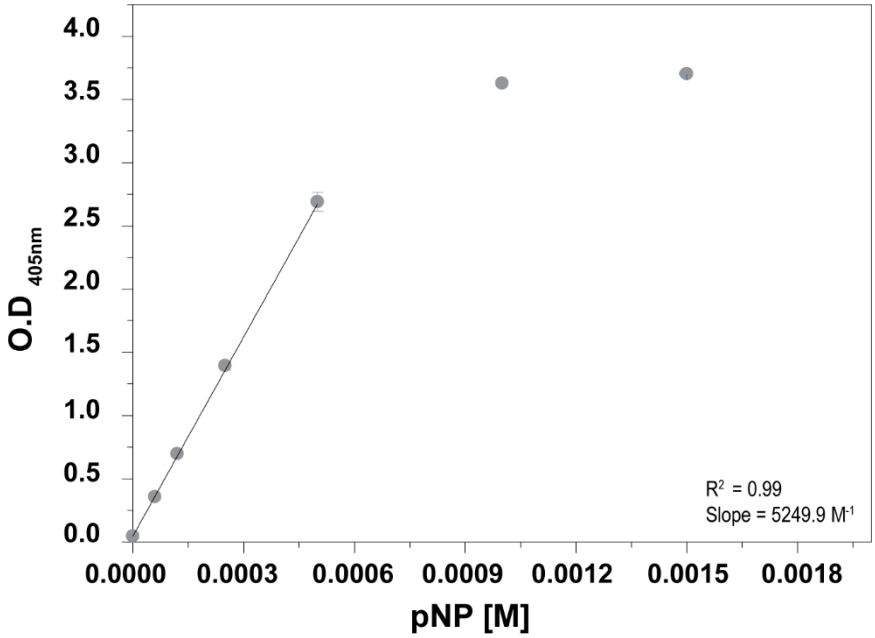


Fig S5. p-Nitrophenyl (pNP) standard curve. The standard pNP curve was used to calculate the concentration of pNP released by Bgl7226 hydrolysis of pNPG. The relationship between $O.D_{405nm}$ and the concentration of pNP in molarity is shown in the graph.

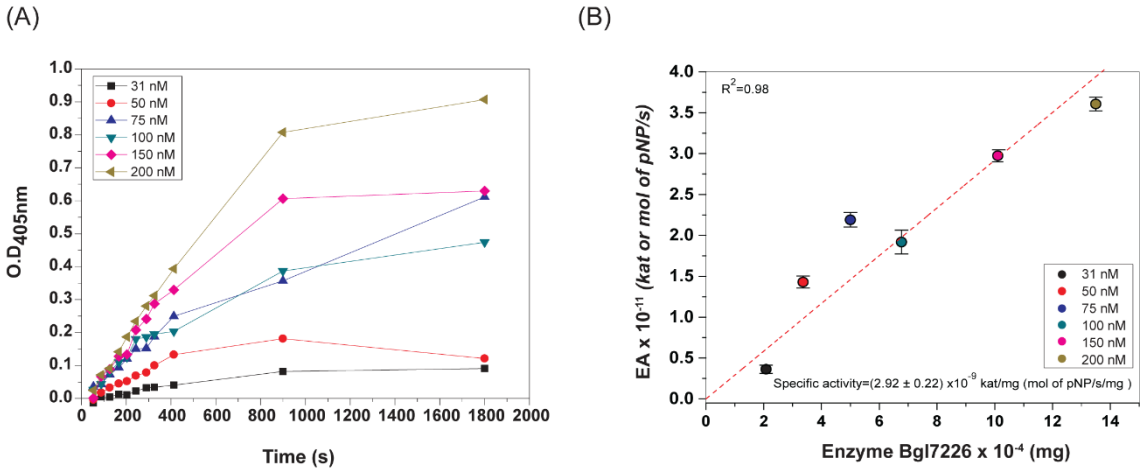


Fig S6. Effect of Bgl7226 concentration on enzymatic reaction. (A) $O.D_{405nm}$ over time at different enzyme concentrations. $O.D_{405nm}$ increases with enzyme concentration. (B) graph of (V_0 vs enzyme concentration). All concentrations were in a straight line. The concentration chosen for enzymatic assays was 100 nM or 0.1 µM.

Conclusão geral

Os estudos voltados ao aproveitamento de materiais renováveis é um dos principais focos da comunidade científica atual, uma vez que trazem impactos grandes ao ecossistema, reduzindo a utilização de fontes fósseis, como o petróleo, para produzir energia e químicos derivados. A lignocelulose, principal componente da parede celular vegetal, é uma das fontes renováveis mais produzidas anualmente, podendo ser aproveitada para a produção de energia limpa e químicos de valor agregado. Nesse trabalho, pensando em contribuir com a bioconversão desse material para aplicações biotecnológicas nós estudamos a microbiota intestinal do cupim, *Syntermes wheeleri*, e observamos que os filos Firmicutes, Spirochaetes e Proteobacteria são os mais abundantes e são responsáveis pela conversão da serapilheira em químicos absorvidos pelo inseto. Esses micróbios produzem uma seleção de enzimas GHs que lideram essa ação de quebra de macromoléculas. Os estudos mostraram uma compartimentalização desse grupo, a depender do proctodeal analisado, com as celulases estando mais presentes no P1 e as hemicelulases no P3. Junto a essas observações os valores de metatranscriptoma apontaram, na seleção para as enzimas, quais poderiam estar sendo mais produzidos no microhabit. As duas enzimas celulases, Exo8574 e Bgl7226, selecionadas e caracterizadas, por meio desses critérios, mostraram capacidade na conversão dos substratos sintéticos que mimetizam as ligações da celulose. Frente a outras enzimas bacterianas do mesmo EC, de cupins, mostraram valores variáveis próximos quanto a cinética enzimática. Para um melhor

entendimento do potencial biotecnológico dessas enzimas, estudos futuros serão desenvolvidos em biomassas pré-tratadas. De forma geral esses estudos trouxeram um esclarecimento sobre a participação dos microrganismos na bioconversão de lignocelulose dentro da espécie de cupim em estudo, apontando a compartimentalização funcional das CAZymes, e como esse microhabit pode ser bioprospectado para a descoberta de novas enzimas que poderiam nunca ser descobertas, por fazerem parte de organismos de difícil cultivo. Dessa forma expandindo o entendimento dessas classes de enzimas, pela forma detalhada de caracterização bioquímica e biofísica realizados, e contribuindo para um entendimento prático de aplicação comercial das mesmas.

Anexos – Participação em eventos