

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**EXPLORANDO UMA COMUNIDADE BACTERIANA
ISOLADA DE UMA PILHA DE BAGAÇO DE CANA-DE-
AÇÚCAR E SEU POTENCIAL FUNCIONAL NA
DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA**

Michelli Inácio Gonçalves Funicelli

Tecnóloga em biocombustíveis

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**EXPLORANDO UMA COMUNIDADE BACTERIANA
ISOLADA DE UMA PILHA DE BAGAÇO DE CANA-DE-
AÇÚCAR E SEU POTENCIAL FUNCIONAL NA
DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA**

Michelli Inácio Gonçalves Funicelli

Orientadora: Profa. Dra. Eliana G. de Macedo Lemos

Coorientador: Dr. Luciano Takeshi Kishi

Dissertação apresentada à Faculdade
de Ciências Agrárias e Veterinárias –
Unesp, Campus de Jaboticabal, como
parte das exigências para a obtenção
do título de Mestre em Microbiologia
Agropecuária

F982e

Funicelli, Michelli Inácio Gonçalves

Explorando uma comunidade bacteriana isolada de uma pilha de
bagaço de cana-de-açúcar e seu potencial funcional na degradação
de biomassa lignocelulósica / Michelli Inácio Gonçalves Funicelli. – –
Jaboticabal, 2018
x, 51 p. : il. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2018

Orientadora: Eliana Gertrudes de Macedo Lemos

Banca examinadora: Mariana Carina Frigieri, Daniel Guariz
Pinheiro

Bibliografia

1. Biocombustíveis. 2. Biomassa. 3. *Chitinophaga*. 4. *Labrys*. 5.
Pandoraea. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e
Veterinárias.

CDU 576.8:633.61

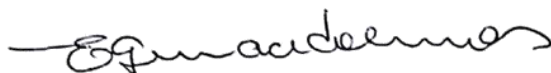
Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da
Informação – Diretoria Técnica de Biblioteca e Documentação - UNESP, Câmpus de
Jaboticabal.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: EXPLORANDO UMA COMUNIDADE BACTERIANA ISOLADA DE UMA PILHA DE BAGAÇO DE CANA-DE-AÇÚCAR E SEU POTENCIAL FUNCIONAL NA DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA

AUTORA: MICHELLI INACIO GONÇALVES FUNNICELLI
ORIENTADORA: ELIANA GERTRUDES DE MACEDO LEMOS
COORIENTADOR: LUCIANO TAKESHI KISHI

Aprovada como parte das exigências para obtenção do Título de Mestra em MICROBIOLOGIA AGROPECUÁRIA, pela Comissão Examinadora:



Profa. Dra. ELIANA GERTRUDES DE MACEDO LEMOS
Departamento de Tecnologia / FCAV / UNESP - Jaboticabal



Profa. Dra. MARIANA CARINA FRIGIERI SALARO
Centro Estadual de Educação Tecnológica Paula Souza / FATEC - Jaboticabal/SP



Prof. Dr. DANIEL GUARIZ PINHEIRO
Departamento de Tecnologia / FCAV / UNESP - Jaboticabal

Jaboticabal, 26 de fevereiro de 2018

DADOS CURRICULARES DO AUTOR

Michelli Inácio Gonçalves Funicelli – Nasceu em 16 de outubro de 1991, em São José do Rio Preto, São Paulo, Brasil, filha de Irene Inácio Gonçalves e Carlos Roberto Funicelli. Graduiu-se em Tecnologia em Biocombustíveis, em dezembro de 2015 pela Faculdade de Tecnologia, Fatec de Jaboticabal. Ingressou em março de 2016 no Curso de Mestrado, área de concentração em Microbiologia Agropecuária, na Faculdade de Ciências Agrárias e Veterinárias – UNESP, Campus de Jaboticabal. Em 26 de fevereiro de 2018 defendeu sua dissertação de mestrado em Microbiologia Agropecuária.

DEDICATÓRIA

Dedico a Deus!

Aos meus familiares, especialmente meus pais, Irene e Carlos (*in memoriam*),
as minhas irmãs, Elaine (Tata), Karine e Barbara, pelo apoio e por todo o amor;

A vocês, dedico meu trabalho e o meu amor.

AGRADECIMENTOS

A Prof^ª. Dr^ª. Eliana Gertrudes de Macedo Lemos, pela amizade, paciência, compreensão, confiança e oportunidade oferecida durante esses anos.

Ao Dr. Luciano Takeshi Kishi pela coorientação, auxílio e elaboração deste trabalho.

Ao Dr. João Carlos Campanharo por todos os ensinamentos, dedicação e por todas as vezes que me ajudou no desenvolvimento deste trabalho e à Dr^ª. Camila Fernandes pelo auxílio e pelos conhecimentos compartilhados.

Aos membros das bancas do exame de qualificação e de defesa, Prof. Dr. Daniel Guariz Pinheiro, Dr. Danilo Oliveira de Alvarenga e Prof^ª. Dr^ª. Mariana Carina Frigieri por compartilharem seus conhecimentos e pelas sugestões oferecidas que contribuíram para o melhoramento desse trabalho.

Às minhas amigas Natália Sarmanho, Pâmela Maldaner, Bárbara Bonfá, Milena Tavares, Gabriela Cabral, Bruna Fernanda e Tatiane Fernanda pelo companheirismo dentro e fora do laboratório, o apoio e as diversas risadas compartilhadas.

À toda equipe do Laboratório de Bioquímica de Microrganismos e Plantas, pela convivência, ensinamentos e pelas contribuições a esse trabalho.

Aos meus amigos Dayanne Fernandes e Silas do Laboratório Enzimologia Imunoquímica e Aplicadas e Luis Teheran, Maria Fernanda e Rafael Correia do Laboratório de Bioinformática pela amizade e troca de conhecimentos.

A todos os meus familiares, especialmente minha mãe, Irene Inácio Gonçalves, às minhas irmãs Elaine Viana Alves, Karine Gonçalves Funicelli e Barbara Gonçalves Funicelli pelo amor e carinho, compreensão e por acreditarem em mim.

Ao Gustavo Alan Siniscalchi pelo companheirismo, carinho, apoio e compreensão durante todos estes anos.

A todos os meus amigos que me acompanharam e apoiaram durante essa caminhada.

Ao programa de pós-graduação em Microbiologia Agropecuária.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

A todos que direta ou indiretamente contribuíram para a realização desse trabalho.

Meus sinceros agradecimentos.

SUMÁRIO

| | Páginas |
|--|---------|
| RESUMO- | iii |
| ABSTRACT- | iv |
| LISTA DE FIGURAS | v |
| LISTA DE TABELAS | vi |
| 1. INTRODUÇÃO | 1 |
| 2. OBJETIVOS | 3 |
| 3. REVISÃO DE LITERATURA | 4 |
| 3.1. Biomassa vegetal e cana-de-açúcar | 4 |
| 3.2. Composição da biomassa lignocelulósica | 4 |
| 3.2.1. Estrutura da celulose | 5 |
| 3.2.2. Estrutura da hemicelulose | 6 |
| 3.2.3. Estrutura da pectina | 6 |
| 3.2.4. Estrutura da lignina | 7 |
| 3.3. Conversão da biomassa lignocelulósica em etanol | 7 |
| 3.4. Enzimas envolvidas na degradação de material lignocelulósico | 8 |
| 3.5. Metagenômica | 11 |
| 4. MATERIAL E MÉTODOS | 12 |
| 4.1. Isolamento de bactérias da comunidade bacteriana CB10b | 14 |
| 4.1.1. Seleção de bactérias produtoras de celulases | 14 |
| 4.1.2. Obtenção do DNA genômico do isolados | 15 |
| 4.1.3. Amplificação das regiões do DNA que codificam o gene 16S rRNA e da região do espaçador transcrito interno (ITS) entre os genes 16S-23S rRNA | 15 |
| 4.1.4. Sequenciamento do gene 16S rRNA e da região do ITS | 16 |
| 4.1.5. Análise do sequenciamento do gene 16S rRNA e ITS | 17 |
| 4.1.6. Análise filogenética do gene 16S rRNA e ITS | 17 |
| 4.2. Extração de DNA total e sequenciamento da comunidade bacteriana CB10b | 17 |
| 4.2.1. Avaliação da melhor estratégia de montagem para o conjunto de dados da comunidade bacteriana (CB10b) | 18 |
| 4.2.2. Montagem, classificação taxonômica e perfil funcional da comunidade bacteriana (CB10b) | 18 |
| 4.2.3. Avaliação do potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica | 19 |

| | |
|---|-----------|
| 4.2.4. Recuperação do genoma parcial a partir do metagenoma do CB10b | 19 |
| 5. RESULTADOS E DISCUSSÃO | 21 |
| 5.1. Caracterização do isolamento de bactérias da comunidade bacteriana CB10b | 21 |
| 5.1.1. Agrupamento filogenético dos isolados da comunidade CB10b | 22 |
| 5.2. Avaliação da melhor estratégia de montagem para o conjunto de dados da comunidade bacteriana (CB10b) | 26 |
| 5.2.1. Montagem do CB10b e predição funcional | 28 |
| 5.2.2. Composição taxonômica e funcional da comunidade bacteriana CB10b | 29 |
| 5.2.3. Potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica | 34 |
| 5.2.4. Recuperação parcial de genomas | 40 |
| 6. CONCLUSÕES | 44 |
| 7. REFERÊNCIAS | 46 |

EXPLORANDO UMA COMUNIDADE BACTERIANA ISOLADA DE UMA PILHA DE BAGAÇO DE CANA-DE-AÇÚCAR E SEU POTENCIAL FUNCIONAL NA DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA

RESUMO- Os biocombustíveis são uma alternativa atraente para a substituição de combustíveis fósseis que impulsionaram o interesse global na conversão de biomassa lignocelulósica. Diversos trabalhos são realizados para otimização das etapas para produção de etanol de segunda geração em termos de custo global para o processo. A demanda pela busca de enzimas específicas utilizadas em processos industriais tem impulsionado a prospecção de novas enzimas microbianas com capacidade de desconstruir biomassa vegetal. Dentre elas, as enzimas do grupo das carbohidrases são as principais atuantes envolvidas no processo de conversão de biomassa. Diante disso, foi realizada a caracterização de uma comunidade microbiana denominada CB10b, que tem potencialidade em degradar a biomassa. Os estudos foram realizados a partir do sequenciamento total do DNA obtido da comunidade e da prospecção de genes úteis nos processos de conversão de biomassa vegetal. Foram isoladas 6 bactérias pertencentes aos respectivos gêneros: *Chitinophaga* (que apresentou atividade enzimática em carboximetilcelulose), *Pandoraea* e *Labrys*, identificado pelas análises do gene 16S rRNA e da região espaçadora intergênica (ITS) entre os genes 16S-23S rRNA. Por meio das análises dos genes foi possível identificar 16.340 ORFs, sendo anotadas 624 ORFs associadas ao banco de dados do CAZy, distribuídas entre as classes glicosil hidrolases (GHs) com 37,98% (237 ORFs); glicosil transferases (GTs) com 21,63% (135 ORFs); polissacarídeo liases (PLs) com 3,18% (20 ORFs); carboidrato esterases (CEs) com 17,00% (106 ORFs); atividades auxiliares (AAs) com 4,81% (30 ORFs) e módulos de ligação a carboidratos (CBMs) com 15,38% (96 ORFs). Adicionalmente foi possível recuperar através da análise do metagenoma, o genoma parcial de uma bactéria do gênero *Pandoraea*. Deste modo, o estudo com a comunidade bacteriana CB10b apresentou genes envolvidos a degradação de biomassa lignocelulósica.

Palavras-chave: biocombustíveis, biomassa, *Chitinophaga*, *Labrys*, *Pandoraea*.

ABSTRACT- Biofuels are an attractive alternative for replacing fossil fuels, bringing the world's attention to the conversion processes of lignocellulosic biomass. Multiple works are currently being done with to optimize the steps in second-generation ethanol production, seeking to reduce the global costs in this process. Demand for specific microbial enzymes is driving the prospection of novel ones able to degrade plant biomass. Among those, carbohydrases are the protagonists in these processes. In this context, we characterized a microbial community called CB10b and tested its potential to degrade biomass. We performed studies on this community by total DNA sequencing followed by prospecting genes related to conversion of plant biomass. Six bacteria were isolated, belonging to the genera: *Chitinophaga*, who presented enzymatic activity for carboxymethyl cellulose; *Pandoraea* and *Labrys*, identified through 16S rRNA and 16S-23S rRNA internal transcribed spacer (ITS) gene analysis. Some 16,340 ORFs were identified, of which 624 were associated with CAZy database, 37.98% (237 ORFs) classified as glycosyl hydrolases (GHs); 21.63% (135 ORFs) as glycosyltransferases (GTs); 17.00% (106 ORFs) were carbohydrate esterases (CEs); 15.38% (96 ORFs) as carbohydrate-binding modules; 4.81% (30 ORFs) as auxiliary activities (AAs) and 3.18% (20 ORFs) as polysaccharide lyase. In addition, it was possible to recover by metagenome analysis the partial genome of a bacterium of the genus *Pandoraea*. Thus, the study with the bacterial community CB10b presented genes involved the degradation of lignocellulosic biomass

Keywords: biofuels, biomass, *Chitinophaga*, *Labrys*, *Pandoraea*.

LISTA DE FIGURAS

| | Páginas |
|--|---------|
| Figura 1. Representação esquemática da estrutura da lignocelulose. | 5 |
| Figura 2. Esquema para a produção de bioetanol a partir bagaço de cana-de-açúcar. | 8 |
| Figura 3. Fluxograma metodológico | 13 |
| Figura 4. Teste para identificação de isolados produtores de celulases. | 21 |
| Figura 5. Árvore filogenética concatenada dos genes 16 rRNA e ITS dos isolados referentes ao gênero <i>Pandoraea</i> | 23 |
| Figura 6. Árvore filogenética do gene 16 rRNA para os isolados <i>Chitinophaga</i> e <i>Labrys</i> | 25 |
| Figura 7. Avaliação da qualidade das diferentes montagens do metagenoma com o programa MetaQuast..... | 27 |
| Figura 8. Distribuição taxonômica da comunidade bacteriana CB10b. | 30 |
| Figura 9. Composição funcional da comunidade bacteriana CB10b com base no banco de dados do Clustes Orthologous Groups (COG). | 33 |
| Figura 10. Distribuição taxonômica das classes de carboidrases associadas ao banco de dados do CAZY na comunidade bacteriana CB10b. | 35 |
| Figura 11. Distribuição das classes de enzimas associadas às carboidrases. | 37 |
| Figura 12. Distribuição taxonômica dos clusters separados pelo programa BusyBee. | 40 |
| Figura 13. Similaridade entre genomas de bactérias do gênero <i>Pandoraea</i> alinhados pelo programa Mauve mostrando as similaridades entre os genomas. | 42 |

LISTA DE TABELAS

Páginas

| | |
|--|----|
| Tabela 1. Oligonucleotídeos iniciadores utilizados na amplificação da região do 16S rRNA e da região do espaçador transcrito interno ITS. | 16 |
| Tabela 2. Descrição da montagem do metagenoma da comunidade bacteriana (CB10b) e predição funcional. | 28 |
| Tabela 3. Comparação entre o genoma de uma <i>Pandoraea pulmonicola</i> com diferentes espécies de <i>Pandoraea</i> deste estudo e outras espécies e <i>Chitinophaga pinensis</i> como controle para comparação. | 43 |

1. INTRODUÇÃO

Os biocombustíveis são uma alternativa atraente para a substituição dos combustíveis fósseis (BAYER; SHOLHAM; LAMED, 2013). O interesse global para utilização de fontes renováveis na conversão de biomassa lignocelulósica é impulsionada principalmente por questões associadas à demanda proveniente do crescimento populacional global e ambientais, referentes às emissões de gases para a atmosfera, como a liberação excessiva de dióxido de carbono.

Neste contexto a biomassa vegetal se torna significativamente interessante por representar uma fonte renovável e abundante para obtenção de químicos, incluindo combustíveis e precursores de polímeros. A sua utilização se torna uma estratégia importante de sustentabilidade e pode representar um avanço para as indústrias de energia e produtos químicos, que podem agregar valor aos subprodutos agrícolas e resíduos florestais e municipais (SWEENEY; XU, 2012).

A biomassa vegetal é composta por lignocelulose, que consiste em uma mistura altamente organizada e interligada de diferentes polímeros, contendo principalmente celulose (35-50%), hemiceluloses (20-50%) e lignina (10-35%) (BRETHAUER; STUDER, 2015). A composição da biomassa varia significativamente entre espécies e dentro de diferentes tecidos de uma espécie (PAULY; KEEGSTR, 2010). O principal desafio na sua utilização está associado à disposição da maior parte do carbono associado à hemicelulose e celulose que estão envolvidos pela lignina, que consiste em polímeros estruturais recalcitrantes da parede celular da planta que proporcionam proteção contra hidrólise (BROWN; CHANG, 2014; DE GONZALO et al., 2016).

A degradação da estrutura do material composto de lignocelulose ocorre por meio da ação de complexos de múltiplas enzimas produzidas por diversos microrganismos principalmente por fungos e bactérias presentes no solo que atuam em sinergia para quebrar as estruturas da lignocelulose (polissacarídeos) em monômeros ou dímeros de açúcares, fornecendo fonte de carbono e energia para os microrganismos (HIMMEL et al., 2010; SWEENEY; XU, 2012; VAN DYK; PLETSCHE, 2012).

Entretanto a recalcitrância da lignina limita a ação de microrganismos, tornando a biomassa lignocelulósica um material resistente à degradação

biológica (TAHERZADEH; KARIMI, 2008). Devido à característica de resistência da lignocelulose são necessários pré-tratamentos físicos, químicos ou biológicos do material lignocelulósico, para romper a estrutura da parede celular vegetal, permitindo o acesso às frações celulósicas da biomassa vegetal (BRETHAUER; STUDER, 2015; DE CASTRO; PEREIRA, 2010).

A utilização de consórcios microbianos que produzem enzimas altamente específicas com a capacidade de quebrar a estrutura da lignocelulose tem obtido resultados promissores como tratamentos para a desconstrução da biomassa lignocelulósica (WONGWILAIWALIN et al., 2013).

Deste modo, uma comunidade obtida da pilha de bagaço de cana-de-açúcar foi objeto deste estudo, no qual realizou-se o isolamento de bactérias desta comunidade, assim como a caracterização dos isolados com base no gene 16S rRNA e da região do espaçador transcrito interno (ITS) e a avaliação da atividade enzimática em placas de petri contendo carboximetilcelulose e avicel no meio de cultura sólido.

Uma comunidade bacteriana pode conter uma riqueza em diversidade maior do que a nossa capacidade de cultivar os isolados, representada por microrganismos em que as necessidades fisiológicas e nutricionais são desconhecidas para que estas possam ser reproduzidas artificialmente, de forma que os mesmos só se desenvolvem quando associados às comunidades. Neste contexto um estudo focado apenas no isolamento e cultivo não poderiam fornecer informação suficientes do papel biológico de diversos membros da comunidade, sendo estas informações melhor alcançadas por uma abordagem metagenômica.

Através das análises realizadas do sequenciamento do metagenoma total, o estudo realizado com a comunidade bacteriana indica que esta apresenta um potencial metabólico para atuar na desconstrução da biomassa lignocelulósica, bem como foi possível recuperar o genoma parcial de bactéria pertencente ao gênero *Pandoraea*.

2. OBJETIVOS

O presente estudo teve como objetivo caracterizar uma comunidade obtida de uma pilha de bagaço de cana-de-açúcar (CB10b) por meio da microbiologia básica e por meio do sequenciamento do DNA total da comunidade e com isso explorar o potencial gênico funcional da comunidade, a fim de descobrir enzimas com grande importância para a desconstrução de material lignocelulósico de interesse para o setor industrial, como no caso da produção do etanol de segunda geração.

3. REVISÃO DE LITERATURA

3.1. Biomassa da cana-de-açúcar

A biomassa vegetal é um termo geral para todo o material orgânico que decorre de plantas (incluindo algas, árvores e culturas) (DEMIRBAS, 2010) e consiste em fonte de energia renovável proveniente de matéria orgânica, que pode ser convertida em bioenergia.

A cultura de cana-de-açúcar resulta em uma biomassa considerada uma das mais importantes no contexto da agricultura brasileira para a obtenção de energia renovável, devido à sua vasta produção como fonte de matéria prima predominante para a indústria de bioetanol (WELFLE, 2017).

O bagaço da cana-de-açúcar é considerado o maior resíduo da agroindústria brasileira (VARGAS BETANCUR; PEREIRA JR., 2010). A composição do bagaço em matéria seca apresenta em média aproximadamente 40% de celulose, 28% de hemicelulose e 22% de lignina (ANTONIO BIZZO et al., 2014).

Com o intuito de ampliar a produção de bioetanol, a biomassa lignocelulósica tem sido reconhecida como uma fonte potencial e sustentável, para a obtenção de açúcares mistos para fermentação na produção de biocombustíveis entre outros biomateriais (HIMMEL et al., 2007).

3.2. Composição da biomassa lignocelulósica

A parede celular vegetal representa uma vasta fonte renovável de carbono presente na biosfera que requer várias estratégias enzimáticas para desconstruir os polissacarídeos estruturais de sua composição (DUTTA; WU, 2014).

A parede celular da planta consiste em uma mistura complexa de polissacarídeos na qual celulose, hemicelulose e lignina são seus principais constituintes e estão interligados, formando uma estrutura rígida oferecendo, assim, sustentação e proteção às células vegetais (BAYER; SHOLHAM; LAMED, 2013) (Figura 1). Nesta figura é possível observar que a parede celular vegetal é composta por microfibrilas contendo celulose recobertas por moléculas de hemicelulose e lignina.

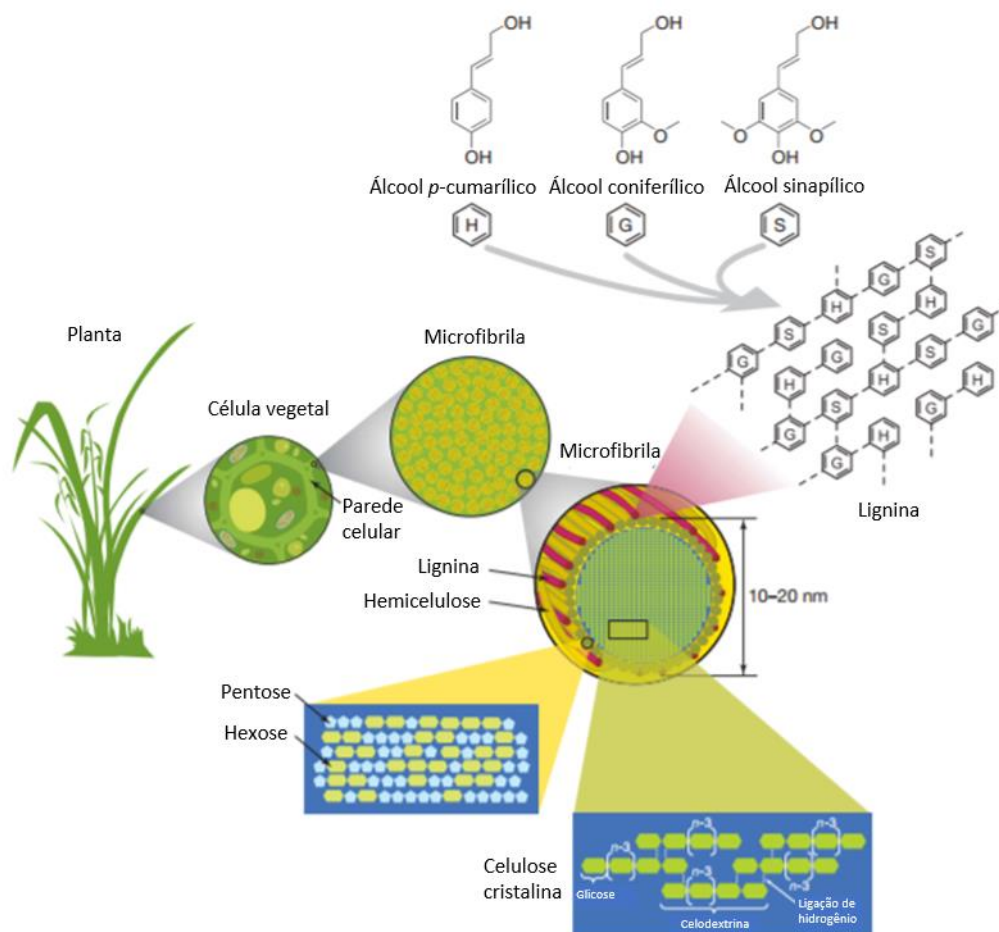


Figura 1. Representação esquemática da estrutura da lignocelulose. Fonte: (RUBIN, 2008).

3.2.1. Estrutura da celulose

A celulose representa o polímero constituinte de matéria vegetal mais abundante do planeta (TAYLOR; BRUMM, 2014). A celulose é um homopolímero estável, constituído por uma cadeia linear formada por monômeros de glicose unidas por ligações glicosídicas do tipo β -1,4 (BAYER; SHOLHAM; LAMED, 2013).

Sua estrutura altamente cristalina se deve pela compactação das fibrilas de celulose que são formadas por meio de ligações de hidrogênio que unem as cadeias de celulose em cadeias de fibras, deste modo a celulose cristalina se refere às fibrilas organizadas e compactas que são mais suscetíveis a

degradação principalmente pela ação de enzimas do que a celulose amorfa que não está organizada desta forma (MATHEWS; PAWLAK; GRUNDEN, 2015).

Além disso, as moléculas de celulose estão envolvidas pelos polímeros de hemicelulose e lignina, o que torna mais difícil o acesso a fração da celulose e serve como proteção, dificultando a hidrólise e degradação da estrutura da parede celular.

3.2.2. Estrutura da hemicelulose

Hemicelulose consiste em heteropolímero interligado às microfibrilas de celulose (SCHÄDEL et al., 2010) e é o segundo componente mais abundante presente na parede celular vegetal (GLASS et al., 2013; ISIKGOR; BECER, 2015).

Este polissacarídeo apresenta uma estrutura mais diversificada que a celulose (KARNAOURI et al., 2014). A classificação da hemicelulose é baseada na composição de sua cadeia principal, que geralmente é composta por um ou dois açúcares (BAYER; SHOLHAM; LAMED, 2013). Os monossacarídeos constituintes do heteropolímero de hemicelulose podem variar entre as pentoses (xilose, ramnose, arabinose), as hexoses (manose, glicose, galactose) e açúcares acetilados (ISIKGOR; BECER, 2015).

A xilana é a classe de hemicelulose mais abundante em todas as plantas superiores, esta possui uma cadeia linear principal composta por ligações glicosídicas β -1,4 de unidades de D-xilose ligadas a cadeias laterais compostas por diferentes tipos de açúcares e resíduos ácidos de açúcares (arabinose, glicose, galactose, ramnose, ácido glucurônico e ácido galacturônico, dependendo da espécie e do tipo de parede)(GUERRIERO et al., 2016; SCHÄDEL et al., 2010).

3.2.3. Estrutura da pectina

A pectina consiste em um polissacarídeo estrutural presente na parede celular vegetal, com a função de regular o desenvolvimento e as respostas de defesa das plantas (BAYER; SHOHAM; LAMED, 2013).

São heteropolissacarídeos compostos por um ácido galacturônico unidos por ligações do tipo α (1-4) (BAYER; SHOLHAM; LAMED, 2013). As classes

estruturais dos polissacarídeos de pectina incluem homogalacturonano (HG), xilogalacturonano (XGA), apiogalacturonano (AGA), Ramnogalacturonano II (RG-II) e Ramnogalacturonano I (RG-I) (CAFFALL; MOHNEN, 2009).

3.2.4. Estrutura da lignina

A lignina tem como papel biológico nas plantas aumentar a integridade da parede celular e promover resistência ao ataque de agentes patogênicos (BROWN; CHANG, 2014). Sua estrutura consiste em um polímero heterogêneo altamente complexo constituído de polímeros aromáticos derivados por álcoois hidroxicinamil (ou monolignóis), entre os principais o álcool coniferil, álcool sinapílico e com quantidades tipicamente menores o álcool *p*-cumaril (RALPH et al., 2004; VANHOLME et al., 2010).

3.3. Conversão da biomassa lignocelulósica em etanol

A crescente demanda por etanol de primeira geração requer alta produção de matéria-prima, levando a preocupações entre a demanda por alimentos e combustível e requer um aumento de uso da terra (BEZERRA; RAGAUSKAS, 2016).

Os biocombustíveis de segunda geração são uma alternativa para evitar muitas dessas preocupações, uma vez que depende de fontes não alimentares, como os materiais lignocelulósicos, que são relativamente baratos e estão disponíveis em grandes quantidades. A biomassa mais comumente avaliada para produção de etanol de segunda geração é bagaço de cana de açúcar.

Para o bioprocessamento da biomassa lignocelulósica é necessário quebrar a recalcitrância das paredes celulares da biomassa, para que seja possível ocorrer a desconstrução do material através da atividade microbiana.

A produção de etanol lignocelulósico consiste basicamente em uma etapa de pré-tratamento, hidrólise enzimática e a fermentação (Figura 2).

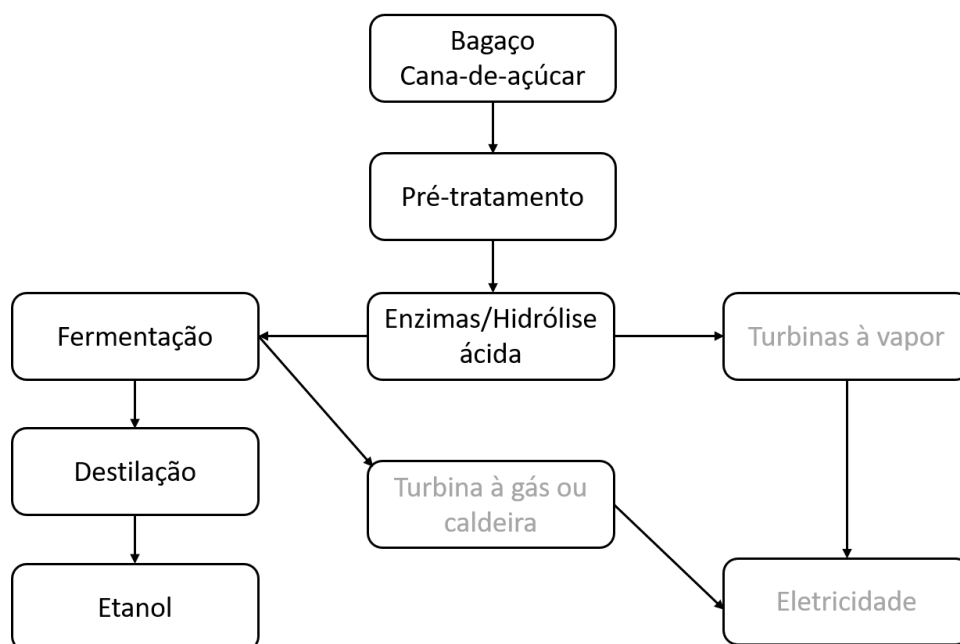


Figura 2. Esquema para a produção de bioetanol a partir de bagaço de cana-de-açúcar adaptado (BEZERRA; RAGAUSKAS, 2016).

Uma variedade de tecnologias de pré-tratamento de materiais lignocelulósicos tem sido investigada, como processos físicos (mecânicos), físico-químicos, químicos, biológicos ou combinados (BRETHAUER; STUDER, 2015; ISIKGOR; BECER, 2015).

Para superar a recalcitrância característica da lignocelulose para a liberação de açúcares fermentescíveis, o pré-tratamento é um dos principais passos na conversão biológica de bagaço em bioetanol (BEZERRA; RAGAUSKAS, 2016; BRETHAUER; STUDER, 2015). A redução da recalcitrância tem sido atribuída a vários fatores, como mudanças nas interações de hemicelulose, celulose e lignina; aumento do acesso a fração da celulose; remoção de carboidratos lignificados e redução da cristalinidade da celulose na parede celular da biomassa, melhorando assim, o acesso das enzimas hidrolíticas aos polímeros de açúcar (BEZERRA; RAGAUSKAS, 2016).

3.4. Enzimas envolvidas na degradação de material lignocelulósico

A desconstrução enzimática do material lignocelulósico ocorre através de complexo de múltiplas enzimas que podem ser produzidas por diversos microrganismos, que se associam para quebrar as macromoléculas (polissacarídeos) em monômeros ou dímeros de açúcares, fornecendo fonte de

carbono e energia para os microrganismos (BAYER; SHOLHAM; LAMED, 2013; MALHERBE; CLOETE, 2002).

Esses sistemas de enzimas altamente especializados, muitas vezes incluem a atividade de celulases, hemicelulases e outras glicosil hidrolases relacionadas, assim como polissacarídeos liases e as esterases de carboidratos para atuar na degradação dos polissacarídeos presentes na parede celular (HIMMEL et al., 2010).

O banco de dados do CAZy (<http://www.cazy.org/>) consiste em uma base de dados que categoriza as enzimas de degradação de polissacarídeos como as Glicosil hidrolases (GHs) que hidrolisam e/ou rearranjam ligações glicosídicas, Glicosil transferases (GTs) que estão associados à formação de ligações glicosídicas, Polissacarídeos Liases (PLs) que clivam ligações glicosídicas de forma não hidrolítica, Carboidratos Esterases (CEs) que hidrolisam ésteres de carboidrato, Atividades Auxiliares (AAs) são enzimas tipo redox que agem em conjunto com as carboidrases e os Módulos de Ligação a Carboidratos (CBMs) que atuam na adesão de carboidratos (ARMSTRONG et al., 2015; LOMBARD et al., 2014).

A degradação enzimática da celulose ocorre mediante a ação de celulases que são enzimas capazes de hidrolisar as ligações glicosídicas β (1 \rightarrow 4) da celulose e são divididas em três tipos principais de atividades enzimáticas, endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.74 / EC 3.2.1.91) e β -glicosidases (EC 3.2.1.21) (LYND et al., 2002).

As endoglucanases atuam randomicamente em regiões internas e amorfas da cadeia da celulose, gerando oligossacarídeos de comprimentos menores, enquanto que, as exoglucanases atuam nas extremidades redutoras e não redutoras da cadeia, liberando unidades de glicose ou celobiose, (LYND et al., 2002), a principal representante deste grupo de enzimas responsáveis pela degradação da celulose microcristalina (SWEENEY; XU, 2012). Já as enzimas β -glicosidases hidrolisam as celobioses liberando moléculas de glicose, atuando somente sobre o substrato solúvel (SWEENEY; XU, 2012). As celulases estão principalmente distribuídas entre as famílias GH5, GH6, GH7 (encontradas em fungos), GH8, GH9, GH12, GH44, GH45, GH48, GH74 e GH124 (BAYER; SHOLHAM; LAMED, 2013).

As hemicelulases representam um grupo diversificado de enzimas que hidrolizam hemiceluloses. Os módulos catalíticos das hemicelulases são representados principalmente pelas classes de glicosil hidrolases (GHs) que hidrolisam ligações glicosídicas e pelas esterases de carboidratos (CEs), que hidrolisam as ligações éster de grupos laterais de acetato ou de ácido ferúlico (SHALLOM; SHOHAM, 2003).

A xilana representa o polímero mais abundante entre as hemiceluloses, sua estrutura é constituída por uma cadeia principal homopolimérica com unidades de D-xilopirranose unidas por ligações glicosídicas do tipo β -1,4 e cadeias laterais podem conter arabinose, galactose, ácido glucurônico, ácido acético, ácido p-cumárico e ferúlico (SAHA; JORDAN; BOTHAST, 2009).

A hidrólise enzimática da estrutura da xilana demanda da atividade de endo-1,4-xilanase (EC 3.2.1.8) e xilosidases (EC 3.2.1.37), sendo que as endo-xilanases atuam aleatoriamente na cadeia principal de xilana e a xilosidase hidrolisam xilobiose e xilo-oligossacarídeos em unidades xilose (SAHA; JORDAN; BOTHAST, 2009).

Entre as enzimas que atuam nas cadeias laterais as L-arabinofuranosidase (EC 3.2.1.55) e glucuronidase (EC 3.2.1.139) atuam nas ramificações para remover resíduos de arabinose e ácido glucurônico respectivamente. Enquanto que as enzimas acetil xilana esterase (EC 3.1.1.72), feruloil esterase (EC 3.1.1.73) e p-cumaril esterase hidrolisam as ligações ésteres entre unidades de xilose da cadeia de xilana e ácido acético ou entre os resíduos da cadeia lateral da arabinose e ácidos fenólicos como ácido ferúlico e ácido p-cumárico (SAHA; JORDAN; BOTHAST, 2009).

As xilanases estão principalmente distribuídas entre as famílias GH10, GH11 e também encontradas nas famílias GH5 e GH8 enquanto que β -xilosidases são agrupadas nas famílias GH3, GH30, GH39, GH43, GH52, GH54, GH116 e GH120 (BAYER; SHOLHAM; LAMED, 2013).

As enzimas envolvidas na degradação da pectina estão agrupadas nas famílias dos polissacarídeo liases, distribuídas principalmente nas famílias PL1, PL3, PL9, PL10 e PL11 representadas pelas pectato liases (EC 4.2.2.2), exopectato liase ou exopoligalacturonato liase (EC 4.2.2.9), pectina liase (EC 4.2.2.10) e ramnogalacturonan liase (EC 4.2.2.23) (BAYER; SHOLHAM; LAMED, 2013).

A degradação da lignina é um processo complexo que envolve a atividade de muitas enzimas hidrolíticas, oxidativas e de algumas proteínas complementares (GUERRIERO et al., 2016). As enzimas lignina-peroxidase (EC 1.11.1.14), manganês-peroxidase (EC 1.11.1.13), peroxidase versátil (EC 1.11.1.16) e lacase (EC 1.10.3.2) estão associadas à degradação de lignina (DATTA et al., 2017). Estas são classificadas como enzimas de atividades auxiliares e estão distribuídas nas famílias AA1, AA2, AA9 e AA10 (LEVASSEUR et al., 2013).

3.5. Metagenômica

O aumento da demanda por enzimas específicas em diversos setores industriais, como os de produção de detergentes, têxteis, indústrias de celulose e papel, síntese orgânica e indústria de biocombustíveis, impulsionou a prospecção de novas enzimas microbianas e, conseqüentemente, tem exigido a busca de novos microrganismos para a obtenção destas (LI et al., 2012).

Neste contexto, a metagenômica surge como uma proposta de acesso à população microbiana e suas informações ainda não alcançadas através de métodos de microbiologia clássica. Esta técnica permite acessar diretamente o conteúdo genético de comunidades microbianas complexas, ou seja, é um método baseado na análise genomas contidos diversos *habitats* (HANDELSMAN et al., 1998; OULAS et al., 2015).

O termo metagenômica foi designado apenas para análise funcional baseada em sequências de um conjunto de genomas contidos em uma amostra ambiental, mas atualmente também abrange uma segunda estratégia baseada na amplificação de genes alvos e subseqüentemente seu sequenciamento (OULAS et al., 2015).

A primeira abordagem é baseada no sequenciamento dos ácidos nucleicos de um ambiente, o DNA total, que permite abordar a estrutura da comunidade microbiana em relação às funções dos microrganismos por meio do seu conteúdo gênico e as interações entre os mesmos. Enquanto a segunda abordagem baseada em genes marcadores consiste apenas em obter informações da distribuição do perfil taxonômico de uma determinada amostra.

As técnicas da metagenômica são ainda impulsionadas pelo avanço no rendimento dos sequenciadores, os chamados Next-Generation Sequencers (NGS), que permitem maior acesso às enzimas, fármacos e produtos microbianos de interesse biotecnológico. Assim, utilizando-se destas tecnologias, busca-se atender a demanda de mercado por estes produtos, mas ainda há muito o que se explorar para concretizar o potencial deste conjunto de técnicas (BROWN et al., 2012).

4. MATERIAL E MÉTODOS

Na Figura 3 tem-se o fluxograma experimental das abordagens utilizadas para avaliar a comunidade bacteriana utilizada no presente estudo.

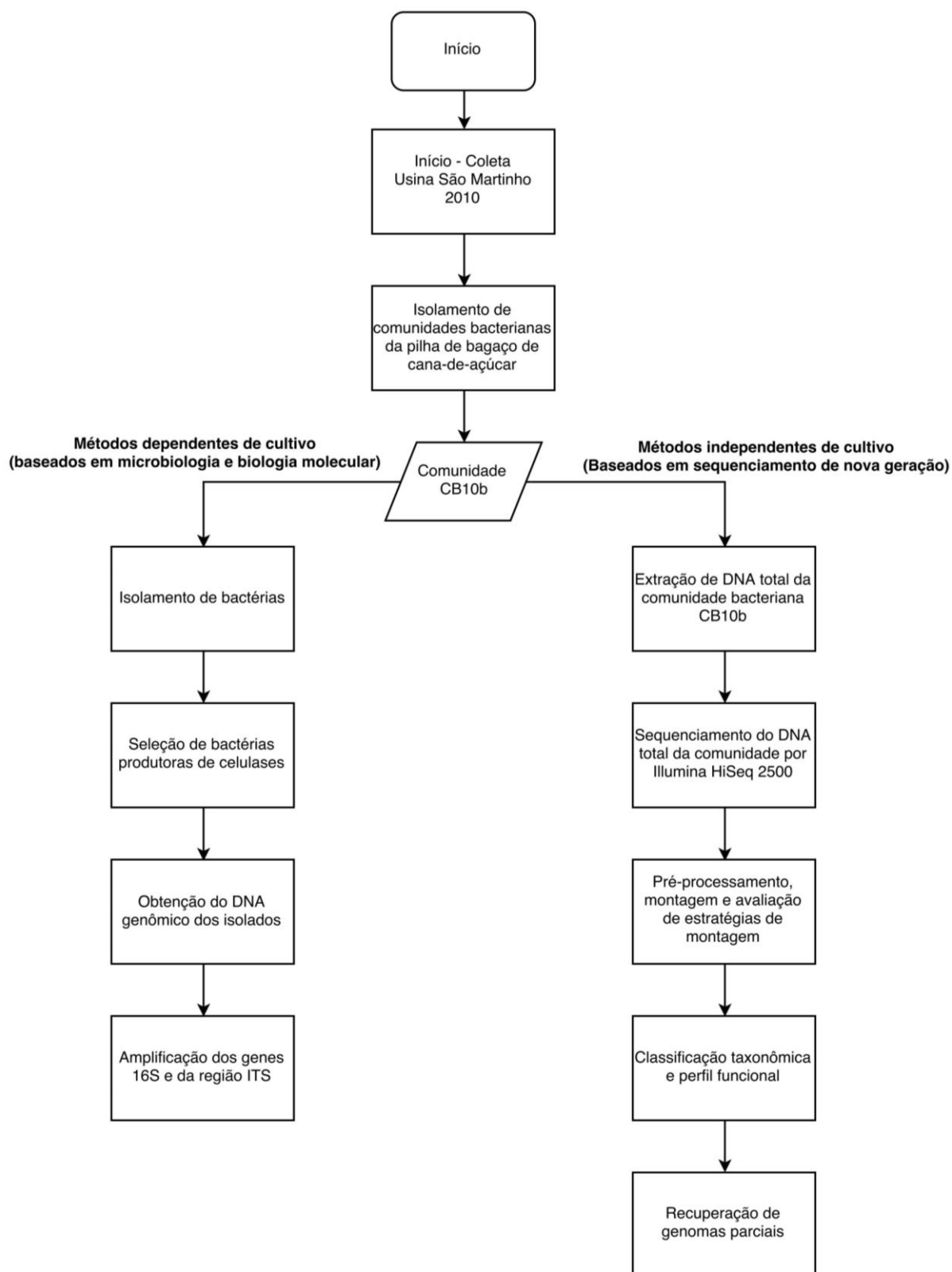


Figura 3. Fluxograma metodológico

4.1. Isolamento de bactérias da comunidade bacteriana CB10b

Os estudos foram realizados a partir de uma comunidade bacteriana isolada de uma pilha de bagaço de cana-de-açúcar e pertencente à coleção de microrganismos do Laboratório de Bioquímica de Microrganismos e Plantas da faculdade de Ciências Agrárias e Veterinárias da UNESP (LBMP) (ALMEIDA, 2011).

Uma alíquota de 100 μL do estoque de uma comunidade bacteriana (CB10b) em glicerol 20% foi adicionada à 900 μL de solução salina 0,85% (m/v) e submetida a diluições seriadas até 10^{-9} . Foram espalhadas 100 μL de cada diluição em placas de petri contendo meio de cultura sólido PGE (0,5 g.L^{-1} K_2HPO_4 ; 0,2 g.L^{-1} $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 0,1 g.L^{-1} NaCl , 1,0 g.L^{-1} de extrato de levedura, 10,0 g.L^{-1} de glicose e 9,0 g.L^{-1} de ágar, pH 6,8). As placas foram incubadas a 30°C durante 72 horas. Após o período de incubação as colônias de bactérias foram isoladas com base nas características morfológicas relacionadas ao tamanho da colônia, forma, elevação, cor e textura. Os isolados selecionados foram cultivados isoladamente no mesmo meio de cultura sob as mesmas condições e usados nos ensaios microbiológicos, bioquímicos e moleculares.

4.1.1. Seleção de bactérias produtoras de celulases

Para avaliar a produção de celulases, os isolados foram cultivados em meio de cultura PGE sólido e incubados a 30°C durante 48 horas. Após este período, foi realizado o preparo de um pré-inóculo em meio PGE líquido para cada isolado em uma densidade ótica com comprimento de leitura de 600 nm ($\text{DO}_{600\text{ nm}}$) de 0,3. Após a obtenção das suspensões, uma alíquota de 10 μL de cada isolado foi transferida para meio CMC sólido e meio AVICEL sólido, sendo então incubados a 30°C por 5 dias. O meio CMC sólido é constituído por 5,0 g.L^{-1} de peptona; 3,0 g.L^{-1} de extrato de levedura e 0,2% (m/v) CMC (Carboximetilcelulose – Synthe) com pH 7,0. Para o meio AVICEL sólido o CMC foi substituído por 0,2% (m/v) AVICEL®-Celulose microcristalina (Fluka analytical). Após o tempo de cultivo as placas contendo os isolados foram coradas com Vermelho Congo 1% (m/v) durante 30 minutos e em seguida foram lavadas com solução de NaCl 1 M estéril durante 10 minutos para retirar o

excesso do corante da placa. A atividade enzimática extracelular ou índice enzimático (IE) foi calculado pela relação entre o diâmetro médio do halo de degradação e o diâmetro médio das colônias (HANKIN; ANAGNOSTAKIS, 1975).

4.1.2. Obtenção do DNA genômico dos isolados

Os isolados foram cultivados em meio PGE durante 48 horas a 30 °C com agitação a 150 rpm. Após o cultivo os isolados bacterianos foram sedimentados por centrifugação (10.000 x g) e uma alíquota da massa celular de cada um foi coletada para extração do DNA genômico utilizando o Kit de purificação QIAamp DNA Mini kit ® (Qiagen). A integridade das amostras de DNA extraídas foi confirmada por eletroforese em gel de agarose, e quantificação por meio da análise de densidade óptica em espectrofotômetro NanoDrop® ND-1000 UV-Vis (ThermoFisher).

4.1.3. Amplificação das regiões do DNA que codificam o gene 16S rRNA e da região do espaçador transcrito interno (ITS) entre os genes 16S-23S rRNA

As amostras de DNA extraídas dos isolados foram amplificadas com oligonucleotídeos iniciadores (*primers*) universais fD1 e o rD1 específicos para o gene 16S rRNA bacteriano (WEISBURG et al., 1991). A reação de amplificação para cada amostra foi realizada sob as seguintes condições: 0,2 mM para cada dNTPs; tampão contendo 20 mM Tris-HCl (pH 8,4) e 50 mM KCl; 1,5 mM de MgCl₂; 10 pmol de cada *primer*; 30 ng de DNA; 0,5U da enzima taq DNA Polimerase em um volume final de 20 µL. As amostras foram amplificadas em um termociclador MJ Research® PTC-100 seguindo o programa: 95 °C, por 1 minuto; 39 ciclos de 95 °C por 2 minutos, 46 °C por 1 minuto, 72 °C por 1 minuto; seguidos de uma extensão final por 5 minutos, a 72 °C. As amostras de DNA dos isolados foram amplificadas com os *primers* FGPS 1490 e FGPS 132 para a região referente ao ITS (LAGUERRE et al., 1996). A reação de amplificação para cada amostra foi realizada sob as seguintes condições: 0,2 mM para cada dNTPs; tampão contendo 20 mM Tris-HCl (pH 8,4) e 50 mM KCl; 1,5 mM de MgCl₂; 5 pmol de cada *primer*; 30 ng de DNA; 0,5 U de enzima taq DNA Polimerase em um volume final de 20 µL, seguindo o programa: um ciclo de 2

min. a 94°C; 30 ciclos de 30 seg. a 94°C, 50°C por 30 seg., 72°C por 1 min; um ciclo final de 5 min. a 72°C. Posteriormente os produtos das PCR foram purificados utilizando o Kit Zymoclean™ Gel DNA Recovery.

4.1.4. Sequenciamento do gene 16S rRNA e da região do ITS

Os produtos das PCRs purificados foram utilizados nas reações de sequenciamento realizadas em microplaca utilizando BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) de acordo com as instruções do fabricante. Para obter a sequência completa do gene do rRNA 16S, foram realizadas cinco reações, com os seguintes *primers*: fD1, rD1, 362f (região alvo 339-362), 786f (região alvo 764-786) e 1203f (região alvo 1179-1203) (MENNA et al., 2006) descritos na Tabela 1. O mesmo programa foi usado com todos os oligonucleotídeos iniciadores: desnaturação inicial a 95 °C por 1 minuto; 39 ciclos de desnaturação a 95 °C por 2 minutos, anelamento a 46 °C por 1 minuto, extensão a 72 °C por 1 minuto; seguidos de uma extensão final por 5 minutos, a 72 °C. As reações de sequenciamento para a região do ITS, também foram realizadas em microplaca utilizando o BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) seguindo as instruções do fabricante utilizando os *primers* FGPS 1490 e FGPS 132'. As reações de amplificação em microplacas foram sequenciadas utilizando o aparelho automático Sanger ABI 3130 XL (Applied Biosystems®) de acordo com as instruções do fabricante.

Tabela 1. Oligonucleotídeos iniciadores utilizados na amplificação da região do 16S rRNA e da região do espaçador transcrito interno ITS.

| Oligonucleotídeos | Sequências | Referência |
|-------------------|---|-------------------------|
| fD1 | 5'-CCGAATTCGTCGACAACAGAGTTTGGATCCTGGCTCAG- 3' | (WEISBURG et al., 1991) |
| 362f | 5' -CTCCTACGGGAGGCAGCAGTGGGG- 3' | (MENNA et al., 2006) |
| 786f | 5' -CGAAAGCGTGGGGAGCAAACAGG- 3' | (MENNA et al., 2006) |
| 1203f | 5' -GAGGTGGGGATGACGTCAAGTCCTC- 3' | (MENNA et al., 2006) |
| rD1 | 5' -CCCGGGATCCAAGCTTAAGGAGGTGATCCAGCC- 3' | (WEISBURG et al., 1991) |
| FGPS 1490 | 5'-TGCGGCTGGATCACCTCCTT-3' | (LAGUERRE et al., 1996) |
| FGPS 132' | 5'-CCGGGTTTCCCCATTTCGG- 3' | (LAGUERRE et al., 1996) |

4.1.5. Análise do sequenciamento do gene 16S rRNA e ITS

As sequências obtidas foram analisadas pelo pacote de programas Phred/Phrap/Consed (EWING et al., 1998; EWING; GREEN, 1998; GORDON; ABAJIAN; GREEN, 1998) com nível de exigência mínima de 400 bases para cada região (para cada *primer* utilizado) com qualidade Phred acima de 20.

Posteriormente, as sequências relacionadas ao gene 16S rRNA foram submetidas à consulta de similaridade de nucleotídeos, com sequências do banco de dados do 'Ribosomal Database Project' (RDP) pela ferramenta 'Classifier' com *Confidence threshold* de 95% e para ITS no banco do NCBI.

4.1.6. Análise filogenética do gene 16S rRNA e ITS

Para compreender melhor as relações evolutivas entre os isolados e outras linhagens relacionadas aos gêneros bacterianos *Pandoraea*, *Labrys* e *Chitinophaga*, as sequências do gene 16S dos isolados bacterianos deste estudo foram comparadas com outras sequências recuperadas do banco de dados do GenBank.

Todas as sequências referentes ao gene 16S rRNA e ITS dos isolados e das sequências recuperadas, foram alinhadas pelo MUSCLE v.3.8 e o melhor modelo de evolutivo foi estimado com jModelTest 2.1.7 (DARRIBA et al., 2012) com base nos critérios de informação de Critério de Informação Akaike. A inferência filogenética foi realizada com o programa Mrbayes v3.2.6 (RONQUIST et al., 2012). O resultado final da árvore filogenética foi visualizado e editado com programa FigTree versão 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

4.2. Extração de DNA total e sequenciamento da comunidade bacteriana CB10b

A comunidade CB10b foi cultivada em meio PGE líquido durante 18 horas, sob agitação constante a 150 rpm à 30°C. A partir da suspensão bacteriana obtida, foi transferida uma alíquota de 1000 µL para um microtubo de 1,5 mL que foi centrifugado (Eppendorf®, 5810 R) a 12000 xg por 20 minutos. Após a centrifugação o sobrenadante foi descartado e o precipitado foi usado para a extração de DNA. O DNA total da comunidade (CB10b) foi extraído utilizando kit da PROMEGA – Wizard® Genomic DNA Purification de acordo com as

recomendações do fabricante. A quantificação do DNA total da comunidade CB10b foi realizada por espectrofotometria no comprimento de onda 260/280 nm (NanoDrop® ND-1000 UV-Vis) e por Fluorometria Qubit® dsDNA BR (Life Technologies Europa, Naerum, Dinamarca). A qualidade do DNA foi avaliada em eletroforese em gel de agarose 0,8% (m/v) contendo brometo de etídeo (10 µg/mL). A biblioteca do DNA metagenômico (CB10b) foi preparada usando o kit da Nextera® XT DNA Library e posteriormente foi submetido ao sequenciamento em uma “lane” em HiSeq Flow Cell v4, com o HiSeq SBS Kit v4 e leituras pareadas de 2 x 100 bp na plataforma Illumina HiSeq 2500.

4.2.1. Avaliação da melhor estratégia de montagem para o conjunto de dados da comunidade bacteriana (CB10b)

Os conjuntos de leituras processados após passar por uma etapa de filtro de qualidade e retirada de adaptadores por meio do pacote de ferramentas do BBmap (<https://sourceforge.net/projects/bbmap/>) foram normalizados com BBnorm. Após a etapa de processamento dois conjuntos de dados seguiram para a etapa de montagem, um com os conjuntos de dados somente processados para a retirada de sequências de baixa qualidade e outro além de processado o conjunto também foi normalizado. Posteriormente os conjuntos de dados foram montados com os seguintes montadores idba_ud v1.1.1 (PENG et al., 2012), megahit v1.0.3-8 (LI et al., 2016), spades v3.10.1 (BANKEVICH et al., 2012), metaspades v3.10.1 (NURK et al., 2017) e pipeline automático de montagem e anotação Atlas, sendo as opções de montadores spades e megahit, apenas para os conjuntos de dados normalizados. Os resultados das montagens foram comparados utilizando o programa MetaQuast v4.4 (MIKHEENKO; SAVELIEV; GUREVICH, 2016).

4.2.2. Montagem, classificação taxonômica e perfil funcional da comunidade bacteriana (CB10b)

Os dados brutos gerados pelo sequenciamento foram processados através do programa Atlas v1.0.2 (WHITE III et al., 2017), analisando quanto à qualidade, sendo retirados eventuais segmentos de sequências contaminantes de regiões de baixa qualidade usando BBDuk que pertence ao pacote de

ferramentas do BBMap (<https://sourceforge.net/projects/bbmap/>). Os dados que passaram pela etapa de qualidade foram normalizados quanto à cobertura de leituras com base na frequência de k-mer formando subconjuntos de leituras usando a ferramenta BBNorm. Os subconjuntos com sequências de alta qualidade foram usados como entrada para o montador SPAdes v3.10.1 (BANKEVICH et al., 2012). Após a montagem, os contigs foram submetidos para a predição das ORFs (*Open Reading Frame*) pelo programa Prodigal (HYATT et al., 2010). As proteínas preditas foram anotadas usando alinhamento de proteína-proteína através do DIAMOND (BUCHFINK; XIE; HUSON, 2015), com filtro padrão > 200 para o *bitscore* e *e-value* de $< 1e^{-7}$ para os bancos do RefSeq (LEARY et al., 2016), EggNOG (HUERTA-CEPAS et al., 2016), dbCAN para famílias CAZy (YIN et al., 2012), ENZYME para comissão enzimática (EC) (BAIROCH, 2000) e COG (TATUSOV et al., 2003), implementados na plataforma do programa Atlas (WHITE III et al., 2017).

4.2.3. Avaliação do potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica

A anotação funcional de ORFs com domínios associados às enzimas de carboidratos ativos (CAZy) foi realizada por busca de similaridade de domínios relacionados às famílias do CAZy com base em modelos ocultos de Markov (HMMs) no dbCAN (<http://csbl.bmb.uga.edu/dbCAN>) (YIN et al., 2012). Foram aplicados os parâmetros de 30% de cobertura do HMM, *e-value* de $1e^{-5}$ para alinhamentos maiores que 80 aminoácidos e de $1e^{-3}$ para alinhamentos menores que 80 aminoácidos. As enzimas preditas com domínios para carboidratos ativos no metagenoma foram adicionalmente comparadas contra os bancos de dados não redundante do GenBank (NCBI), UNIProt (APWEILER, 2009) e o InterProscan (ZDOBNOV; APWEILER, 2001) considerando os parâmetros *e-value* $< 10^{-5}$ e um valor mínimo de 50% de identidade para inferir função às proteínas putativas.

4.2.4. Recuperação do genoma parcial a partir do metagenoma do CB10b

O arquivo de contigs gerados na montagem foi carregado no portal Web Busybee versão 2a7a1ac (2017-08-25) (<https://ccb-microbe.cs.uni->

saarland.de/busybee) para recuperação de genomas e análise de qualidade por meio do programa CheckM (v1.0.7) para avaliar o grau de integridade, contaminação e heterogeneidade dos conjuntos de dados recuperados. Para a análise no BusyBee foi habilitada apenas a anotação taxonômica e os outros parâmetros seguiu-se o padrão do programa. Após a separação dos conjuntos de clusters, estes foram filtrados e analisados para a recuperação do genoma com o maior índice de cobertura em relação a um genoma de referência. A comparação do genoma selecionado com outros genomas foi realizada por um alinhamento múltiplo utilizando o algoritmo Progressive Mauve (DARLING et al., 2004), os genomas selecionados para realizar a comparação foram recuperados do banco de dados do GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). O alinhamento dos genomas foi construído usando os parâmetros padrão do Mauve Aligner. As sequências mais próximas do organismo deste estudo foram carregadas no pacote do software JSpecies (<http://www.imedeia.uib.es/jspecies>) para realizar cálculos de comparação entre os genomas e para calcular a identidade média de nucleotídeos (ANI) com um limiar de 95% para delineamento de espécies (KONSTANTINIDIS; TIEDJE, 2004; RICHTER; ROSSELLO-MORA, 2009).

5. RESULTADOS E DISCUSSÃO

5.1. Caracterização do isolamento de bactérias da comunidade bacteriana CB10b

Através do isolamento por diluição seriada dos componentes da comunidade CB10b, obteve-se 6 isolados morfológicamente distintos e denominados respectivamente; CB10b_01, CB10b_02, CB10b_03, CB10b_04, CB10b_05 e CB10b_06.

A atividade de endoglucanases (CMCase) dos isolados e da comunidade bacteriana CB10b foi avaliada em meio CMC e Avicel (Figura 4). Foi possível identificar isolados bacterianos que possuem atividade de endoglucanase pela formação de halo de hidrólise ao redor das colônias em meio com CMC (Figura 4A). Para os isolados CB10b_01, CB10b_04 e para a comunidade CB10b que tiveram atividade em placa, foi realizada a medição do índice enzimático, onde estes isolados apresentaram índices de 1,71 cm, 1,87 cm e 0,67 cm respectivamente. Por outro lado, não foi detectada atividade enzimática em nenhum dos isolados e nem na comunidade bacteriana através do ensaio com meio contendo Avicel (Figura 4B).

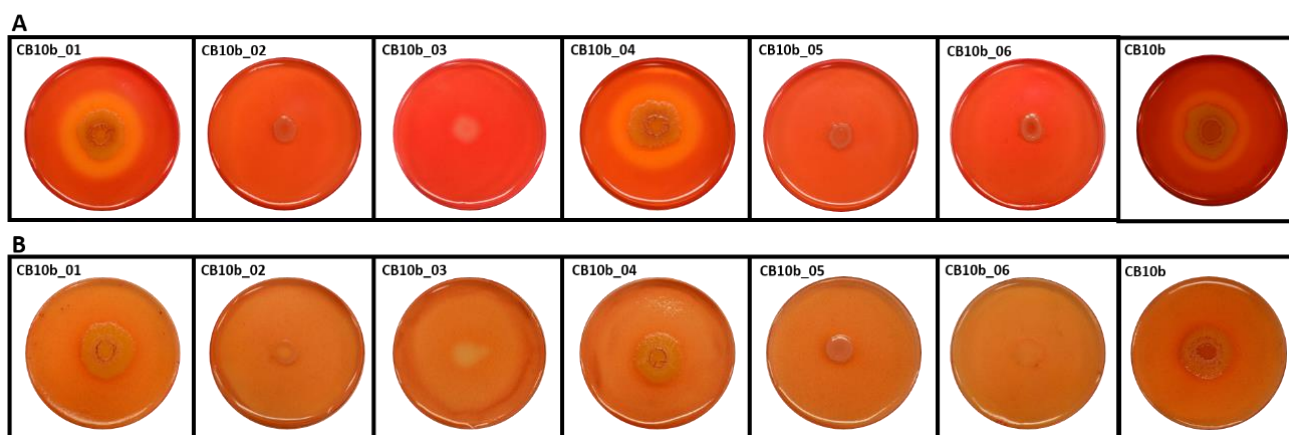


Figura 4. Teste para identificação de isolados produtores de celulases. Placas coradas com vermelho congo 1% após 5 dias de cultivo em BOD à 30°C. Teste com os 6 isolados e a comunidade CB10b em meio de cultura sólido contendo CMC 0,2% (A) e em meio de cultura sólido contendo AVICEL 0,2% (B).

A celulose microcristalina conhecida como avicel consiste em uma celulose purificada e parcialmente despolimerizada, obtida de polpa de material vegetal fibroso com ácidos minerais e tratada com alfa-celulose (tipo I_β) (THOORENS et al., 2014). A estrutura da celulose possui hidroxilas que estão

envolvidas em ligações de hidrogênio intra e intermolecular, que resultam em vários arranjos cristalinos ordenados ao longo de sua estrutura (PARK et al., 2010). Essa estrutura cristalina altamente organizada dificulta o acesso de enzimas celulolíticas, necessitando da ação de enzimas do tipo exoglucanases (celobiohidrolase I e celobiohidrolase II) que atuam sobre as extremidades redutoras e não redutoras da cadeia de celulose e módulos de ligação a carboidratos (CBMs), os quais auxiliam na penetração e ancoragem do módulo catalítico na fração cristalina do substrato (celulose), aumentando a atividade da enzima (ARANTES; SADDLER, 2010; GAO et al., 2013).

Já o composto carboximetilcelulose de sódio consiste em um polímero aniônico derivado da celulose quimicamente modificado com hidróxido de sódio e monocloroacetato de sódio, que o torna altamente solúvel em água e menos complexo que o avicel, favorecendo uma melhor interação entre substrato e enzimas, especialmente para as endoglucanases e beta-glicosidades.

5.1.1. Agrupamento filogenético dos isolados da comunidade CB10b

Por meio da análise de similaridade das sequências do gene 16S rRNA dos isolados utilizando o banco de dados do RDP pela ferramenta “*Classifier*” com um limite de 95% de confiança, foi possível identificar os isolados CB10b_01 e CB10b_04 como sendo do gênero *Chitinophaga*, o isolado CB10b_03 como sendo do gênero *Labrys* e os isolados CB10b_02, CB10b_05 e CB10b_06 como sendo do gênero *Pandoraea*.

Através da classificação taxonômica dos isolados relacionados ao gênero *Pandoraea*, os isolados CB10b_02, CB10b_05 e CB10b_06 se agruparam em um clado distinto de várias espécies de *Pandoraea* que juntos compartilham um ancestral comum com *Pandoraea terrae* (Figura 5).

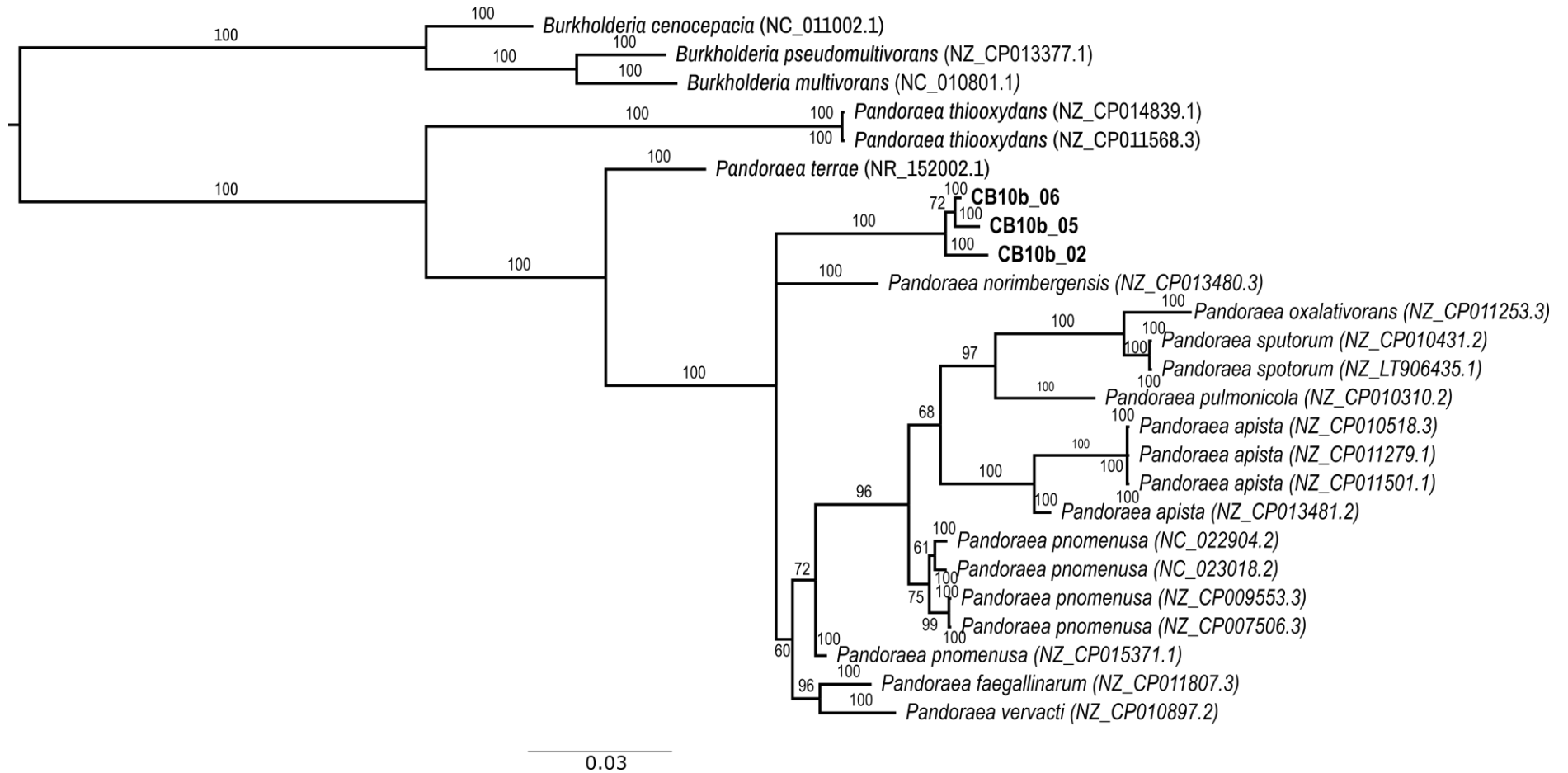


Figura 5. Árvore filogenética concatenada dos genes 16 rRNA e ITS dos isolados referentes ao gênero *Pandoraea*. As seqüências deste trabalho estão destacadas em negrito. A árvore foi construída usando o método de inferência Bayesiana com 1.000.000 gerações. Os números dos ramos indicam as probabilidades posteriores em porcentagem. Como grupo externo foram utilizadas seqüências correspondentes ao gênero *Burkholderia* que pertencem ao filo Proteobacteria.

O gene 16S do RNA ribossomal (rRNA) constitui-se de uma região conservada entre bactérias e a análise deste gene consiste em um método padrão estabelecido para identificação de gêneros bacterianos (AMANN et al., 1995; WOESE, 1987). Este método não permite a diferenciação entre subespécies, devido ao alto nível de similaridade do DNA entre as espécies. Entretanto, a região espaçadora intergênica 16S-23S consiste em uma região que ocorrem variações consideráveis entre espécies tanto em conteúdo da sequência, quanto em comprimento (tamanho) (GURTLER; STANISICH, 1996).

Sendo assim, com o propósito de distinguir os isolados relacionados a um mesmo gênero em espécies distintas, foi realizada uma análise adicional, baseada no sequenciamento da região espaçadora do transcrito interno dos isolados. A análise dos genes concatenados 16S rRNA e a região espaçadora entre os genes 16S-23S rRNA foi possível verificar que os isolados CB10b_02, CB10b_05 e CB10b_06 apresentam diferenças em relação as sequências recuperadas no banco de dado do GenBank pertencentes ao gênero *Pandoraea* (Figura 4).

Por meio da análise filogenética do gene 16S para o gênero *Labrys*, foi observado que o isolado deste estudo, apresentou-se agrupamento junto ao clado das espécies de *L. ginsengisoli*, *L. neptuniae* e *L. portucalensis* (Figura 6A). Enquanto para o gênero *Chitinophaga* os isolados CB10b_01 e CB10b_04, apresentaram agrupamento junto às espécies de *C. jiangningensis*, *C. dinghuensis* e *C. terrae* (Figura 6B). Segundo a análise filogenética do gene 16S rRNA estes isolados apresentam evidências de serem possíveis novas espécies dos gêneros *Labrys* e *Chitinophaga*. Porém não foi possível obter mais informações como foi realizado para o gênero *Pandoraea* que teve a região do ITS analisada. Estes gêneros não foram analisados quanto a região do ITS devido à indisponibilidade de informações no banco de dados do GenBank sobre as sequências depositadas relacionadas a estes dois gêneros.

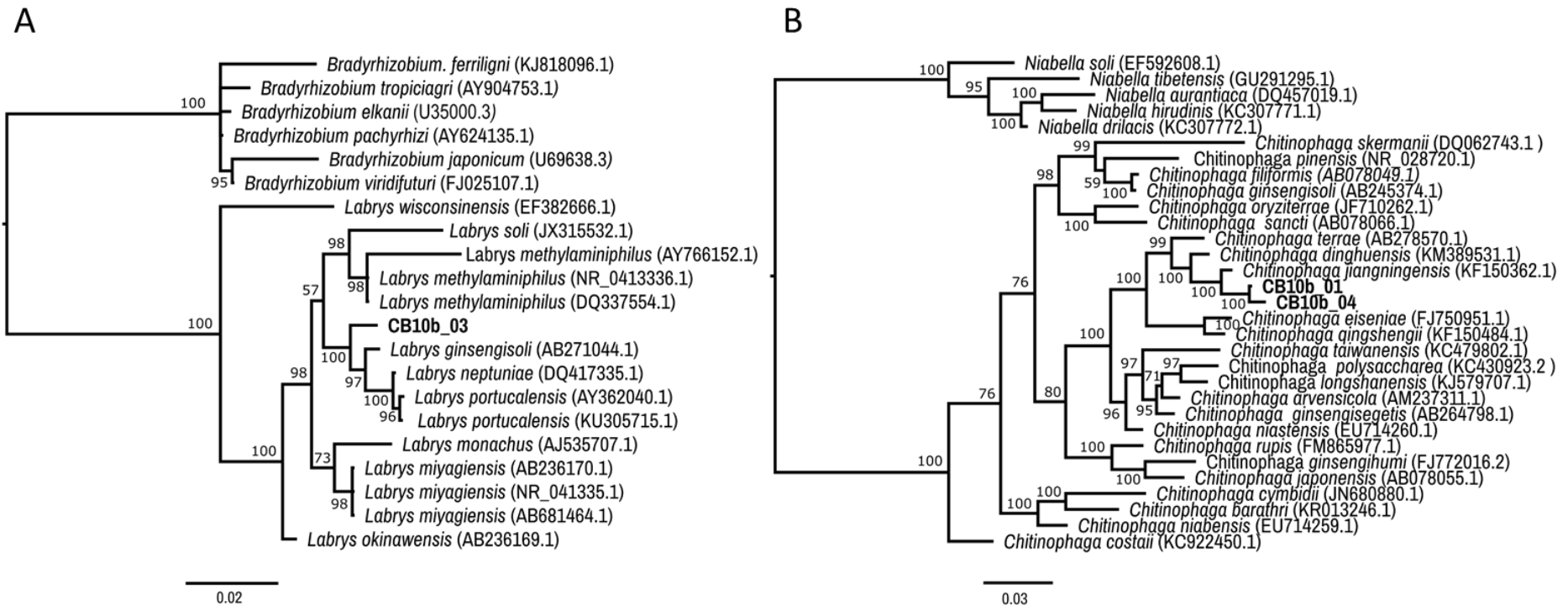


Figura 6. Árvore filogenética do gene 16 rRNA para os isolados *Chitinophaga* e *Labrys*. (A) Sequências relacionadas ao gênero *Labrys* e como grupo externo foram utilizadas sequências de *Bradyrhizobium spp.*, (B) sequências relacionadas ao gênero *Chitinophaga* e como grupo externo foram utilizadas sequências de *Niabella spp.* As árvores foram construídas usando o método de inferências bayesiana com 1.000.000 de gerações. Os números dos ramos indicam as probabilidades posteriores em porcentagem.

5.2. Avaliação da melhor estratégia de montagem para o conjunto de dados da comunidade bacteriana (CB10b)

Os resultados das montagens realizadas para os conjuntos de dados da comunidade bacteriana são mostrados na Figura 7. Considerando os parâmetros analisados pelo programa MetaQuast, os melhores resultados foram apresentados pelas montagens idba_ud com dados normalizados, megahit com dados normalizados, megahit com dados normalizados e opção de meta-sensitive, spades atlas e megahit atlas.

| | IDBA_UD | MEGAHIT | MEGAHIT (meta-sensitive) | METASPADES | SPADES | IDBA_UD | MEGAHIT | MEGAHIT (meta-sensitive) | METASPADES | SPADES | ATLAS (Spades) | ATLAS (Megahit) |
|--|------------|------------|--------------------------|------------|----------|------------|------------|--------------------------|------------|-----------|----------------|-----------------|
| Estatísticas dos genomas | | | | | | | | | | | | |
| Fração genômica (%) | 0,417 | 0,44 | 0,442 | 0,104 | 0,034 | 0,407 | 0,408 | 0,408 | 0,852 | 0,471 | 0,449 | 0,419 |
| Taxa de duplicação | 1,033 | 1,036 | 1,037 | 1,374 | 1,4 | 1,039 | 1,036 | 1,04 | 1,198 | 1,22 | 1,05 | 1,055 |
| Maior alinhamento | 21.403 | 21.403 | 21.403 | 1.064 | 891 | 21.403 | 21.403 | 21.403 | 3649 | 1.331 | 21.403 | 21.403 |
| Tamanho total dos alinhamentos | 112.597 | 113.014 | 114.268 | 37.592 | 12.970 | 109.557 | 109.556 | 109.888 | 301.191 | 148.921 | 115.742 | 115.624 |
| <i>Burkholderia cepacia</i> | 422 | 590 | 762 | 164 | 166 | 341 | 422 | 422 | 2.273 | 426 | 797 | 505 |
| <i>Chitinophaga jiangningensis</i> | 14.426 | 14.272 | 14.709 | 7.734 | 4.090 | 14.164 | 14.476 | 14.375 | 52.334 | 16.471 | 14.368 | 14.460 |
| <i>Chitinophaga terrae Kim and Jung 2007</i> | 5.771 | 6.579 | 6.579 | - | 211 | 6.099 | 6.044 | 6.329 | 5.697 | 2.167 | 7.066 | 7.073 |
| <i>Pandoraea pnomenusa</i> | 46.886 | 46.492 | 47.267 | 4.222 | 527 | 46.966 | 46.544 | 46.766 | 46.152 | 26.828 | 48.407 | 48.032 |
| <i>Pandoraea pulmonicola</i> | 64.805 | 64.693 | 64.563 | 25.547 | 7.644 | 61.753 | 61.844 | 61.844 | 206.133 | 112.672 | 65.021 | 65.141 |
| <i>Pandoraea sputorum</i> | 6.783 | 6.459 | 6.459 | 2.157 | 495 | 6.613 | 6.695 | 6.695 | 11.494 | 5.508 | 6.874 | 6.830 |
| # Montagens incorretas | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Tamanho dos contigs montados incorretamente | 1.352 | 1.352 | 1.352 | 0 | 0 | 1.352 | 1.352 | 1.352 | 892 | 1.290 | 1.352 | 1.352 |
| # mismatches 100 kbp | 2.476,78 | 2.387,97 | 2.402,53 | 3.399,62 | 3.463,24 | 2.434,80 | 2.433,85 | 2.436,88 | 3.120,31 | 2.710,60 | 2.401,45 | 2.488,64 |
| # substituição de base a cada 100 kbp | 229,51 | 222,83 | 225,37 | 45,19 | 38,06 | 224,8 | 228,15 | 225,33 | 95,7 | 57,29 | 249,96 | 227,78 |
| # N's por 100 kbp | 0 | 0 | 0 | 683,41 | 77,6 | 0 | 0 | 0 | 314,07 | 17,74 | 0 | 0 |
| Medidas de distribuição | | | | | | | | | | | | |
| Número de contigs | 3.546 | 3.353 | 3.362 | 2.098 | 666 | 3.855 | 4.173 | 4.146 | 9.664 | 4.729 | 3.316 | 3.314 |
| Maior contig | 826.303 | 610.084 | 826.259 | 2.403 | 1.355 | 826.375 | 469.036 | 469.036 | 8.856 | 1.944 | 610.197 | 610.106 |
| Tamanho total | 14.570.471 | 14.220.265 | 14.238.566 | 1.553.969 | 399.488 | 16.937.785 | 16.752.065 | 16.792.455 | 10.500.713 | 3.044.175 | 14.210.191 | 14.211.574 |
| Tamanho total (>= 1000 bp) | 12.939.649 | 12.680.694 | 12.696.651 | 312.540 | 5.515 | 15.795.031 | 15.365.612 | 15.432.976 | 6.396.674 | 179.954 | 12.688.332 | 12.688.332 |
| Tamanho total (>= 10000 bp) | 11.059.806 | 11.020.078 | 11.021.049 | 0 | 0 | 11.146.486 | 11.096.581 | 11.109.401 | 0 | 0 | 11.047.774 | 11.038.173 |
| Tamanho total (>= 50000 bp) | 10.308.081 | 9.995.317 | 99.55.698 | 0 | 0 | 10.504.326 | 10.330.272 | 10.360.693 | 0 | 0 | 10.023.901 | 10.130.058 |

Figura 7. Avaliação da qualidade das diferentes montagens do metagenoma com o programa MetaQuast. As montagens agrupadas nas colunas em preto (■) indicam os conjuntos de dados normalizados, em cinza escuro (■) indicam as montagens utilizando os conjuntos de dados não normalizados e em cinza claro (■) indicam os conjuntos de dados normalizados utilizando a ferramenta Atlas.

Analisando os resultados obtidos pelo MetaQuast a montagem com o SPADES através da ferramenta Atlas foi escolhida para seguir as análises metagenômicas da comunidade. Considerando que o montador K consiste em uma ferramenta automática de montagem e anotação de metagenomas, publicado em março de 2017, esta ferramenta facilita a manipulação dos dados metagenômicos.

5.2.1. Montagem do CB10b e predição funcional

O sequenciamento do DNA total da comunidade bacteriana CB10b gerou aproximadamente 54 milhões de leituras. Após o processamento das sequências e normalização, estas reduziram aproximadamente 17 milhões de leituras paired-ends, distribuídos em 4.627 *contigs* com tamanho médio de 3.192 pb e N50 de 115.752 pb (Tabela 2).

Tabela 2. Descrição da montagem do metagenoma da comunidade bacteriana (CB10b) e predição funcional.

| Metagenoma Comunidade Bacteriana CB10b | | | |
|--|------------|--|--------|
| Montagem | | Predição funcional | |
| Total de leituras | 53.952.118 | ORFs preditas | 16.340 |
| Leituras normalizadas | 17.347.294 | ORFs anotadas | 10.213 |
| Total de contigs | 4.627 | ORFs hipotéticas ou desconhecidas | 3.858 |
| Tamanho total dos <i>contigs</i> (pb) | 14.769.957 | ORFs anotadas RefSeq | 10.213 |
| Maior <i>contigs</i> (pb) | 610.106 | ORFs anotadas Enzyme | 1.561 |
| Menor <i>contigs</i> (pb) | 252 | ORFs anotadas EggNog | 2.954 |
| <i>Contigs</i> > 1 kb | 1.110 | ORFs anotadas COG | 8.191 |
| <i>Contigs</i> > 10 kb | 104 | ORFs anotadas CAZy | 399 |
| <i>Contigs</i> >100 kb | 42 | ORFs anotadas Uniprot | 11.253 |
| Tamanho médio dos <i>contigs</i> (pb) | 3.192 | | |
| N50 contigs (pb) | 115.752 | | |

Através da predição funcional um total de 16.340 ORFs foram preditas e destas 10.213 ORFs apresentaram uma função atribuída com base nos bancos de dados de referência utilizados neste estudo (Tabela 2), enquanto que 3.858 apresentaram função hipotética ou desconhecida.

5.2.2. Composição taxonômica e funcional da comunidade bacteriana CB10b

A análise taxonômica da comunidade bacteriana CB10b a partir do metagenoma apresentou dois filos mais abundantes, Proteobacteria com 72,08 % e Bacteroidetes com 26,63 % (Figura 8). Os filos que tiveram abundância relativa com valores inferiores a 1% foram agrupados em uma categoria denominada como “outros” apresentando abundância relativa de 0,071 %, enquanto que para as sequências que não tiveram classificação em nível de filo tiveram abundância relativa de 1,22 % (Figura 8). Estudos de Kanokratana *et. al*, (2013) e Mhuantong *et. al*, (2015) também relacionados à análise de comunidades microbianas isoladas da pilha de bagaço de cana-de-açúcar apontaram elevada abundância principalmente dos filos Proteobacteria e Bacteroidetes que estão associados à degradação do material lignocelulósico do bagaço (KANOKRATANA *et al.*, 2013; MHUANTONG *et al.*, 2015).

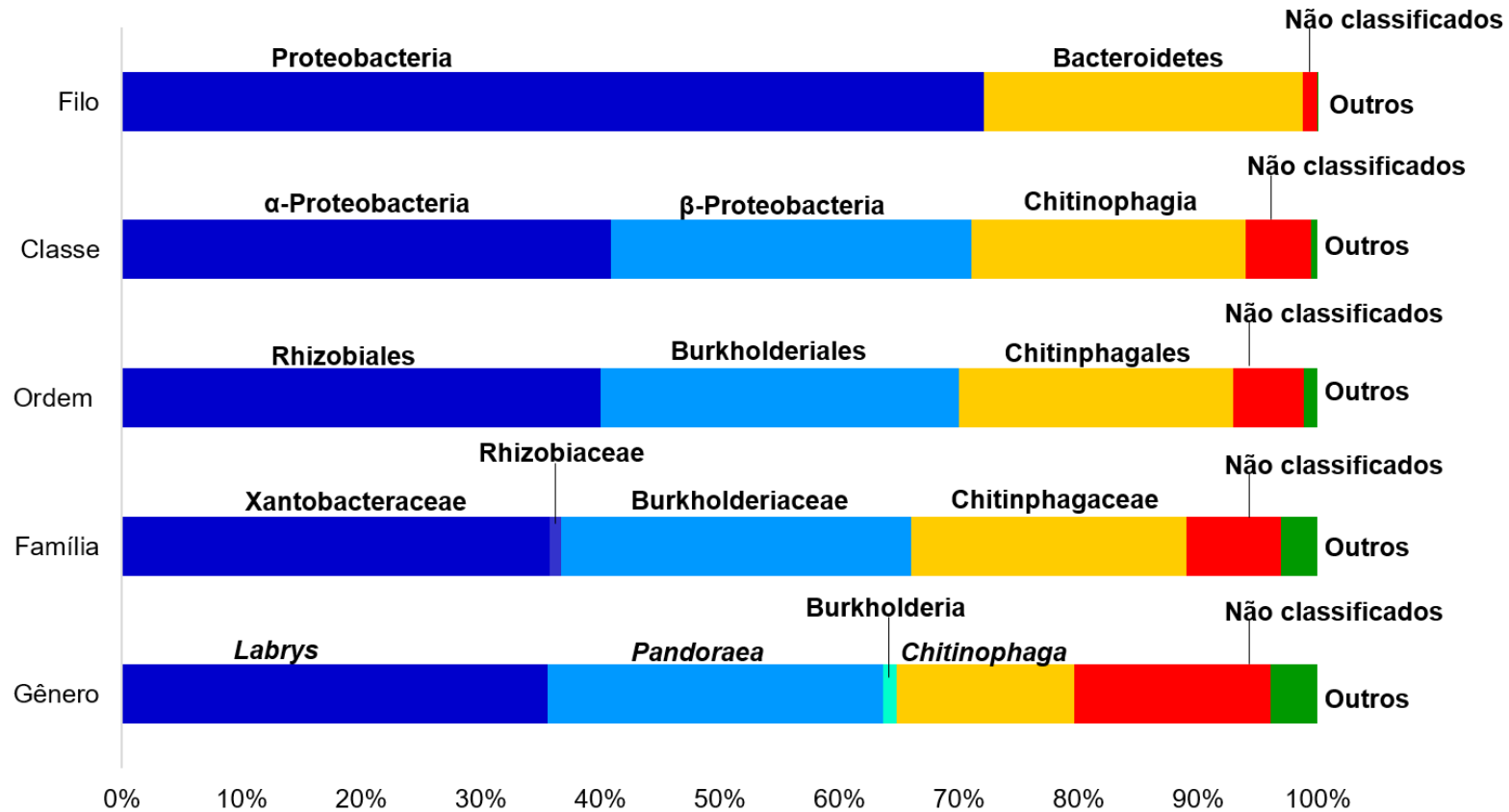


Figura 8. Distribuição taxonômica da comunidade bacteriana CB10b. Para todas as categorias indicadas em vermelho as sequências não obtiveram uma classificação na categoria analisada; categorias representadas em verde indicam sequências que tiveram abundância relativa inferior a 1%.

A distribuição taxonômica em nível de classe foi de 40,91% para α -Proteobacteria; 30,12% para β -Proteobacteria, 22,94% para Chitinophagia; 5,49% de sequências não classificadas e 0,532% para a categoria outros.

Dentre estas classes, as distribuições taxonômicas em nível de ordem foram de 40,02% para Rhizobiales, 29,98% para Burkholderiales, 22,95% para Chitinophagales, 5,92% para sequências que não tiveram classificação em nível de ordem e 3,07% para categoria “outros” (Figura 8). Em nível de famílias, a comunidade está distribuída em Xantobacteraceae, com 35,72%, e Rhizobiaceae, com 1%, que pertencem a ordem Rhizobiales, seguidas das famílias Burkholderiaceae, com 29,31%, pertencente à ordem Burkholderiales, e Chitinophagaceae, com 22,95%, pertencendo à ordem Chitinophagales, neste nível, a categoria de sequências não classificadas corresponde a 7,96%, e de outros, 3,07%. Com base na classificação taxonômica em nível de gênero, a comunidade bacteriana é representada pelos gêneros *Labrys*, com 35,60%, *Pandoraea*, com 28,06% e *Chitinophaga*, com 14,92%. As sequências não classificadas em nível de gênero apresentaram abundância relativa de 16,40% e as sequências classificadas na categoria outras apresentaram 3,94% neste nível (Figura 8).

Alguns dos estudos associados ao gênero bacteriano *Labrys* relatam o potencial deste gênero para aplicação biotecnológica em processos de biorremediação de compostos como pentacloronitrobenzeno e antibióticos associados ao grupo das fluoroquinolonas (ofloxacina, norfloxacina e ciprofloxacina) (AMORIM et al., 2014; LI et al., 2011). A primeira identificação deste gênero como bactéria celulolítica foi realizada em um estudo que isolou bactérias do intestino de larvas de *Holotrichia paralela* (HUANG; SHENG; ZHANG, 2012).

Os gêneros *Pandoraea* e *Burkholderia*, pertencentes à classe β -Proteobacteria, são apontadas como bactérias capazes de atuar na degradação de lignina. A estirpe *Pandoraea* sp. B-6, isolada de lâminas de bambu, apresentou capacidade de degradar a lignina kraft por meio da lignina peroxidase (SHI et al., 2013). Um estudo associado à mineração de genes degradadores de hemicelulose e lignina em um biorreator anaeróbio alimentado com diferentes resíduos (palha de arroz, resíduos vegetais e resíduos alimentares) identificou

os gêneros *Pandoraea* e *Burkholderia* que apresentaram elevada abundância de genes associados à degradação de lignina (PANDIT et al., 2016).

O gênero *Chitinophaga* está principalmente associado às comunidades microbianas de ambientes agrícolas e florestais, onde a biomassa vegetal e fúngica é degradada (LARSBRINK et al., 2017), sendo intensamente relacionado com a decomposição de quitina (como sugerido pela denominação do gênero); entretanto, este também apresenta atividade celulolítica (MCKEE; BRUMER, 2015; PROENÇA et al., 2017; SANGKHOBOL; SKERMAN, 1981).

A composição funcional da comunidade bacteriana CB10b com base nas categorias funcionais COG (*Clusters Orthologous Groups*), estão distribuídas principalmente entre as categorias de “metabolismo e transporte de aminoácidos” (categoria COG [E]), representando categoria funcional com a maior abundância relativa no metagenoma com 12,3% das ORFs; “metabolismo e transporte de carboidratos” (categoria COG [G]) com 9,20%; “transcrição” (categoria COG [K]) com 8,60%; e “produção e conversão de energia” (categoria COG [C]) abundância relativa de 6,24% das ORFs. As categorias relacionadas à “apenas predição funcional” (categoria COG [R]) com 7,51%, “função não conhecida” (categoria COG [S]) com 4,73% e “sem atribuições” para o COG (categoria COG [X]) com 0,84%, apresentam potencial para a identificação de possíveis novos genes (Figura 9).

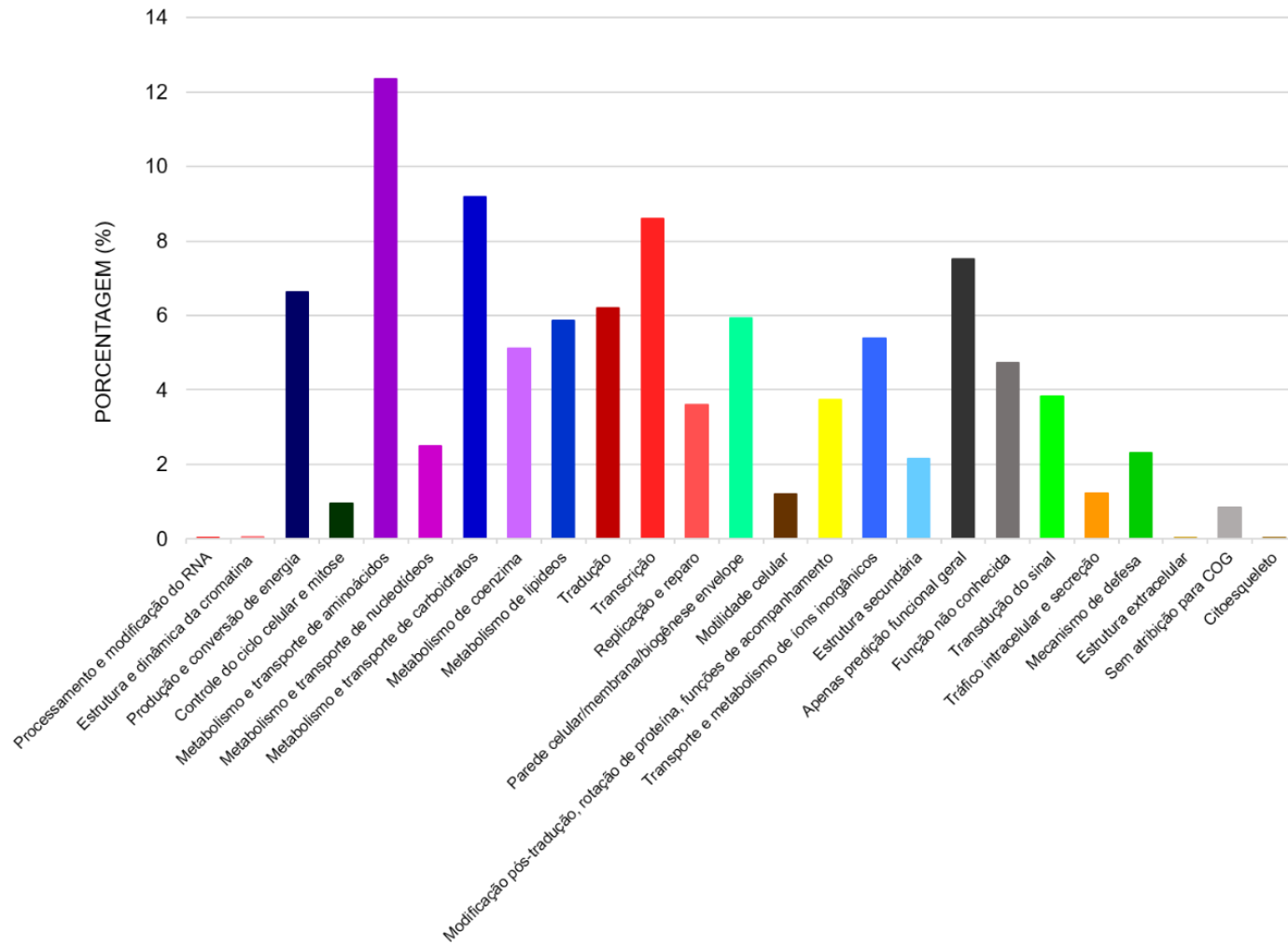


Figura 9. Composição funcional da comunidade bacteriana CB10b com base no banco de dados do Clustes Orthologous Groups (COG).

5.2.3. Potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica

A anotação funcional das ORFs com domínios associados ao CAZy realizada pelo programa Atlas excluiu algumas ORFs com domínios menores, porém foi possível identificar novas ORFs utilizando abordagem pela plataforma online do dbCAN.

As ORFs preditas foram comparadas com enzimas do banco de dados do CAZy, apresentando 624 ORFs associadas a enzimas ativas para carboidratos. Em relação às classes funcionais, estas foram distribuídas em glicosil hidrolases (GHs) com 37,98% (237 ORFs); glicosil transferases (GTs) com 21,63% (135 ORFs); polissacarídeo liases (PLs) com 3,18% (20 ORFs); carboidrato esterases (CEs) com 17,00% (106 ORFs); atividades auxiliares (AAs) com 4,81% (30 ORFs) e módulos de ligação a carboidratos (CBMs) com 15,38% (96 ORFs). Estas classes de carboidrases foram representadas principalmente pela família da Chitinophagacea com 46,80% das ORFs distribuídas para o gênero *Chitinophaga* e não classificados desta família, sendo o principal táxon representando no metagenoma do CB10b. Além disso, os gêneros *Pandoraea* da família Burkholderiaceae e *Labrys* da família Xanthobacteraceae apresentaram respectivamente 14,42 e 12,66 % das classes destas enzimas (Figura 10).

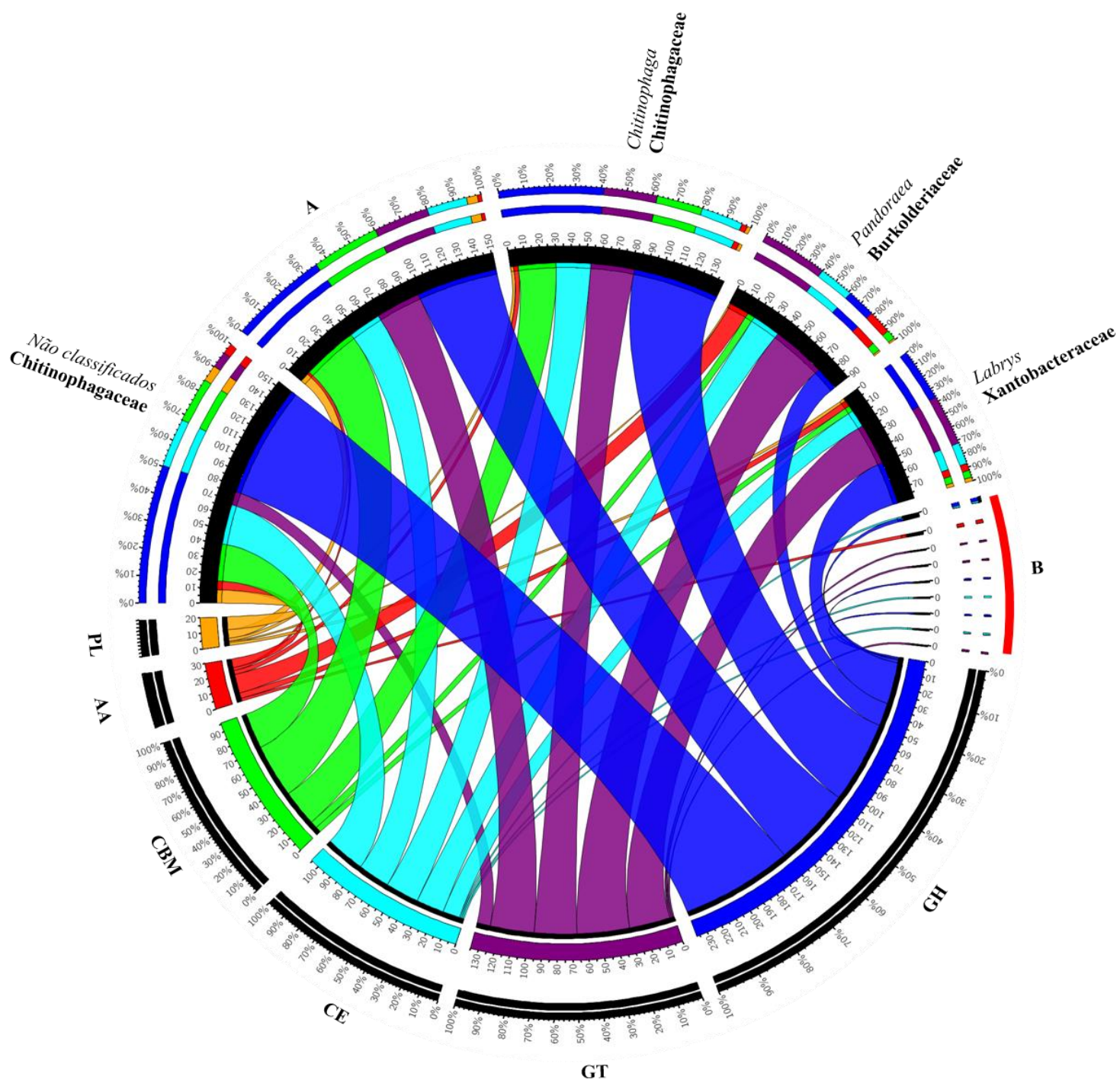


Figura 10. Distribuição taxonômica das classes de carboidrases associadas ao banco de dados do CAZY na comunidade bacteriana CB10b. (A) ORFs não classificadas, (B) gêneros bacterianos *Burkholderia*, *Methylobacterium*, *Microvirga*, *Phyllobacterium*, *Pseudomonas*, *Agrobacterium*, *Rhizobium*, *Sinorhizobium* e *Inquillinus* (representados, sentido horário). A largura das barras de cada espécie ou família bacteriana e família enzimática funcional indica sua abundância relativa no metagenoma.

Um total de 12,30% (77 sequências) ORFs anotadas para estas classes de enzimas foram recuperadas e classificadas como carboidrases, utilizando a abordagem de anotação através da plataforma do dbCAN, mas sem atribuição taxonômica. Tal resultado se deve ao fato de que o programa Atlas executa um filtro com *cut-off* padrão para sequências maiores que 200 aminoácidos para anotação funcional e taxonômica com relação aos bancos de dados utilizados pelo programa.

A análise do perfil funcional da comunidade bacteriana CB10b com base no banco de dados do CAZy classificou as ORFs em diferentes classes de enzimas, distribuídas entre glicosil hidrolases, glicosil transferases, carboidratos esterases, polissacarídeos liases, atividades auxiliares e módulos de ligação à carboidratos (Figura 11).

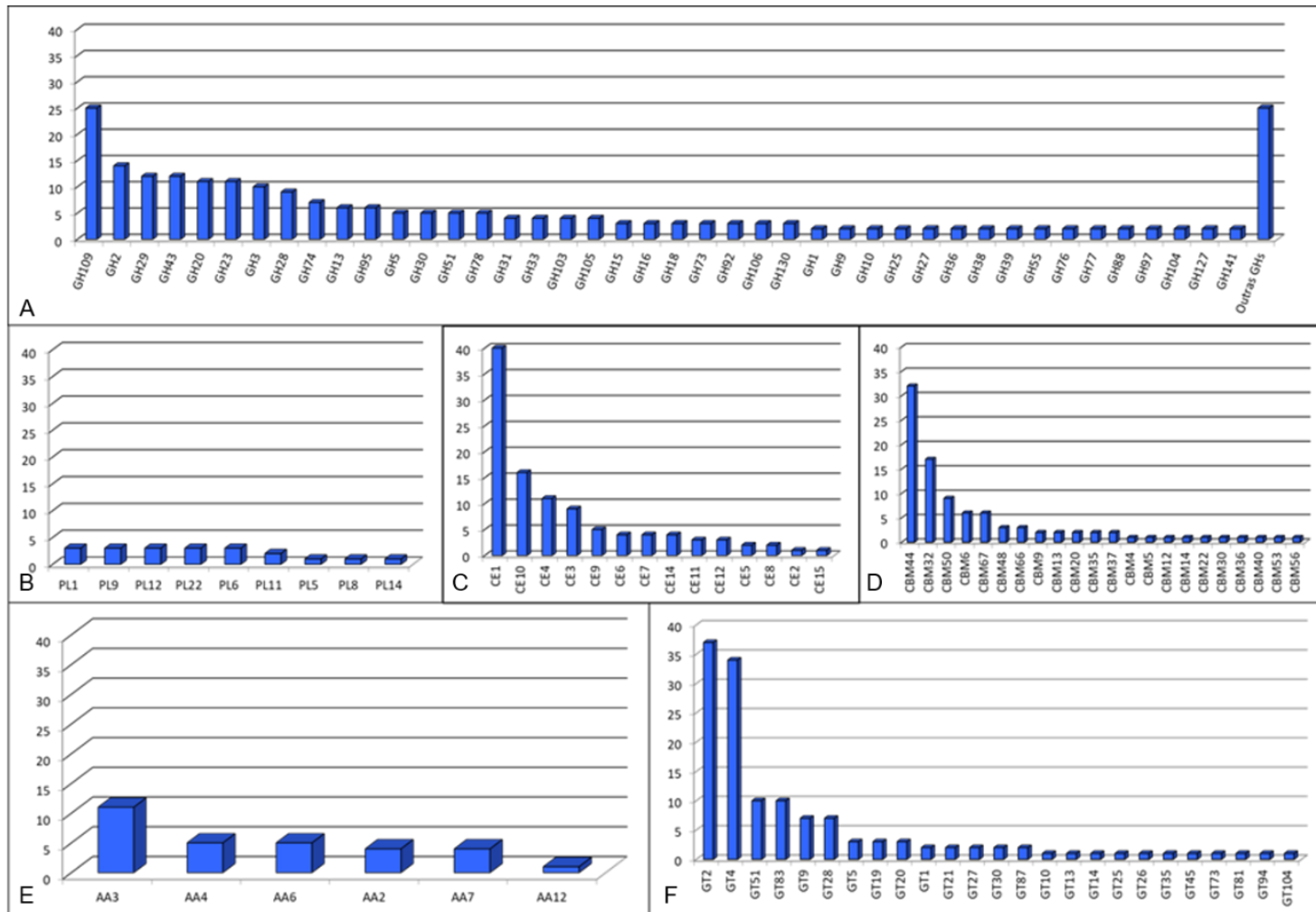


Figura 11. Distribuição das classes de enzimas associadas às carboidrases. A: Glicosil Hidrolases, B: Polissacarídeo liases, C: Caboidratos Esterases, D: Módulos de Ligação a Carboidratos, E; Atividades Auxiliares e F: Glicosil Transferases identificadas na comunidade bacteriana CB10b.

A classe das enzimas glicosil hidrolases constitui a classe mais abundante na comunidade bacteriana apresentando 68 famílias diferentes. São demonstradas 43 famílias e uma classe denominada como outras GHs, que representam as ORFs únicas de cada família, distribuídas em GH11, GH19, GH24, GH35, GH37, GH42, GH67, GH82, GH89, GH99, GH102, GH108, GH114, GH115, GH116, GH117, GH123, GH125, GH137, GH138, GH139, GH140, GH142, GH143, GH144 (Figura 11A).

A classe das GHs é mais abundante e tem sido reportada em vários estudos associados à degradação de biomassa (KANOKRATANA et al., 2013; MHUANTONG et al., 2015; MONTELLA et al., 2017) e está associada principalmente a desestruturação dos polímeros de celulose e hemicelulose.

As famílias GH3 e GH9 estão principalmente associadas às enzimas beta-glucosidases que também foram encontradas em outras famílias de “glicosil hidrolases” (GH1, GH16, GH116 e GH144).

A família GH5 (predita como endo-1,4-beta-glucanase) atua randomicamente em ligações internas da cadeia de celulose e a GH15 (glucan 1,4- α -glucosidase) atuam nas extremidades não redutora da cadeia de celulose, liberando unidades de β -D-glucose. As GH1, GH3, GH16 e GH116 (β -glucosidase) atuam sobre as moléculas de celobiose e de celbio-oligossacarídeos, para a liberação final de monômeros de glicose e outros oligossacarídeos.

A estrutura da hemicelulose contém uma variação maior de carboidratos, que demanda uma maior amplitude de enzimas de endo-ação (MHUANTONG et al., 2015). As GH5, GH10 e GH11 (endo-1,4- β -xilase) atuam na hidrólise de ligações internas da cadeia principal de xilana, o principal polissacarídeo da hemicelulose. Além destas enzimas, a xilana é degradada por β -xilosidases de múltiplas famílias incluindo GH31, GH39 e GH43, que atuam nas cadeias laterais liberando unidades de xilose.

As GH2 (β -galactosidase), GH27, GH36 (α -galactosidase) e GH92 (1,2- α -mannosidase), por sua vez, atuam nas frações de galactano e manose.

Também foram encontradas as famílias GH78 e GH106 que são enzimas associadas à hidrólise de resíduos de L-ramnose. Já a GH105 atua na decomposição das ramificações da ramnogalacturano. Ambas estão relacionadas à degradação da pectina, assim como a GH51 (α -L-

arabinofuranosidase) que pode atuar tanto na hidrólise de resíduos de α -L-arabinofuranosídeos de terminais não-redutores de hemicelulose quanto na fração de pectina.

Outra classe identificada no metagenoma foi a de “polissacarídeo liases” (PLs), que promovem a quebra de polissacarídeos que contenham ácido urônico por eliminação β , produzindo um açúcar 4,5-insaturado na nova extremidade não redutora do produto (GARRON; CYGLER, 2014). As polissacarídeo liases encontradas neste estudo são representadas por 9 famílias (Figura 11B). Considerando as atividades principais em cada família, as PL1 e PL9 (pectato liases) provavelmente estão associadas a quebra eliminativa de ligações 1,4- α -D-galacturonanos, enquanto as PL11 (ramnogalacturano-endo-liases) estão relacionados a quebras de ligações 1,4 de α -L-ramnopiranosil na cadeia de ácido galacturônico, as quais estão relacionadas a decomposição da estrutura da pectina.

A classe das “carboidrato esterases” engloba 14 famílias (Figura 11C), as famílias CE1, CE3, CE4 e CE10 (carboxil esterase, acetil esterase, aril esterase e quitina esterase, respectivamente) representam as principais atividades preditas na classe de carboidrato esterases. Enquanto as famílias CE8, CE11 e CE12 (pectina esterase, N-acetilglucosamina deacetilase e ramnogalacturonano acetilesterase) foram preditas em menor abundância. Alguns componentes não celulósicos da parede celular possuem grupos O-acetil que dificultam a hidrólise dos polímeros em seus respectivos monômeros (PAWAR et al., 2013), o que significa que classes de carboidrase estereases preditas estão relativamente associadas a deacetilação dos polímeros associados, principalmente a pectina.

Os módulos de ligação a carboidratos estão distribuídos entre 22 famílias (Figura 11D). Este grupo de proteínas não catalíticas promove associação das enzimas com o substrato (BORASTON et al., 2004). O módulo de ligação CBM32 está associado as GH5 (endoglucanase), GH106 (α -L-ramnosidase) e outras com função não identificada (GH20 e GH29). Também foram identificados outros módulos, como CBM48 que está associado a uma GH13 (1,4- α -glucano enzima de ramificação), CBM67 associado a GH78 (α -L-ramnosidase), CBM6 associado a GH43 (β -xilosidase) e CBM6 ligado a uma PL6 (pectato liase).

A classe de enzimas classificada como “atividades auxiliares”, por sua vez, é representada pelas famílias de enzimas ligninolíticas de polissacarídeos e mono-oxigenases (LEVASSEUR et al., 2013). Estas enzimas auxiliam principalmente as GHs na decomposição do material lignocelulósico. Foram identificadas na comunidade CB10b 6 famílias de atividades auxiliares (Figura 11E), no entanto apenas a família AA3 obteve uma predição funcional. A família AA3 pertence às oxidoreduções de glicose-metanol-colina (GMC). Para esta família foram preditas as subfamílias AA3_2 (aril álcool oxidase e glicose 1-oxidase) e AA3_3 (álcool oxidase).

Por fim, a classe das “glicosil transferases” está relacionada a transferência de regiões estereoespecíficas de vários doadores de açúcar para uma variedade de biomoléculas importantes, incluindo glicanos, lipídios, peptídeos e pequenas moléculas (CHANG et al., 2011). Esta classe está mais relacionada à biossíntese de carboidratos do que a sua degradação. As famílias GT2, GT4 e GT51 são as famílias mais abundantes encontradas no metagenoma (Figura 11F).

5.2.4. Recuperação parcial de genomas

A recuperação de conjuntos de clusters executado pelo programa BusyBee obteve a separação de três conjuntos de sequências (clusters) denominados como “bins” (Figura 12).

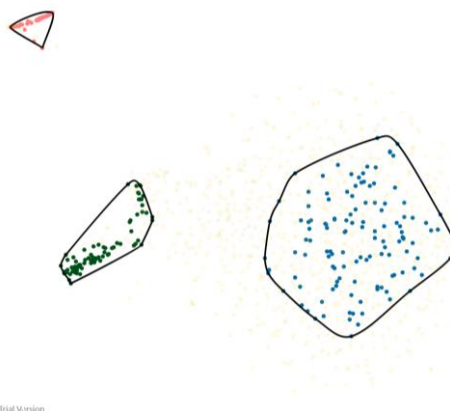


Figura 12. Distribuição taxonômica dos clusters separados pelo programa BusyBee. Azul representa o conjunto de sequências (bin 1), verde representa o conjunto de sequências (bin 2) e rosa representa o conjunto de sequências (bin 3).

Os índices de integridade e contaminação pelo programa CheckM, realizados para os conjuntos de sequências recuperados através do programa BusyBee, foram respectivamente 31,53% e 0% para *bin 1 (Labrys)*, 55,50% e 5,41% para o *bin 2 (Pandoraeea)* e 95,50% e 0% para o *bin 3 (Chitinophaga)*. Através da análise dos conjuntos de sequências no metagenoma CB10b foi possível recuperar genomas parciais para os gêneros, *Labrys*, *Pandoraeea* e *Chitinophaga*, onde essa *Chitinophaga* foi identificada por KISHI *et. al*, (2017). No presente trabalho descrevemos o genoma parcial da *Pandoraeea sp.* da classe de β -Proteobacteria, pelo fato de que foi o segundo genoma que apresentou maior índice de integridade, com um índice de contaminação de 5,41 %. Este gênero mediante a análise de conjuntos de clusters apresentou 4.840.980 pb, distribuídos em 87 contigs e com um total de 3.827 ORFs. Mediante ao alinhamento progressivo com o programa Mauve, foi possível identificar grande similaridade entre o genoma deste estudo com um genoma de uma *Pandoraeea pulmonicola* (Figura 13).

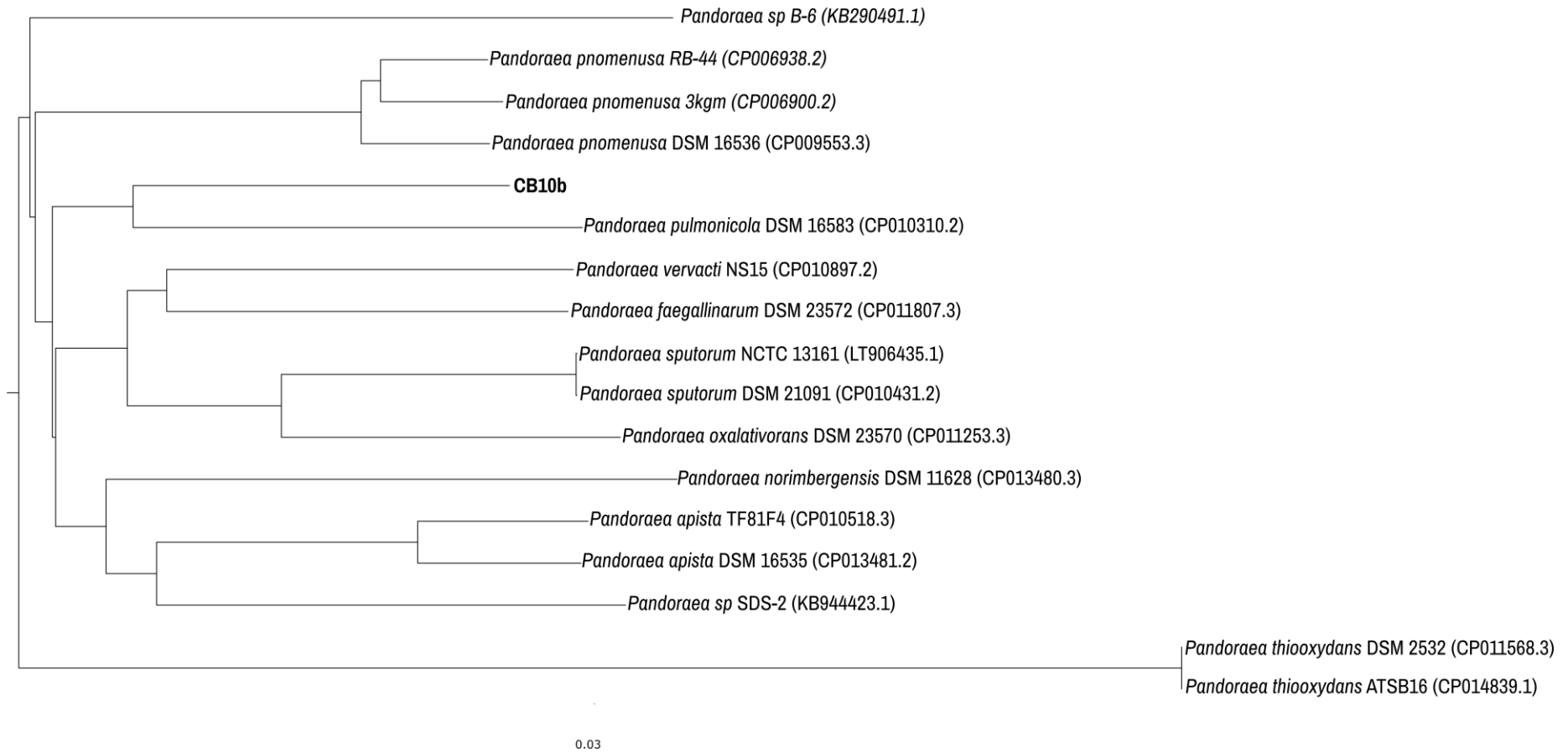


Figura 13. Similaridade entre genomas de bactérias do gênero *Pandoraea* alinhados pelo programa Mauve mostrando as similaridades entre os genomas. A árvore foi calculada usando o método de Neighbor-Joining pelo programa Mauve e visualizada pelo programa FigTree. A barra indica o número de substituições por posição de nucleotídeo.

A análise média de identidade de nucleotídeos (ANI) de genes conservados e compartilhados, com suporte do coeficiente de correlação de frequência tetranucleotídica (TETRA), tem sido reportada como método confiável e representativo substituto para hibridização DNA-DNA (DDH) (KISHI et al., 2017; SCORTICHINI et al., 2013). Um valor de ANI de 95% \pm 0,5% de identidade corresponde a 70% de DDH (GORIS et al., 2007), valor geralmente recomendado para delimitar espécies.

Os resultados comparativos entre os genomas demonstram, que a *Pandoraea pulmonicola* é mais similar à *Pandoraea sp.* CB10b (Tabela 3). Porém o genoma parcial da *P. sp.* CB10b com 82% de cobertura em relação ao genoma da *P. pulmonicola*, pelos parâmetros da análise de ANI apresentou-se no limite inferior para ser considerado da mesma espécie.

Tabela 3. Comparação entre o genoma de uma *Pandoraea pulmonicola* com diferentes espécies de *Pandoraea* deste estudo e outras espécies e *Chitinophaga pinensis* como controle para comparação.

| Organismo | Genoma [pb] | GC[%] | ANIb [%] | ANIm [%] | Tetra [%] |
|-----------------------------------|-------------|-------|----------|----------|-----------|
| <i>P. sp</i> CB10b | 4.840.980 | 65,9 | 87,72 | 89,62 | 98,37 |
| <i>P. pnomenusa</i> 3KGM | 5.429.298 | 64,7 | 83,44 | 87,18 | 98,26 |
| <i>P. pnomenusa</i> DSM16536 | 5.389.285 | 64,9 | 83,42 | 87,14 | 97,95 |
| <i>P. sp</i> SD6-2 | 5.772.015 | 62,5 | 83,33 | 86,92 | 97,41 |
| <i>P. apista</i> DSM16535 | 5.493.967 | 62,7 | 83,26 | 86,88 | 96,41 |
| <i>P. faegallinarum</i> DSM23572 | 5.241.671 | 63,7 | 83,16 | 86,86 | 95,28 |
| <i>P. oxalativorans</i> DSM23570 | 5.630.311 | 63,1 | 83,05 | 86,83 | 94,70 |
| <i>P. vervacti</i> NS15 | 5.636.000 | 63,6 | 82,86 | 86,78 | 94,04 |
| <i>P. sp</i> B6 | 4.323.500 | 63,4 | 82,79 | 86,76 | 93,89 |
| <i>P. spotorum</i> DSM21091 | 5.742.997 | 62,8 | 82,71 | 86,73 | 92,93 |
| <i>P. norimbergensis</i> DSM11628 | 6.167.370 | 63,1 | 82,4 | 86,55 | 92,74 |
| <i>P. thiooxydans</i> DSM5325 | 4.464.186 | 63,2 | 73,59 | 83,53 | 90,67 |
| <i>C. pinensis</i> DSM2588 | 9.127.347 | 45,2 | 63,07 | 80,76 | 40,63 |

ANIb: Identidade média de nucleotídeos baseada no algoritmo do BLAST

ANIm: Identidade média de nucleotídeos baseada no algoritmo do MUMmer

Tetra: Índice de correlação de assinaturas tetra-nucleotídica

6. CONCLUSÕES

A partir da comunidade bacteriana CB10b foi possível obter 6 isolados bacterianos, onde, por meio das análises dos dados de sequenciamento dos genes de 16S e ITS foram identificados os isolados CB10b_01 e CB10b_04 como sendo do gênero *Chitinophaga*, o isolado CB10b_03 como sendo do gênero *Labrys* e os isolados CB10b_02, CB10b_05 e CB10b_06 como sendo do gênero *Pandoraea*. As análises filogenéticas baseadas no gene 16S rRNA e na região espaçadora entre os genes 16S-23S para o gênero *Pandoraea* e análise do gene 16S rRNA para os gêneros *Labrys* e *Chitinophaga* evidenciam que os isolados apresentam diferenças em relação as sequências recuperadas no banco de dados do GenBank.

O ensaio baseado na atividade de enzimas capazes de degradar carboximetilcelulose apenas isolados CB10b_01 e CB10b_04 pertencentes ao gênero *Chitinophaga* apresentaram atividade enzimática, que foi possível ser detectada no meio de cultura sólido.

Além da caracterização dos isolados, verificou-se que estes organismos são de grande importância reportados pela literatura e estes podem ser veículos de novos estudos associados à obtenção de enzimas envolvidas na degradação de biomassa vegetal e biomoléculas.

Com as análises do metagenoma foi possível definir que a comunidade é predominantemente composta pelos gêneros *Labrys*, *Pandoraea* e *Chitinophaga*, onde análises de recuperação de genomas no metagenoma da comunidade CB10b, possibilitaram recuperar genomas parciais para o gênero *Pandoraea*, *Labrys* e *Chitinophaga*.

Através da análise da composição funcional com base no banco de dados do CAZy foi possível encontrar 624 ORFs distribuídas entre as classes glicosil hidrolases (31,98%), glicosil transferases (21,63%), polissacarídeo liases (3,18%), carboidrato esterases (17,00%), atividades auxiliares (4,81%) e módulos de ligação a carboidratos (15,38%), indicando que a comunidade possui um conjunto de enzimas com capacidade de atuar na desconstrução de biomassa lignocelulósica.

Além da caracterização dos isolados, verificou-se que estes organismos são de grande importância reportada pela literatura e estes podem ser veículos de

novos estudos associados à obtenção de enzimas envolvidas na degradação de biomassa vegetal e biomoléculas.

7. REFERÊNCIAS

ALMEIDA, M. L. M. **DIVERSIDADE BACTERIANA DE UM CONSÓRCIO DEGRADADOR DE RESÍDUO CELULÓSICO**. [s.l.] Faculdade de Tecnologia de Jaboticabal, 2011.

AMANN, R. I. et al. Phylogenetic identification and in situ detection of individual microbial cells without cultivation . Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. **American Society for Microbiology**, v. 59, n. 1, p. 143–169, 1995.

AMORIM, C. L. et al. Biodegradation of ofloxacin, norfloxacin, and ciprofloxacin as single and mixed substrates by *Labrys portucalensis* F11. **Applied Microbiology and Biotechnology**, v. 98, n. 7, p. 3181–3190, 2014.

ANTONIO BIZZO, W. et al. The generation of residual biomass during the production of bio-ethanol from sugarcane, its characterization and its use in energy production. **Renewable and Sustainable Energy Reviews**, v. 29, p. 589–603, 2014.

APWEILER, R. The universal protein resource (UniProt) in 2010. **Nucleic Acids Research**, v. 38, n. SUPPL.1, p. 190–195, 2009.

ARANTES, V.; SADDLER, J. N. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. **Biotechnology for Biofuels**, v. 3, n. 1, p. 4, 2010.

ARMSTRONG, Z. et al. Biocatalysts for biomass deconstruction from environmental genomics. **Current Opinion in Chemical Biology**, v. 29, p. 18–25, 2015.

BAIROCH, A. The ENZYME database in 2000. **Nucleic acids research**, v. 28, n. 1, p. 304–5, 2000.

BANKEVICH, A. et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012.

BAYER, E. A.; SHOLHAM, Y.; LAMED, R. **The Prokaryotes**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. v. 2

BEZERRA, T. L.; RAGAUSKAS, A. J. A review of sugarcane bagasse for second-generation bioethanol and biopower production. **Biofuels, Bioproducts and Biorefining**, v. 10, n. 5, p. 634–647, 2016.

BORASTON, A. B. et al. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **The Biochemical journal**, v. 382, n. 3, p. 769–81, 2004.

BRETHAUER, S.; STUDER, M. H. Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals – A Review. **CHIMIA International Journal for Chemistry**, v. 69, n. 10, p. 572–581, 2015.

BROWN, B. L. et al. Editorial: From Genes to Species: Novel Insights from Metagenomics. **Standards in genomic sciences**, v. 6, n. 1, p. 1181, 2012.

BROWN, M. E.; CHANG, M. C. Exploring bacterial lignin degradation. **Current Opinion in Chemical Biology**, v. 19, n. 1, p. 1–7, 2014.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature methods**, v. 12, n. 1, p. 59–60, 2015.

CAFFALL, K. H.; MOHNEN, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. **Carbohydrate Research**, v. 344, n. 14, p. 1879–1900, 2009.

CHANG, A. et al. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. **Current Opinion in Biotechnology**, v. 22, n. 6, p. 800–808, 2011.

DARLING, A. C. E. et al. Mauve: multiple alignment of conserved genomic sequence with rearrangements. **Genome research**, v. 14, n. 7, p. 1394–403, 2004.

DARRIBA, D. et al. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, v. 9, n. 8, p. 772–772, 2012.

DATTA, R. et al. Enzymatic Degradation of Lignin in Soil: A Review. **Sustainability**, v. 9, n. 7, p. 1163, 2017.

DE CASTRO, A. M.; PEREIRA, N. Produção, propriedades e aplicação de celulases na hidrólise de resíduos agroindustriais. **Química Nova**, v. 33, n. 1, p. 181–188, 2010.

DE GONZALO, G. et al. Bacterial enzymes involved in lignin degradation. **Journal of Biotechnology**, v. 236, p. 110–119, 2016.

DEMIRBAS, A. Fuels from Biomass. In: **Biorefineries: For biomass upgrading facilities**. London: Springer, 2010. p. 33–73.

DUTTA, S.; WU, K. C.-W. Enzymatic breakdown of biomass: enzyme active sites, immobilization, and biofuel production. **Green Chem.**, v. 16, n. 11, p. 4615–4626, 2014.

EWING, B. et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome research**, v. 8, n. 3, p. 175–85, 1998.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1998.

GAO, D. et al. Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. v. 110, n. 27, p. 10922–10927, 2013.

GARRON, M.-L.; CYGLER, M. Uronic polysaccharide degrading enzymes. **Current Opinion in Structural Biology**, v. 28, p. 87–95, 2014.

GLASS, N. L. et al. Plant Cell Wall Deconstruction by Ascomycete Fungi. **Annual Review of Microbiology**, v. 67, n. 1, p. 477–498, 2013.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: A Graphical Gool for Sequence Finishing. **Genome research**, v. 8, n. 3, p. 195–202, 1998.

GORIS, J. et al. DNA-DNA hybridization values and their relationship to whole-

genome sequence similarities. **International Journal of Systematic and Evolutionary Microbiology**, v. 57, n. 1, p. 81–91, 2007.

GUERRIERO, G. et al. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. **Engineering in Life Sciences**, v. 16, n. 1, p. 1–16, 2016.

GURLER, V.; STANISICH, V. A. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. **Microbiology**, v. 142, n. 1, p. 3–16, 1996.

HANDELSMAN, J. et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chemistry & Biology**, v. 5, n. 10, p. R245–R249, 1998.

HANKIN, L.; ANAGNOSTAKIS, S. L. The Use of Solid Media for Detection of Enzyme Production by Fungi. **Mycologia**, v. 67, n. 3, p. 597–607, 1975.

HIMMEL, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. **Science**, v. 315, n. 5813, p. 804–7, 2007.

HIMMEL, M. E. et al. Microbial enzyme systems for biomass conversion: emerging paradigms. **Biofuels**, v. 1, n. 2, p. 323–341, 2010.

HUANG, S.; SHENG, P.; ZHANG, H. Isolation and Identification of Cellulolytic Bacteria from the Gut of *Holotrichia parallela* Larvae (Coleoptera: Scarabaeidae). **International Journal of Molecular Sciences**, v. 13, n. 12, p. 2563–2577, 2012.

HUERTA-CEPAS, J. et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, v. 44, n. D1, p. D286–D293, 2016.

HYATT, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, v. 11, p. 119, 2010.

ISIKGOR, F. H.; BECER, C. R. Lignocellulosic Biomass: a sustainable platform for production of bio-based chemicals and polymers. **Polymer Chemistry**, v. 6, p. 4497–4559, 2015.

KANOKRATANA, P. et al. Phylogenetic Analysis and Metabolic Potential of Microbial Communities in an Industrial Bagasse Collection Site. **Microbial Ecology**, v. 66, n. 2, p. 322–334, 2013.

KARNAOURI, A. et al. Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila*. **Frontiers in Microbiology**, v. 5, n. JUN, p. 1–22, 2014.

KISHI, L. T. et al. Draft Genome Sequence of a Chitinophaga Strain Isolated from a Lignocellulose Biomass-Degrading Consortium. **Genome Announcements**, v. 5, n. 3, p. e01056-16, 2017.

KONSTANTINIDIS, K. T.; TIEDJE, J. M. Trends between gene content and genome size in prokaryotic species with larger genomes. **Proceedings of the National Academy of Sciences**, v. 101, n. 9, p. 3160–3165, 2004.

LAGUERRE, G. et al. Typing of Rhizobia by PCR DNA Fingerprinting and PCR-Restriction Fragment Length Polymorphism Analysis of Chromosomal and Symbiotic Gene Regions: Application to *Rhizobium leguminosarum* and Its

- Different Biovars. **Applied and Environmental Microbiology**, v. 62, n. 6, p. 2029–2036, 1996.
- LARSBRINK, J. et al. Proteomic insights into mannan degradation and protein secretion by the forest floor bacterium *Chitinophaga pinensis*. **Journal of Proteomics**, v. 156, p. 63–74, 2017.
- LEARY, N. A. O. et al. Reference sequence (RefSeq) database at NCBI : current status , taxonomic expansion , and functional annotation. v. 44, n. November 2015, p. 733–745, 2016.
- LEVASSEUR, A. et al. Expansion of the enzymatic repertoire of the CAZY database to integrate auxiliary redox enzymes. **Biotechnology for biofuels**, v. 6, n. 1, p. 41, 2013.
- LI, D. et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. **Methods (San Diego, Calif.)**, v. 102, p. 3–11, 2016.
- LI, R. et al. Biodegradation of Pentachloronitrobenzene by *Labrys portucalensis* pcnb-21 Isolated from Polluted Soil. **Pedosphere**, v. 21, n. 1, p. 31–36, 2011.
- LI, S. et al. Technology Prospecting on Enzymes: Application, Marketing and Engineering. **Computational and Structural Biotechnology Journal**, v. 2, n. 3, p. 1–11, 2012.
- LOMBARD, V. et al. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research**, v. 42, n. D1, p. 490–495, 2014.
- LYND, L. R. et al. Microbial Cellulose Utilization: Fundamentals and Biotechnology. **Microbiology and Molecular Biology Reviews**, v. 66, n. 3, p. 506–577, 1 set. 2002.
- MALHERBE, S.; CLOETE, T. E. Lignocellulose biodegradation: Fundamentals and applications. **Reviews in Environmental Science and Biotechnology**, v. 1, n. 2, p. 105–114, 2002.
- MATHEWS, S. L.; PAWLAK, J.; GRUNDEN, A. M. Bacterial biodegradation and bioconversion of industrial lignocellulosic streams. **Applied Microbiology and Biotechnology**, v. 99, n. 7, p. 2939–2954, 2015.
- MCKEE, L. S.; BRUMER, H. Growth of *Chitinophaga pinensis* on plant cell wall glycans and characterisation of a glycoside hydrolase family 27 β -L-arabinopyranosidase implicated in arabinogalactan utilisation. **PLoS ONE**, v. 10, n. 10, p. 1–22, 2015.
- MENNA, P. et al. Molecular phylogeny based on the 16S rRNA gene of elite rhizobial strains used in Brazilian commercial inoculants. **Systematic and Applied Microbiology**, v. 29, n. 4, p. 315–332, 2006.
- MHUANTONG, W. et al. Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. **Biotechnology for biofuels**, v. 8, p. 16, 2015.
- MIKHEENKO, A.; SAVELIEV, V.; GUREVICH, A. MetaQUAST: Evaluation of

- metagenome assemblies. **Bioinformatics**, v. 32, n. 7, p. 1088–1090, 2016.
- MONTELLA, S. et al. Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. **Scientific Reports**, v. 7, n. April 2016, p. 42623, 2017.
- NURK, S. et al. metaSPAdes: a new versatile metagenomic assembler. **Genome Research**, v. 27, n. 5, p. 824–834, 2017.
- OULAS, A. et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. **Bioinformatics and biology insights**, v. 9, p. 75–88, 2015.
- PANDIT, P. D. et al. Mining of hemicellulose and lignin degrading genes from differentially enriched methane producing microbial community. **Bioresource Technology**, v. 216, p. 923–930, 2016.
- PARK, S. et al. Cellulose crystallinity index : measurement techniques and their impact on interpreting cellulase performance. p. 1–10, 2010.
- PAULY, M.; KEEGSTRA, K. Plant cell wall polymers as precursors for biofuels. **Current Opinion in Plant Biology**, v. 13, n. 3, p. 304–311, 2010.
- PAWAR, P. M.-A. et al. Acetylation of woody lignocellulose: significance and regulation. **Frontiers in Plant Science**, v. 4, n. May, p. 1–8, 2013.
- PENG, Y. et al. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics**, v. 28, n. 11, p. 1420–1428, 2012.
- PROENÇA, D. N. et al. Draft genome sequence of the cellulolytic endophyte *Chitinophaga costaii* A37T2T. **Standards in Genomic Sciences**, v. 12, n. 1, p. 53, 2017.
- RALPH, J. et al. Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. **Phytochemistry Reviews**, v. 3, n. 1–2, p. 29–60, 2004.
- RICHTER, M.; ROSSELLO-MORA, R. Shifting the genomic gold standard for the prokaryotic species definition. **Proceedings of the National Academy of Sciences**, v. 106, n. 45, p. 19126–19131, 2009.
- RONQUIST, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. **Systematic biology**, v. 61, n. 3, p. 539–42, 2012.
- RUBIN, E. M. Genomics of cellulosic biofuels. **Nature**, v. 454, n. 7206, p. 841–845, 2008.
- SAHA, B. C.; JORDAN, D. B.; BOTHAST, R. J. Enzymes, Industrial (overview). In: **Encyclopedia of Microbiology**. [s.l.] Elsevier, 2009. p. 281–294.
- SANGKHOBOL, V.; SKERMAN, V. B. D. *Chitinophaga*, a New Genus of Chitinolytic Myxobacteria. **INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY**, v. 31, n. 3, p. 285–293, 1981.
- SCHÄDEL, C. et al. Hemicellulose concentration and composition in plant cell

- walls under extreme carbon source-sink imbalances. **Physiologia Plantarum**, v. 139, n. 3, p. 241–255, 2010.
- SCORTICHINI, M. et al. A Genomic Redefinition of *Pseudomonas avellanae* species. **PLoS ONE**, v. 8, n. 9, p. e75794, 2013.
- SHALLOM, D.; SHOHAM, Y. Microbial hemicellulases. **Current Opinion in Microbiology**, v. 6, n. 3, p. 219–228, 2003.
- SHI, Y. et al. Biochemical investigation of kraft lignin degradation by *Pandoraea* sp. B-6 isolated from bamboo slips. **Bioprocess and Biosystems Engineering**, v. 36, n. 12, p. 1957–1965, 2013.
- SWEENEY, M. D.; XU, F. Biomass Converting Enzymes as Industrial Biocatalysts for Fuels and Chemicals: Recent Developments. **Catalysts**, v. 2, n. 4, p. 244–263, 2012.
- TATUSOV, R. L. et al. The COG database: an updated version includes eukaryotes. **BMC bioinformatics**, v. 4, p. 41, 2003.
- TAYLOR, P.; BRUMM, P. J. Biofuels Bacterial genomes : what they teach us about cellulose degradation Bacterial genomes : what they teach us about cellulose degradation. v. 7269, 2014.
- THOORENS, G. et al. Microcrystalline cellulose, a direct compression binder in a quality by design environment—A review. **International Journal of Pharmaceutics**, v. 473, n. 1–2, p. 64–72, 2014.
- VAN DYK, J. S.; PLETSCHKE, B. I. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. **Biotechnology Advances**, v. 30, n. 6, p. 1458–1480, 2012.
- VANHOLME, R. et al. Lignin biosynthesis and structure. **Plant physiology**, v. 153, n. 3, p. 895–905, 2010.
- VARGAS BETANCUR, G. J.; PEREIRA JR., N. Sugar cane bagasse as feedstock for second generation ethanol production. Part I: Diluted acid pretreatment optimization. **Electronic Journal of Biotechnology**, v. 13, n. 3, p. 10–11, 2010.
- WEISBURG, W. G. et al. 16S ribosomal DNA amplification for phylogenetic study. **J. Bacteriol.**, v. 173, p. 697–703, 1991.
- WELFLE, A. Balancing growing global bioenergy resource demands - Brazil's biomass potential and the availability of resource for trade. **Biomass and Bioenergy**, v. 105, p. 83–95, 2017.
- WHITE III, R. A. et al. ATLAS (Automatic Tool for Local Assembly Structures) - a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. p. 1–11, 2017.
- WOESE, C. R. Bacterial evolution. **American Society for Microbiology**, v. 51, n. 2, p. 221–271, 1987.
- WONGWILAIWALIN, S. et al. Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading

consortia. **Applied Microbiology and Biotechnology**, v. 97, n. 20, p. 8941–8954, 2013.

YIN, Y. et al. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. **Nucleic Acids Research**, v. 40, n. W1, p. 445–451, 2012.

ZDOBNOV, E. M.; APWEILER, R. InterProScan - an integration platform for the signature-recognition methods in InterPro. **Bioinformatics**, v. 17, n. 9, p. 847–848, 2001.