UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

UNRAVELLING THE DIVERSITY, METABOLIC AND BIOTECHNOLOGICAL POTENTIAL OF A LIGNOCELLULOSE-DECOMPOSING BACTERIAL COMMUNITY BY GENOMIC-CENTERED METAGENOMICS

Bruno Weiss

Biological Scientist

2021

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

UNRAVELLING THE DIVERSITY, METABOLIC AND BIOTECHNOLOGICAL POTENTIAL OF A LIGNOCELLULOSE-DECOMPOSING BACTERIAL COMMUNITY BY GENOMIC-CENTERED METAGENOMICS

Discente: Bruno Weiss

Orientador: Prof.Dr. Alessandro de Mello Varani

Co-orientadora: Profa. Dra. Lucia Carareto Alves

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

2021

W429u

Weiss, Bruno

Unravelling the diversity, metabolic and biotechnological potential of a lignocellulose-decomposing bacterial community by genomic-centered metagenomics / Bruno Weiss. -- Jaboticabal, 2021

90 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Alessandro de Mello Varani

Coorientadora: Lucia Carareto Alves

1. Metagenomics. 2. Lignocelulose. 3. Biotechnology. 4. Biological communities. 5. Metabolism. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: UNRAVELLING THE DIVERSITY, METABOLIC AND BIOTECHNOLOGICAL POTENTIAL OF A LIGNOCELLULOSE-DECOMPOSING BACTERIAL COMMUNITY BY GENOMIC-CENTERED METAGENOMICS

AUTOR: BRUNO WEISS
ORIENTADOR: ALESSANDRO DE MELLO VARANI
COORIENTADORA: LUCIA MARIA CARARETO ALVES

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

Prof. Dr. ALESSANDRO DE MELLO VARANI (Participaçao Virtual)
Departamento de Tecnologia / FCAV / UNESP - Jaboticabal

Profa. Dra. CLAUDIA BARROS MONTEIRO VITORELLO (Participaçao Virtual)
Departamento de Genética-ESALQ/USP / Piracicaba/SP

Prof. Dr. DANIEL GUARIZ PINHEIRO (Participaçao Virtual)
Departamento de Tecnologia / FCAV / UNESP - Jaboticabal

Prof. Dr. LEANDRO MARCIO MOREIRA (Participaçao Virtual)
Departamento de Ciências Biológicas-ICEB/UFOP / Ouro Preto/MG

Prof. Dr. JACKSON ANTONIO MARCONDES DE SOUZA (Participaçao Virtual)
Departamento de Biologia Aplicada à Agropecuária / FCAV / UNESP - Jaboticabal

Jaboticabal, 29 de outubro de 2021

CURRICULAR DATA OF THE AUTHOR

Bruno Weiss – born on May 10th, 1988, in Piracicaba – São Paulo, Brazil, son of Eugen Weiss and Sueli Totti Weiss. Graduated in Biological Sciences in 2010 at the University of São Paulo (USP), "Luis de Queiroz" College of Agriculture (ESALQ) campus. Obtained a Master's Degree in Sciences (Microbiology) at the graduate program of Agricultural Microbiology in 2017, at University of São Paulo (USP), "Luis de Queiroz" College of Agriculture (ESALQ) campus. Joined São Paulo State University "Julio de Mesquita Filho" (UNESP) for the candidature of Doctorate's Degree in the graduate program of Agricultural Microbiology, at the Faculty of Agricultural and Veterinary Sciences (FCAV) campus. On October 29th, 2021, defended this Doctoral Thesis.

This being human is a guest house.
Every morning a new arrival.
A joy, a depression, a meanness,
some momentary awareness comes
As an unexpected visitor.
Welcome and entertain them all!
Even if they're a crowd of sorrows,
who violently sweep your house
empty of its furniture,
still treat each guest honorably.
He may be clearing you out
for some new delight.
The dark thought, the shame, the malice,
meet them at the door laughing,
and invite them in.
Be grateful for whoever comes,
because each has been sent
as a guide from beyond.

The Guest House - Jalaluddin Rumi, XIII century

DEDICATION

To those that will still walk with us,
Gordo, Branco, Menina, Princesa, Miranda, Shiva, Leozinho, Dorothy, Petunia e Barnabé
I offer this work.

To those that will also walk with us, but from the other side of the rainbow,
Leonardo, Yves, Zuleika, Vilma, Ursa, Luís, Isabela e Lara
I dedicate this work.

ACKNOWLEDGMENTS

It is difficult to orderly aggregate all persons that directly and indirectly contributed to this work. It is relevant that every contribution to the growth of my character and intellect, in one way or another, had implications on this work. I consider that this is the product mainly of educational, personal, and professional excellence contributed by many teachers. Some of those are professors, many are friends, and surely, some are family members – and as such, the list grows exponentially. I am, thus, forced to reduce, as everything worthwhile for the mind, heart, and soul.

I thank my mother for the unbounded support for all my endeavors, and the love. It was, is, and will always be the substrate for my confidence and valor over my work and life.

I thank my brother, sister and nieces for their encouragement and support. Although in many very important events I could not be present, your patience and recognition of the importance of this work allowed me the see the same importance clearly – particularly needed in many times of difficulties.

I thank professor Alessandro de Mello Varani for accepting me as his student, and for all his teachings – technical, scientific, and professional – which served me as a guiding map to overcome the natural (nonetheless ominous) difficulties of a Doctorate Course. I also appreciate his patience through my oscillating focus that too often wobbled between existentialist deliberations, epistemological ramblings, and (often excessive) scientific reductions.

I thank professor Lucia Maria Carareto Alves for also accepting me as her student and support. Without her microbiological and biochemical teachings and suggestions, this work wouldn’t be possible, literally. I also thank Dr. Milena Tavares and M.Sc. Anna Carolina Souza, for their highly refined wet-lab skills and knowledge, and the resulting high-quality data.

I thank all my friends, Tadeu, Danillo, Heloísa, Natália, Caio, Thiago, Helen, Wallynson, Ana, Bruno, Michelli, Saura, Rafael, for all your support and patience. Without you, and the unpredictable chaotic events that were mostly driven by you all (I

may have some secondary participation, allegedly), the path would have been unbearably predictable.

I thank Célia Maria Fernandes, for all emotional support and comprehension. Such guidance, in relevant moments of my path, contributed heavily to this accomplishment.

I had the support of Programa de Pós-Graduação em Microbiologia Agropecuária. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. FAPESP financed the overarching process (16/16624-1), as also CNPq (under process number 303061/2019-7).

SUMMARY

REVELANDO O POTENCIAL METABÓLICO E BIOTECNOLÓGICO DE UMA COMUNIDADE BACTERIANA DECOMPOSITORA DE LIGNOCELULOSE POR METAGENOMIA CENTRALIZADA EM GENOMAS

RESUMO - Os processos de produção de biocombustíveis de segunda geração estão em alta demanda. Na natureza, a lignocelulose pode ser rapidamente desconstruída. A eficácia do fenômeno na natureza pode ser devido à divisão de trabalho bioquímico efetuado nas comunidades bacterianas. Aqui, projetamos um experimento para coletar e enriquecer seletivamente um consórcio bacteriano obtido do solo e sobras de palha seca da cana-de-açúcar de uma usina de cana-de-açúcar. Procedeu-se ao sequenciamento de duas amostras provenientes da 2ª e 20ª semanas (fração bacteriana anexada as fibras da cana-de-açucar e a fração bacteriana livre disposta no meio de cultivo), visando definir a dinâmica composicional do consórcio, sua composição metagenômica, espécias, abundâncias e potencial metabólico em relação às suas capacidades desconstrutoras de lignocelulose e de controlar ecofisiologicamente a dinâmica da comunidade. Fomos capazes de separar 52 genomas das leituras metagenômicas, apresentando alta completude e baixa contaminação, estimando a abundância relativa de cada espécie representada por tais genomas montados a partir dos dados metagenômicos. Observamos a dinâmica do consórcio, em que ~46% das espécies não mostrou nenhuma modificação relevante em sua abundância. Três espécies encontradas na 2ª semana estavam mormente ausentes na 20ª semana, enquanto todas as espécies altamente abundantes encontradas no início mantiveram alta abundância na 20ª semana. Também, foi possível reconstruir o metabolismo de cada genoma. Encontramos sequências relacionadas à desconstrução da lignina na maioria dos genomas, e muitas das espécies mais abundantes em todas as amostras e frações são relacionadas a gêneros de organismos conhecidos por serem capazes de desconstruir a lignina. Propusemos um modelo metabólico qualitativo que nos permite ver as diferentes potencialidades cada genoma e grupo taxonômico no que diz respeito à desconstrução de polímeros sacarídicos e lignina, e também esclarecer a divisão de trabalho bioquímico que este consórcio pode utilizar para desconstruir a lignocelulose. Estas informações poderão auxiliar na estruturação de decisões para esforços de engenharia e biologia sintética visando o aprimoramento ou seleção de partes do consórcio para a execução de etapas específicas na desconstrução de biomassa lignocelulósica.

Palavras-chave: metagenômica, lignocelulose, biotecnologia, comunidades biológicas, metabolismo

UNRAVELLING THE DIVERSITY, METABOLIC AND BIOTECHNOLOGICAL POTENTIAL OF A LIGNOCELLULOSE-DECOMPOSING BACTERIAL COMMUNITY BY GENOMIC-CENTERED METAGENOMICS

ABSTRACT - Processes for the production of second-generation biofuels are in high demand. In nature, lignocellulose can be rapidly deconstructed. The effectiveness of the phenomenon in nature may be due to the division of biochemical labor effectuated in bacterial communities. We designed an experiment to collect and selectively enrich a bacterial consortium obtained from soil and sugarcane dry straw leftover from a sugarcane milling plant. This consortium was cultivated for 20 weeks in aerobic conditions, where the only source of carbon was sugarcane lignocellulosic biomass. We proceeded to sequence three samples from the 2nd and 20th weeks, Attached and Free fraction, to define the consortia's compositional dynamics, metagenomic composition, species, abundances, and metabolic potentials capacities deconstruct the lignocellulose and to control its dynamics ecophysiologically. We separated 52 genomes from the metagenomic reads with high completeness and low contamination, estimating each species' relative abundance represented by such metagenome-assembled genomes. We observed the consortium dynamics, in which ~46% of species showed no relevant modification in their abundance. Three species found in the 2nd week were mostly absent in the 20th week, while all the highly abundant species found in the beginning kept such high abundance in the 20th week. Also, it was possible to reconstruct the metabolism of each genome. We found sequences related to lignin deconstruction in most genomes. Many of the most abundant species in all samples and fractions are genera-related to organisms known to deconstruct Lignin. We proposed a qualitative metabolic model allowing us to see each genome and taxonomic group's different potentialities regarding the deconstruction of saccharidic polymers and lignin and clarifying the division of biochemical labor this consortium may utilize to deconstruct the lignocellulose. This information helps structure decisions for engineering and synthetic biology efforts to improve or select parts of the consortium to execute specific steps in the deconstruction of lignocellulosic biomass.

Keywords: metagenomics, lignocellulose, biotecnology, biological comunities, metabolism

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

The growing demand for renewable fuels as a product of society's insight for less environmentally impacting solutions has been nursing research aiming for renewable fuel production improvements. First-generation ethanol production from sugarcane is relevant for the production of renewable fuel. Although this product has a relatively high yield, it can also generate large residue quantities, as lignocellulosic biomass or bagasse, a byproduct of harvesting and processing. This biomass contains more than 65% of the plants' fixed energy in organic polymers. This biomass is organized in a highly sophisticated manner through various polysaccharides and other biopolymers (Pippo *et al.*, 2011). At the industrial scale, the lignocellulose can be partially deconstructed into fermentable sugars, increasing the overall fuel yield through its use in second-generation ethanol production (Limayem *et al*., 2012; Bond *et al.*, 2013; Puentes-Tellez *et al*., 2018). Several efforts are in course for lignocellulosic biomass use for fuel generation on an industrial scale. However, its deconstruction into other polymers is still a very challenging process.

The polymers found in the lignocellulosic biomass are mostly cellulose, hemicellulose, pectin, and lignin (Zhao *et al*., 2012). Cellulose is the most abundant organic polymer on Earth, and generally, the main component of lignocellulosic biomass (around 25%). The cellulose is a linear polysaccharide composed of glucose units (French, 2017), forming long-chain affords a rich intermolecular interaction web, resulting in a very rigid, crystalline, or quasi-crystalline alternating with regions of less crystalline structure (Bayer *et al*., 2013).

Conversely, hemicellulose is amorphous and flexible overall structure, more amenable to deconstruction (Zhou *et al*., 2017). It attaches itself to this wooden cellulose structure, but being a more heterogeneous polymer, highly branched, and composed of various pentoses (e.g., D-xylose, D-arabinose) and hexoses (e.g., D-galactose, D-mannose, and D-glucose) (Zhou *et al*., 2017). Pectin, an essential component of other plant tissues (e.g., fruits), is of little interest in studying biomass deconstruction because it is usually absent or present at minimal amounts in mature sugarcane tissues (Ridley *et al*., 2001). Beyond being the second most abundant organic polymer on Earth, Lignin is of high relevance for the process of deconstruction of lignocellulolytic biomass. It acts as a retardant to cellulose and hemicellulose breakdown (dos Santos *et al*., 2018). Lignin is a highly branched, phenolic polymer composed of diverse monomers, and so its decomposition does not result in fermentative sugars (Kamimura *et al*., 2019). Notably, Lignin is rich in aromatic units (i.e., Coniferyl, sinapil, *p*-coumaryl alcohols). Its composition varies between plant species, tissues, and age, lacking a definitive primary structure (Isikgor *et al.*, 2015). Therefore, lignocellulose confers rigidity and impermeability to the plant cell wall and recalcitrance to the biomass (Bayer *et al*., 2013). For that reason, various enzymes are necessary for lignocellulose deconstruction at an industrial scale, reflecting the complexity of compounds found in the lignocellulose biomass. Some of such enzymes are grouped under Glycosyl Hydrolases protein families by the Enzyme Commission (EC).

However, in nature, bacterial communities or consortia can efficiently deconstruct lignocellulose (Wilhelm et al., 2018, Boer et al., 2005). These bacterial consortiua may synthesize oxidative and hydrolytic enzymes that synergistically act, breaking down the lignocellulosic biomass (Sweeney *et al*., 2012; Beckham *et al*., 2016). This synergistic operation of biochemical deconstruction may also assist the maintenance of the

biogeochemical cycles. For instance, the Laccase oxidative enzyme (EC 1.10.3.2) performs the deconstruction of cell-wall component lignin, allowing other hydrolytic proteins, such as cellulases and xylanases, to access polysaccharides and oligosaccharides (Carvalho *et al*., 2010). The hydrolysis of polysaccharides reduces sugars, particularly monosaccharides as glucose and xylose, which may serve as a carbon source for the community's organisms.

Cellulases (EC 3.2.1.-) can hydrolyze glycosidic bonds ß-(1,4) between cellulose dimers (cellobiose) and monomers (glucose). Depending on the way of action, cellulases are classified as endoglucanases (EC 3.2.1.4), cellobiose hydrolases, or exo-glucanases (EC 3.2.1.91), and ß-glucosidase (EC 3.2.1.21). Each cellulase act in different ways to break down the plant cell wall (Perez *et al*., 2002, Sweeney *et al*., 2012). Cellulases belong to the Glycosyl Hydrolase family (e.g., GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH48, GH74, and GH124) (Bayer et al., 2013). Hemicellulases present many modes of action and specific substrates, considering that hemicellulose polymer is composed of a plurality of hexoses and pentoses monomers. Hemicellulose hydrolyzing enzymes may be found included in GH3, GH5, GH8, GH10, GH11, GH30, GH39, GH43, GH51, GH52, GH54, GH62, GH116, GH129, and GH127 families (Bayer et al., 2013).

Interestingly, these cellulolytic and hemicellulolytic enzymes can also be found free or as multimeric complexes called cellulosomes. Cellulosomes include carbohydrate-binding modules (CBM), which bonds to the complex substrate surface (i.e., the polysaccharides), increasing the hydrolysis' efficiency (Lynd *et al*., 2002). Another plethora of enzymes may contribute to cellulases and hemicellulases into the deconstruction of lignocellulosic biomass. These are grouped into the Glycosyl Transferases (GTs), Auxiliary Activity (AA), and Polysaccharide Lyases (PLs) enzymes.

In this study, we isolated lignocellulose deconstructing bacterial consortium obtained from above-ground sugarcane straw. This bacterial consortium was cultivated *in-vitro* through 20 weeks using lignocellulosic biomass (sugarcane bagasse) as the only carbon source. Using bioinformatics and metagenomics approaches, we evaluated the consortium's ability to deconstruct and use sugarcane bagasse as a carbon source. For this goal, we separated two contrasting time samplings (2^{nd} and 20^{th} weeks of the lignocellulose deconstructing *in-vitro* community cultivation), coupled with an identification of the bacterial community closely associated with the sugarcane bagasse fibers (Attached-fraction) and non-closely-associated to the sugarcane bagasse fibers (Free-fraction). We sought to define the consortium's taxonomic composition, function, and metabolic potential. Moreover, we identified sequences of enzymes related to polymers deconstruction, biogeochemical cycles, and secondary metabolites potentially involved in the consortium's interactions.

Furthermore, we offered a model of taxonomy-defined Division of Biochemical Labor (DoBL) for this consortium. Here, we define DoBL as the concatenation of syntrophic relationships between a bacterial consortium's components, observing each step in the execution of a defined complex biochemical reaction designated to one or a few members, being these species or groups, of such community. Our results clarify the metabolic process potentially involved in lignocellulose deconstruction and the relationship between the parts of this system, contributing to future improvements for the use of this technology in the production of biomass-derived biofuels.

2. LITERATURE REVIEW

2.1. The ecological, economic, and environmental importance of lignocellulose

Due to the perception of global warming, environmental impact, and the global warming connection to unsustainable practices and the use of fossil fuels, there is a growing tendency to a partial or total abandon of products and energy obtained from petroleum. However, such a transition will not occur quickly, as energy sources with characteristics similar to those derived from petroleum are necessary to adequate the world economy (Hill *et al.*, 2006). An appropriate solution is polymers of biological origin for energy generation (Hill et al., 2006; Rubin, 2008).

Lignocellulose sources have been recently studied, aiming to answer this economic directive tendency. The production of the second generation (2G) ethanol from corn, wheat, and rice straw, sweet sorghum, and sugarcane bagasse have being scrutinized for their economic potential (Milanez *et al.*, 2015; Puentes-Tellez *et al*., 2018). In Brazil, conventional ethanol production, beyond sugar production, abundantly results in lignocellulosic biomass, which focuses on 2G ethanol research. For each tonne of processed sugarcane, between 270 and 280 kg of bagasse is generated (Rodrigues *et al*., 2003). It is estimated that more than 65% of the total energy found in sugarcane is allocated to the lignocellulose biomass (Pippo *et al.*, 2011). In this manner, it is clear that using this material to produce biofuels is an attractive environmental solution with economic significance.

Lignocellulose is the generic term used to designate the fraction of diverse polymers found in biological material residues derived from plant tissues (Ogeda *et al*., 2010). Such polymers are synthesized in the formation of plant tissues during the development of the plant cell wall. The plant cell wall is a rigid polysaccharide matrix, recalcitrant to decomposition,

which confers rigidity and durability. It is composed of cellulose, the most abundant biopolymer on Earth, used in the agroindustry to produce renewable biofuels. It also acts as one of the largest biologically fixated carbon pools on the carbon cycle (Aristidou *et al*., 2000; Fang *et al.*, 2019). In the particular case of sugarcane, lignocellulosic biomass is composed of the straw produced in the mechanical harvest and the bagasse produced as a byproduct of the sugarcane processing at the conventional sugar alcohol production (Canilha *et al.*, 2012).

Brazil is the largest sugarcane producer and is the most significant conventional sugarcane ethanol producer (Bordonal et al., 2018). Although it is an affluent opportunity for Brazil to increase its biofuel production and mitigate the environmental impact, lignocellulose deconstruction is a very complicated process. It is worth noting that, in nature, this process often occurs, as it is part of biogeochemical cycles. In this way, understating and applying bacterial communities to transform polysaccharides into fermentable sugars may emerge as a potential new solution for the agroindustry and production of renewable fuels (Hill *et al.*, 2006; Rubin, 2008).

2.2. Composition and complexity of lignocellulose

Lignocellulose composition is highly complex, encompassing diverse hetero and homopolymers, mainly carbohydrates (Figure 1). This composition is variable between tissues on the same plant, between plant species, and between tissues from specimens on different maturation and developmental stages (Cooper *et al.*, 2016). Its main components are cellulose (35 to 50%), hemicellulose (20-35%), and pectin, all polysaccharides linked to Lignin (15-20%).

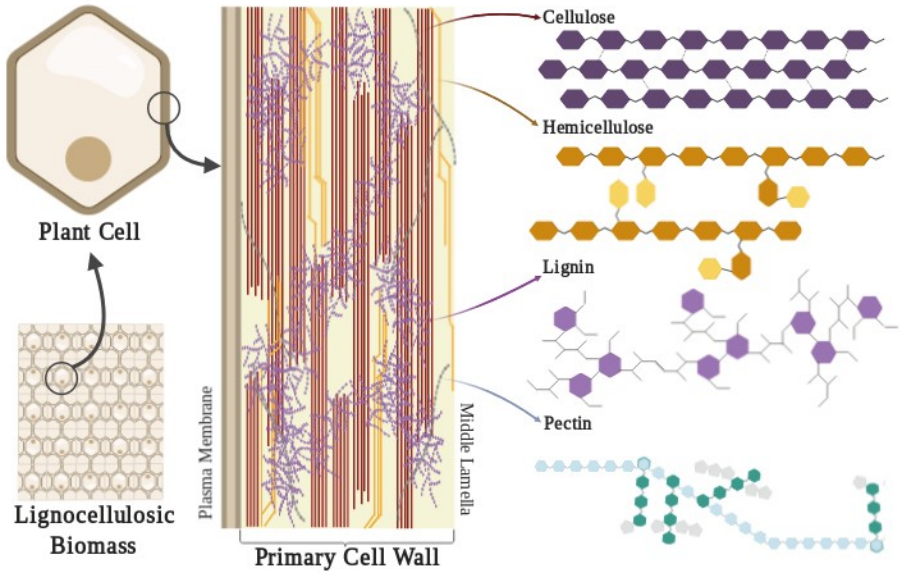

Figure 1. Schematic representation of the plant cell wall and the lignocellulosic biomass concerning its diverse composing molecules. Source: Weiss (2021)

Cellulose is a linear and highly stable homopolymer. Structurally composed of glucose monomers linked by glycosidic bond β-(1,4) (Kang *et al*., 2014; Brethauer *et al*., 2015). This relatively simple composition is structured in individual chains composed of hundreds to thousands of glucose units. These units are organized as microfibril clusters mainly through a plethora of hydrogen bonds, ordered in highly rigid regions alternated by less crystalline, amorphous, flexible regions, which are more available to hydrolysis (Cooper *et al*., 2016) (Figure 2).

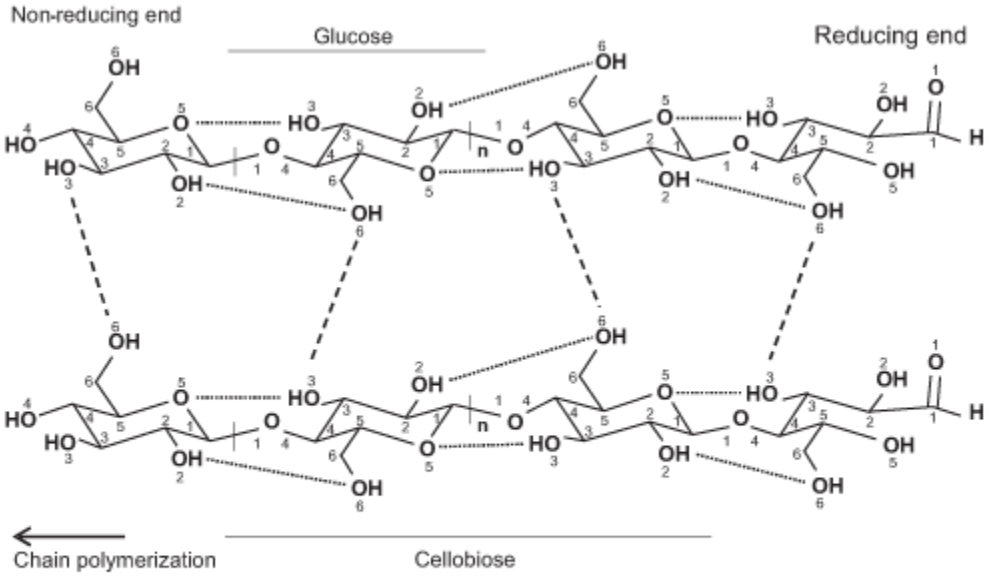

Figure 2. Cellulose structure, showing many glucose monomers, organized as cellobiose units, that compose the polymer. Dashed lines represent hydrogen bonds. Source (modified): Festucci-Buselli *et al*., 2007.

Hemicellulose is a linear or ramified heteropolymer, usually showing lower molecular weight than cellulose. It is associated with cellulose and lignin in the plant cell wall. The hemicellulose is composed of pentoses (xylose, arabinose, rhamnose), hexoses (dextrose, mannose, galactose), and uranic acid, presenting an overall amorphous structure (Kang *et al*., 2014; Brethauer *et al*., 2015).

Hemicellulose links strongly to cellulose through hydrogen bonds, forming a porous and fibrous web responsible for the plant cell wall's mechanical resistance capacity (Cooper *et al.*, 2016). It is interesting to note that the hemicellulose chemical and structural diversity contrast to the cellulose (Figure 1). The highly crystalline and rigid cellulose structure is embedded in the amorphous and flexible hemicellulose matrix (Bayer *et al*., 2013) (Figure 3).

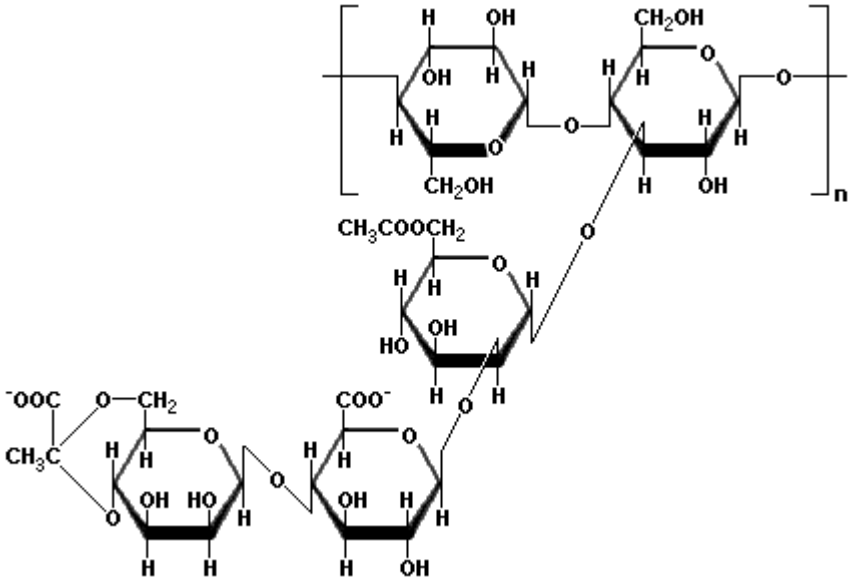

Figure 3. Example of hemicellulose structure. In contrast to cellulose, hemicellulose presents more diversity in its composing monomers, which beyond glucose can include xylose, arabinose, mannose, and many others (see text) (Source: Kulkarni *et al*., 2012).

Pectin is a structural polysaccharide that is mainly involved in the mediation of the plant's defense response and regulating plant tissue development (Cooper *et al.*, 2016). It is composed of highly hydrophilic galacturonic acid heteropolymers linked by α-(1,4) bonds and often substituted in diverse groups such as galactose, rhamnose, and xylose methylated to the carboxyl group (Cooper *et al.*, 2016). The pectin can link to cations, attracting water molecules, resulting in a gelatinous matrix in which other plant cell wall polymers may be immersed (Cooper *et al.*, 2016). During healthy plant tissue development, the quantity, density, composition, and pectin structure may change (Palin *et al*., 2012). For instance, in mature plant cells, as is the case of sugarcane lignocellulosic biomass, pectin is found in scarce amounts (Bayer *et al.*, 2013; Cooper *et al*., 2016).

Lignin is a non-saccharidic, highly recalcitrant phenolic heteropolymer, showing high molecular weight (Mood *et al*., 2013; Brethauer *et al.*, 2015). It is composed of phenyl propane monomers and coniferyl, sinapyl, and paracoumaryl alcohols, requiring oxygen and a specific set of enzymes for its degradation (Kang *et al*., 2014; Brethauer *et al*., 2015). Although lignin is not a polysaccharide, it is a primary compound of lignocellulose. Therefore, lignin is a component largely responsible for hindering the biomass's deconstruction (Figure 4).

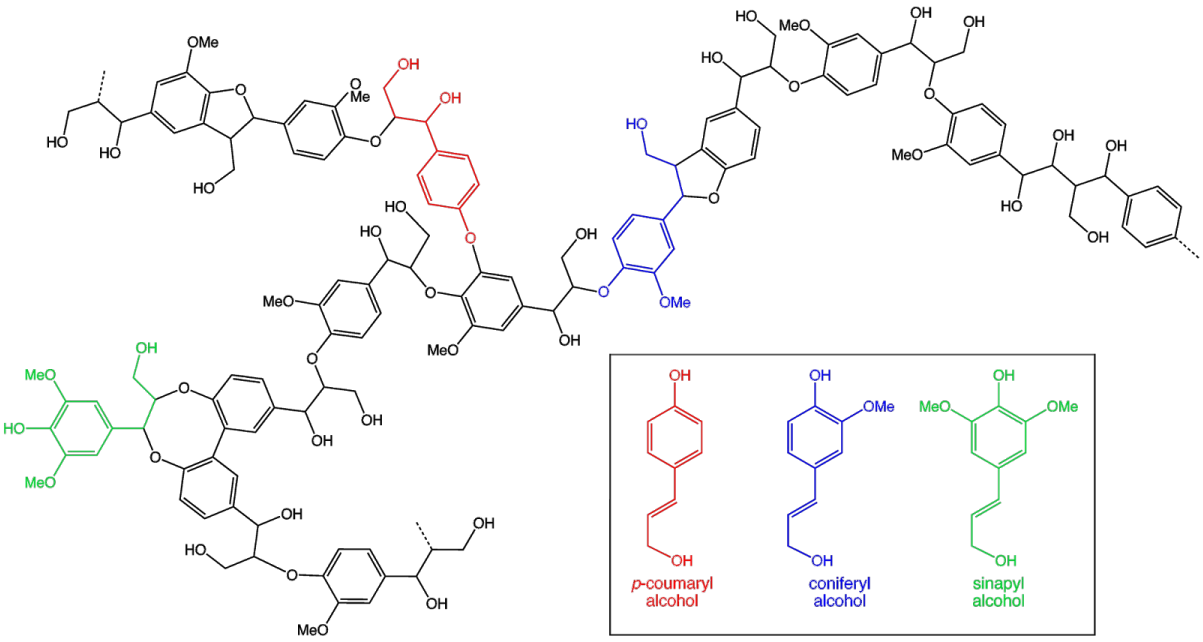

Figure 4. Example of lignin structure, and its composing monomers. Source: www.howaboutlignin.blogspot.com

2.3. Lignocellulose deconstruction by bacterial consortia

In nature, the deconstruction of lignocellulose polymers occurs through a broad diversity set of enzymes, reflecting the structural and chemical complexity found in such biomass. These enzymes act synergistically and are synthesized by bacteria, although not

exclusively (Sweeney *et al.*, 2012; Beckham *et al*., 2016). The oxidative enzymes, such as laccase (EC 1.10.3.2) and lignin-peroxidase (LiP) (EC 1.11.1.14) works mainly at the deconstruction of lignin, allowing other enzymes (e.g., cellulases and xylanases) to access the polysaccharides and oligosaccharides over which they may act (Carvalho *et al*., 2010).

The polymers' hydrolysis reduces sugars, including glucose, a carbon source to most organisms that may comprise a consortium. A broad number of enzymes are related to deconstructing the lignocellulose, many are grouped into Glycosyl Hydrolases (GH) families by the Enzyme Commission (EC) (Bayer *et al*., 2013).

The deconstructing enzymes of cellulose, cellulases (EC 3.2.1.-) can hydrolyze glycosidic bonds β-(1,4) between the monomers of cellulose, and are classified into three groups according to its mode of action over the substrate: i) endoglucanases or endo-1,4-β-glucanases (EC 3.2.1.4), ii) cellobiose hydrolases are included in the exoglucanases or Exo-1,4-β-glucanases group, and iii) β-glycosidases (EC 3.2.1.21), which hydrolyze cellobiose, resulting in glucose molecules (Perez *et al*., 2002; Sweeney *et al*., 2012; Zabed *et al*., 2019; Houfani *et al*., 2020). The cellulases are classified into the Glycosyl Hydrolases families such as GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH48, GH74, and GH124 (Bayer *et al*., 2013) (Table 1).

Table 1. Enzymes and their Families can deconstruct the polymers found in the plant cell wall. GH: Glycoside Hydrolases; CE: Carbohydrate Esterases; AA: Auxiliary Activities; PL: Polysaccharide Lyases.(continues)

Cellulases and Hemicellulases	E.C. Number	GH Families
α-amylase	3.2.1.1	13, 14, 57, 119
α-galactosidase	3.2.1.22	4, 27, 31, 36, 57, 97, 110
α-glucosidase	3.2.1.20	4, 13, 31, 63, 97, 122
α-glucuronidase	3.2.1.139	67
α-L-arabinofuranosidase	3.2.1.55	51, 54, 62
α-L-fucosidase	3.2.1.51	29, 59
α-N-acetylgalactosaminidase	3.2.1.49	36
α-N-acetylglucosaminidase	3.2.1.50	89
α-1,2-mannosidase	3.2.1.24	31, 89, 92
α-xylosidase	3.2.1.177	31
β-galactosidase	3.2.1.23	1, 2, 3, 35, 42, 50
β-glucosidase	3.2.1.21	1, 3
β-glucuronidase	3.2.1.31	1, 2
β-mannosidase	3.2.1.25	1, 2, 5
β-xylosidase	3.2.1.37	3, 30, 39, 43, 52, 54
Arabinase	3.2.1.99	43
Cellobiohydrolase	3.2.1.91	6, 9
Endo-1,4-β-galactanase	3.2.1.89	53
Endo-1,6-β-galactanase	3.2.1.164	5
Endo-1,3(4)-β-glucanase	3.2.1.6	16
Endo-1,3-β-xylanase	3.2.1.32	10, 11, 26
Endo-1,4-β-xylanase	3.2.1.8	5, 8, 10, 11, 43
Endoglucanase	3.2.1.4	5, 6, 7, 9, 12, 44, 45, 51, 74
Exo-1,4-β-glucanase	3.2.1.74	9
Glucan 1,3-β-glucosidase	3.2.1.58	3, 5, 17, 55
Glucan 1,4-β-glucosidase	3.2.1.74	3
Galactan 1,3-β-galactosidase	3.2.1.145	43
Glucoamylase	3.2.1.3	15, 97
Mannan endo-1,4-β-D-mannosidase	3.2.1.78	5, 26, 113
Maltodextrin glucosidase	3.2.1.20	4, 13
Malto-oligosyltrehalose trehalohydrolase	3.2.1.141	13
Trehalase	3.2.1.28	3, 15, 37, 65
Xylan α-1,2-glucuronidase	3.2.1.131	67, 115
Xylanase	3.2.1.8	5, 8, 10, 11, 43

Table 1. Enzymes and their Families can deconstruct the polymers found in the plant cell wall. GH: Glycoside Hydrolases; CE: Carbohydrate Esterases; AA: Auxiliary Activities; PL: Polysaccharide Lyases. (end)

Cellulases and Hemicellulases	E.C. Number	CE Families
Acetyl xylan esterase	3.1.1.72	1, 2, 3, 4, 5
Feruloyl esterase	3.1.1.73	1
Pectin acetylesterase	3.1.1.72	12
Pectin mettylesterase	3.1.1.11	8
Ligninases	**E.C. Number**	**AA Families**
Catalase	1.11.1.6	2
Lignin peroxidase	1.11.1.14	2
Manganese peroxidase	1.11.1.13	2
Peroxidase	1.11.1.7	2
Pectinases	**E.C. Number**	**PL Families**
Pectate lyase	4.2.2.2	1, 3, 9, 10
Rhamnogalacturonan lyase	4.2.2.-	11

Adapted from Bayer *et al*., 2013, *Lignocellulose-Decomposing Bacteria and Their Enzyme Systems*; Carbohydrate Actives Enzymes (http://www.cazy.org/) (Lombard *et al*., 2014).

Cellulose and hemicellulose deconstructing enzymatic systems can be found as free enzymes or bundled in multimeric complexes called cellulosomes, which may comprehend modules beyond the hydrolytic one (Lynd *et al*., 2002). This is the case of carbohydrate-binding modules (CBM), which attach the protein to the polymer substrate's surface, increasing the efficiency of the hydrolysis accomplished by the hydrolytic module (Lynd *et al*., 2002; Dash *et al*., 2019). These enzymes' synergistic activity allows the deconstruction of the insoluble polymers of difficult access for most enzymes, frequently constituted of up to thousands of monomers, into unitary molecules of sugars.

The enzymes lignin peroxidase, manganese peroxidase, and laccase can deconstruct the heterogeneous, ramified, and aromatic lignin structure (Wang *et al*., 2016). Such enzymes act oxidatively instead of hydrolytically as Glycosyl Hydrolases. They are also accountable for the deconstruction of cellulose and hemicellulose, evidencing the inherent difficulties of

biomass deconstruction. These enzymes were found in a broad diversity of organisms, as in the white-rot fungi (*Trametes hirsuta*) and brown-rot fungi (*Gloeophyllum trabeum*), which degrades lignin, leaving the wood with altered coloration. Such processes may also be affected by other groups, such as bacterial genera *Streptomyces*, *Pseudomonas*, *Rhodococcus*, *Burkholderia*, *Microbacterium*, *Acinetobacter*, and *Sphingomonas*, along with many others. However, the knowledge about ligninolytic bacteria mechanisms is still limited (Wang *et al*., 2016; Basak *et al*., 2020). Although the lignin does not result in monosaccharides by the end of its deconstruction, the lignin recalcitrance to the biotechnological transformation of lignocellulosic biomass into fermentable sugars is relevant to the industry.

Therefore, lignin in the biomass is the main factor in the elevated price of the deconstruction of lignocellulose for biofuel production, rising to 100% on the products' price (Lindy *et al.*, 2008). The industrial use of lignocellulolytic enzymes in high concentrations demands alteration in temperatures and use of extreme pH, which also elevates the cost of preparation of the biomass, reducing the commercial appeal for the biofuel production (Merino *et al*., 2007; Shi *et al*., 2009; Chylenski *et al.*, 2019). Considering the broad diversity of organisms able to deconstruct the polymers found in the biomass, the use of synergistic microbial communities for such an end seems promising (Taherzadeh *et al*., 2008; Wongwilaiwalin *et al*., 2013).

2.4. Data and search for lignocellulose deconstruction related enzymes in genomes and metagenomes

The CAZy (carbohydrate-active enZymes) Database is the standard accession hub for lignocellulolytic domain sequences analysis. This Database organizes and classifies all

known CAZymes (carbohydrate-active enZymes). CAZymes are enzymes that act over glycans, glycoconjugates, and polysaccharide assembly (Glycosyl Transferase), breakdown (Glycoside Hydrolases, Polysaccharide Lyases, Carbohydrate Esterases), and modification (http://www.cazy.org/) (Lombard *et al*., 2014; Terrapon *et al*., 2017).

CAZyme Database (CAZyDB) is frequently revised and reorganized as the field expands and changes following new evidence, and improvements are proposed. Such data is also manually gathered and curated by specialists, refining the knowledge obtained through the high number of published studies (Terrapon *et al*., 2017). Table 2 presents a summary of information obtained in the database as of June 2020, based on 150 manually curated and more than 18,000 genomes and metagenomes (Table 2, http://www.cazy.org/Genomes.html). The number of sequences of all domain types is *circa* 1.8 million at present.

Table 2. Domains found in CAZy Database (http://www.cazy.org/) as of June 2020.

Domains	Taxonomic Designation				Unclassified	Structure	Experimentally Characterized	Families
	Archaea	Bacteria	Eukaryota	Virus				
GH	3,441	611,087	52,425	120,292	2,494	1,471 (39)	7,377	167
GT	10,035	598,223	60,187	3,004	978	269 (2)	2,045	111
PL	37	21,586	1,566	457	29	84 (2)	347	40
CE	2,155	74,953	3,486	8	74	100 (1)	262	18
AA	6	5,849	8,327	233	4	116 (1)	241	16
CBM	580	200,483	13,118	1,774	220	404 (11)	2,438	87
Total	16,254	1,512,181	139,109	125,768	3,799	2,444 (56)	12,710	439

All main types of CAZymes domains are summarized: Glycoside Hydrolases (GH), Glycosyl Transferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Auxiliary Activities (AA), and Carbohydrate-Binding Modules (CBM). Each domain shows the number of sequences found in Archaea, Bacteria and Eukaryota taxonomic Domains, and Virus. Also, such quantification is shown for taxonomically unassigned sequences ("Unclassified," sixth column). The "Structure" (seventh) column indicates the number of sequences that had their structure described. In parenthesis, it is shown the number of crystallographic-level structures resolved for such a family of peptides. The number of experimentally characterized sequences and the number of families found in each domain are indicated in the eighth and ninth columns.

Although the CAZyDB is an essential tool for the organization of carbohydrate-active enzymes, it cannot overcome some difficulties that emerge as the gargantuan amount of data

and metadata is rapidly generated in genomics and metagenomics. To best apply all the evolving concepts found in the CAZyDB to this fast-generating-data field, the dbCAN (automated Carbohydrate-active enzyme ANnotation) meta server may be used to propagate the curated information using a high-quality methodology (Yin *et al*., 2012).

In its most recent embodiment, dbCAN2 integrates three different databases/methodologies for annotation of new sequences derived from genomics and metagenomics projects and based on CAZyDB data, being as I) HMMER to define domain boundaries based on a CAZy HMM database (Wheeler *et al*., 2013), ii) DIAMOND fast blast against CAZyDB (Buchfink *et al*., 2015), and iii) Hotep for conserved peptide pattern search in the PRP library (Busk *et al*., 2017).

Combining these three approaches is recognized as an optimal attempt to find the domain sequences, and identify the families they may pertain to while reducing the chance of false positives, and also being fast enough to keep up with the genomics/metagenomics sequencing and publication pace.

2.5. Metagenomics in the studies of bacterial communities

Traditional genomic procedures require axeny, and so often fail to capture the microbial community members' totality (Alvarenga *et al*., 2017). Although informative as axenic-based studies have been in the past, and still are in other situations, currently cultivation-free methods are likely to provide information about this complex and synergistic system (Wood *et al*., 2014). Considering the complexity and the synergy of lignocellulolytic bacterial communities' activity, the metagenomics methods to access the diversity and genomic content are ideal to further the knowledge of such activities' biochemistry (Steele *et al*., 2009; Christoserdova, 2010).

Metagenomics is the field of science that deals with obtaining total genomic information about the microbiological communities, without isolating the species that compose such communities (Christoserdova, 2010). This allows the genomic study of both previously known and unknown species or groups. In addition to the compilation of species, this approach permits access to all genomic information, enabling the inference of all information that can be stored in genomic/genetic format, such as the potential metabolism. This is in contrast to Metataxonomic studies, also a cultivation-free method, in which only one (*a priori*) universal marker gene (16S sequence) is sequenced from the total pool DNA obtained from the community, leading to answer about the composition and abundance of species fond in the community.

Often Metagenomics and Metataxonomics are banded together, and called ‘Metagenomic studies’, although inappropriately. Metagenomics refers to the totality of genes found in a community, while Metataxonomics refers to the richness and abundance of species based on only one marker gene sequence. The discernment obtained from Metataxomomics is a result of the variations found in the 16S universal marker gene for each taxonomic group. This confusion may also have historical origins, as Metataxonomics was the first cultivation-free method utilized, and was already inappropriately called Metagenomics in the past. In Metagenomics, it is also possible to study a particular metabolism of each community's component, through the separation (“binning”) of individual genomes of each specific component of the community, and the inferring of its particular metabolism inside the community’s shared metabolism. Such genomes are called Metagenomic Assembled Genomes (MAGs) – in a sense, Metagenomics can answer ‘Who is there’ and also ‘What they *might* be doing’.

The shared metabolism is the set of metabolic transformation processes divided among the different groups that comprise the community. The division of processes between the community's components is a phenomenon found in systems such as biogeochemistry, in which many biochemical processes are necessary for the completeness of the transformations, thus requiring the division of such activities between species (Handelman, 2004; De Souza *et al*., 2013).

As an example of the potential for genomic resolution of non-cultivable organisms in Metagenomic study, Parks and colleagues (2017) reported the reconstruction of 7,903 bacterial and archaean genomes from 1,550 metagenomes available in public databases (Parks *et al*., 2017). Among the Phyla found, 17 bacterial and 3 Archaean were represented by their first to be known members (Parks *et al*., 2017). Also, many of the already known Phyla accessed were enriched with new representatives. The implications of such explorations of bacterial and archaean diversity and functionality are a direct product of the impossibility of separation of many of the organisms in a community, which curb our comprehension of the community's internal operations, and demand isolation of its components – which is often impossible or very difficult. It is estimated that less than 1% of bacteria and archaea are cultured in the laboratory, in simplified axenic systems. The other 99% uncultivable is the so-called "Microbial Dark Matter," of which access is mostly only possible through indirect observations or cultivation-free techniques (Amann *et al.*, 1995; Streit *et al.*, 2004; Marcy *et al.*, 2007; Hedlund *et al.*, 2014; Nobu *et al.*, 2014).

With the progress of sequencing technologies and computational methods that deal with data from sequencing, metagenomics has become a frequent procedure for studying the diversity and functionality of microbial communities, revealing a greater variety of organisms

not yet accessed through techniques based on isolation and axeny, and their genomic aspects (Von Mering *et al*., 2007; Christoserdova, 2010; Simon *et al*., 2010).

2.6. Metabolic models of bacterial communities

Metabolic models describe organisms' capacities for transforming matter and energy available in the environment through metabolic processes inferred from their genomic content. Such a strategy is considered "bottom-up" by the Systems Biology discipline since it integrates information from the parts that comprise the system (O'Brian *et al*., 2015). This extrapolation of phenotype from genotype is possible mainly due to the accumulation of knowledge about the functioning and transmission of genetic/genomic information to cellular biochemistry. Such knowledge is the copious product of the Omics Sciences (Henry *et al*., 2010; Sander *et al*., 2015, Thor *et al*., 2017).

The very large amount of omics data that has been routinely generated in biology laboratories worldwide demands not only the maintenance retrieval of such data but also its organization. The association of such data with the cellular activity performed according to such genetic information and metadata to the phenotype allows for the organization of the systems under scrutiny (Franke *et al*., 2005; O'Brian *et al*., 2015). Metabolic modeling is a particular scientific methodology for information integration activity, typically when the amount of information and the systems' complexity makes the integration very difficult, and as such is very appropriate to the act of integrating genomic and phenotypic data. Metabolic models, thus, allows for the comprehension of complex systems, through the organization of omics data.

The scientific community recognizes the central participation of microorganisms in the execution of biochemical transformations and the maintenance of complex systems or

processes such as human health, processes in agriculture, and biogeochemical cycles (Greenblum *et al*., 2012; Biggs *et al.*, 2015; Sander *et al*., 2015; Cook *et al.*, 2017). These activities require a large number of biochemical steps. One or a few organisms hardly could perform all such steps, but often this can occur through communities. The transformations proceed in a shared and synergistic fashion, leading to completion (Cardona *et al*., 2016). Thus, the models of metabolic pathways of communities reflect the degree of complexity of the communities in which they are found, with expected simplifications as to every theoretical model, and kept organized as to facilitate the comprehension of the processes under study (Santos *et al*., 2011; Sander *et al*., 2015).

Metabolic modeling, like all modeling activities, begin with one or a few primary objectives or aspects that must be represented, allowing the systematic abstraction of complicating factors that are naturally part of the system under scrutiny but presumably are of little or no significance for the aspect under observation (Santos *et al*., 2011). Oberhardt and colleagues (Oberhardt *et al*., 2009) point five general and non-exclusive categories of metabolic, biochemical, genomic, or microbiological aspects, which can be elucidated through metabolic modeling, namely: **i**) contextualization of data of high-performance procedures, **ii**) support for metabolic engineering, **iii**) direct discoveries generating *ab initio* hypotheses, **iv**) question multi-species or multi-component relationships, and **v**) discovery of particular emergent properties of metabolic networks (Oberhardt et al., 2009).

From this perspective, the investigation of the structure of the metabolic network of a lignocellulose-degrading microbial consortium from metagenomic data directly benefits the aspects worked on in topics **i**, **iv**, and **v**, but some aspects like **iii** and eventually, in subsequent works, aspects of category **ii** could be promoted or advanced, based on

appropriate initial modeling. This type of modeling also allows an interactive reconstruction of network models, where the proposed model generates predictions and testable hypotheses. It is logically coupled with experiments and new information systematically collected. It feeds the metabolic models and gradually promotes more accurate descriptions of the system's properties by each iterative cycle (Aittokalio *et al*., 2006). Thus, the metabolic modeling of microorganisms and communities is a procedure that provides a favorable framework for the construction of a genotype-phenotype bridge and allows a better understanding of interactions in microbial communities responsible for highly complex and synergistic activities, such as the deconstruction of the biopolymers found in the lignocellulolytic biomass.

3. MATERIAL AND METHODS

3.1. Lignocellulose Deconstructing Bacterial consortium sampling

The bacterial consortia were obtained by sampling a sugarcane plantation soil, following the methods described by Constancio and colleagues (2020). For enrichment, a 500 µl supernatant aliquot was inoculated in a sterile Bushnell Haas Broth (BHB) medium containing 0.5% of autoclaved milled sugarcane bagasse as the only carbon source.

3.2 Adaptation of the bacterial consortium from the field to culture medium containing sugarcane bagasse

The bacterial consortium's adaptation process to the medium containing milled sugarcane bagasse was conducted as follows: the culture suspension was incubated for

seven days (at 30°C) under constant shaking (150 rpm). After this period, a 500 µl aliquot was transferred to a new enriched medium containing 50 ml of BHB+sugarcane bagasse. This procedure was executed for eight weeks, and by each aliquot step, 800 µl of the samples were stocked in glycerol at -80°C. For the proceeding of this work, a sample stored in the freezer from the first week of cultivation was thawed and recovered, then cultivated under the same conditions described above, through 20 weeks (Figure 5).

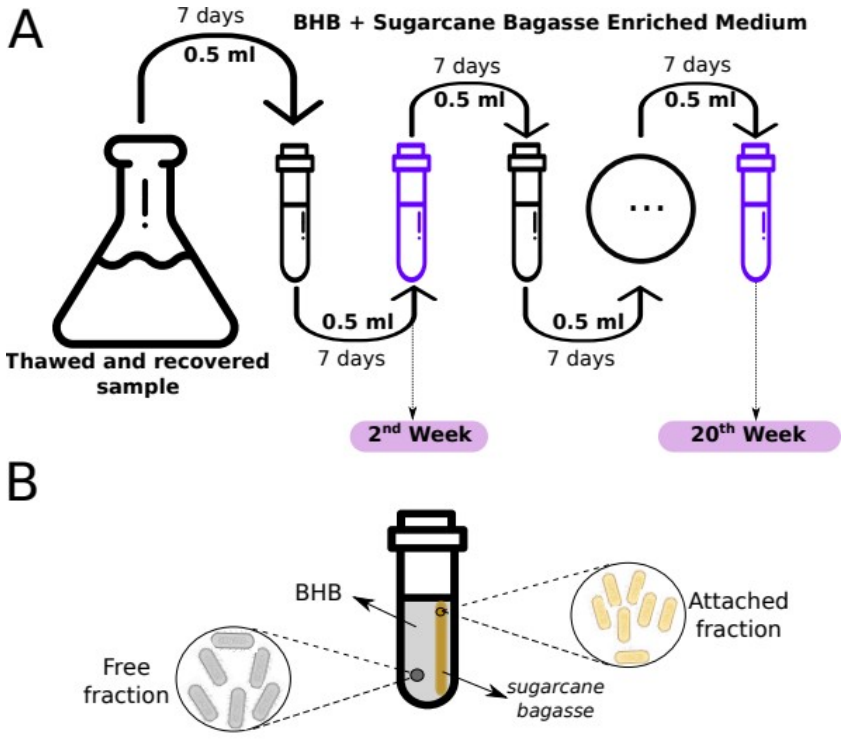

Figure 5. Detailing of the methodology from the adaptation of the consortium, and the time samples, 2nd and the 20th weeks of consortium cultivation (A), and the process to re-cover bacteria that were closely associated (Attached-fraction) and non-closely-associated (Free-fraction) to the sugarcane bagasse fibers (B).

3.3. Extraction of metagenomic DNA and sequencing

We adopted two contrasting time sampling (2nd and the 20th weeks of bacterial consortia cultivation). For both, we recovered the bacteria that are closely associated with the sugarcane bagasse fibers (Attached-fraction) and non-closely-associated to the sugarcane bagasse fibers (Free-fraction) from the medium containing 50 ml of BHB+sugarcane bagasse.

The Attached and Free-fractions of the culture were separated by filtration on Whatman No. 1 filter paper. This procedure was performed using sterile material. The bacteria found in the liquid phase (Free-fraction) were recovered by centrifugation, and the bacteria found attached to the bagasse fibers (Attached-fraction) were recovered by the method that follows: 30 mL of a solution of NaCl (0.85%) containing 0.2% Tween®20 (Polysorbate) was added to the bagasse fibers, then homogenized vigorously by 60 seconds, then centrifuged for 10 minutes under 1,000 Xg, 4°C. The supernatant was transferred to a new tube under ice-cooling. This procedure was executed two more times. After that, glass beads were added to the bagasse fibers, homogenized vigorously, centrifuged for 10 minutes under 1,000 Xg, and 4°C. All the supernatants were collected and transferred to an ice-refrigerated tube, centrifuged at 27,000 g for 30 minutes, 4°C. The supernatant from this step was discarded, and the solid part proceeded to the DNA extraction step. For each week of bacterial consortia cultivation (2nd and the 20th week), 1 ml aliquots from the samples were utilized for DNA extraction, and 800 µl aliquot was stocked in glycerol -80°C.

The total DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corporation), following the manufacturer's instructions. Next, the metagenomic DNA was submitted to gel electrophoresis in the Max Cell EC 360M system, submerged in TBE 1X buffer solution (Tris 89 mM, Boric Acid 89 mM, EDTA 2.5 mM, pH 8.3), in a 0.8% agarose gel, with the addition of 0.5 mg/ml ethidium bromide during ~2 h, under constant

voltage of 90 V. The electrophoretic profile was checked under UV light, and registered on GEL DOC Universal Hood II (BIO-RAD).

The purity of samples was verified on a NanoDrop ND-1000 (Thermo Scientific). The DNA concentration was quantified utilizing the Qubit® Fluorometer (Life Technologies) equipment with Qubit TM DNA kit BR Assay (Invitrogen®), following the manufacturer's instructions.

The total metagenome DNA libraries for sequencing procedures were prepared with Illumina Nextera DNA Library Prep Kit. The sequencing was executed at Illumina® HiSeq 2500 platform, on a Flow Cell v. 4, using HiSeq SBS v. 4 kits (Illumina), 2x100 bp paired-end reads. Sequenced reads quality was evaluated with FastQC 0.11.4 tool, and the removal of adapters was performed with Trimmomatic v. 0.36 (Bolger *et al*., 2014). Only reads larger than 50bp and with PHRED values above 23 were considered.

3.4. Sugarcane bagasse fibers scanning electron microscopy

Four different circumstances were compared: only the fiber-rich sterile medium alone (i) and kept under shaking (150rpm) (ii), and the fiber-rich medium with the community and kept under shaking (150rpm) during the 2^{nd} (iii) and 20^{th} (iv) weeks of cultivation. Both samples were thawed and recovered, then cultivated in BHB medium with sugarcane bagasse for five days. These samples were filtered and separated in the attached and free fractions. Post-fixation was performed with osmium tetroxide (OsO_4) and ethanol dehydration (99% to 10%). The fixed samples were then assembled over the support and coated with gold (20-30 mm). Two sterile mediums were used (30°C, 150 rpm, and only at 30°C) for negative control. Imaging was performed under the scanning electron microscope (Joel JSM6610LV) in a range of magnification from 1,000x to 5,000x.

3.5. Evaluation of the decomposition of lignocellulosic biomass

The analysis of cellulose, hemicellulose, and lignin in sugarcane bagasse fibrous fractions was performed using the Ankom filter bag technique and an Automated Fiber Analyzer (ANKOM Technology, USA). The 2nd and 20th weeks only from the sugarcane bagasse fibers were analyzed. The Neutral Detergent Fiber (NDF), Acid Detergent Fiber (ADF), and lignin content were measured using procedures described by Goering and Van Soest (1970).

For the NDF analyses, 100 ml of neutral detergent solution (30.0 g Sodium sulfate + 10.0 mg Ethylene glycol + 18.0 g EDTA + 6.81 g Sodium borate + 4.56 g Sodium phosphate) were added to 1 L of distilled water with the biomass. For the ADF analyses, 100 ml of acid detergent solution (28.5 ml Sulfuric acid + 20.0 g of cetyltrimethylammonium bromide (CTAB) was added to the biomass. The biomass samples were previously weighed for both analyses, followed by boiling for one hour in a fiber digester (MA- 455 Marconi®). The samples were vacuum filtered and washed three times with distilled hot water and washed two times with pure acetone under ambient temperature. The samples were transferred to a kiln under 105 °C and weighed. All analyses were evaluated in six replicates for the 2nd and 20th weeks of cultivation.

3.6. HPLC and sugar yields in culture medium

Fifteen milliliters of sample supernatants from the 2nd and 20th weeks of cultivation were collected by centrifugation (Sorvall centrifuge at 16,266 ×g for 96 minutes at 4° C) and concentrated (Eppendorf AG 22331 Hamburg Concentrator Plus) to a final volume of 1 mL. All samples were filtered through a 0.45 µM cellulose ester filter and further analyzed by

liquid chromatography on a High-Performance Liquid Chromatography (HPLC) system equipped with a refractive index detector (RID) (Shimadzu, model 100 RID – 10A). Sugar separation was performed by a Supelcosil LC-NH_2 column (25 cm x 4.6 mm) with a constant flow rate of 1 $mL \cdot min^{-1}$ using acetonitrile: H_2O buffer (75: 25, v: v) at 35° C. The sugar yields hydrolysis of lignocellulosic biomass was calculated according to Zhu et al. (2011). All analyses were evaluated in triplicates.

3.7. Metagenomic procedures - binning, quality assessment, and taxonomy designation

The bacterial consortium composition, diversity, and abundance of each metagenome-assembled genome (MAGs) was defined as the MetaWRAP v. 1.2.2 pipeline using standard parameters (Uritskiy *et al*., 2018). Firstly, in the MetaWRAP pipeline, megahit v1.0.6-gfb1e59b (Li *et al*., 2016) was used for the metagenome assembly. Then the binned genomes' quality (completeness and contamination) was estimated using CheckM v.1.4.0 (Parks *et al.*, 2015). GTDBtk v. 0.3.0 (Chaumeil *et al*., 2018), together with Kraken2 v. 2.0.8 (Wood *et al.*, 2014) using the complete GenBank RefSeq Database (O'Leary *et al*., 2016), was used to designate each identified MAGs taxonomically.

3.8. Functional, Metabolic Pathways and Carbohydrate Enzymes Annotation and Analysis

Each MAG was annotated using the RAST server (Overbeek *et al*., 2014) and KEGG GhostKOALA (Kanehisa *et al*.,2016). The dbCAN2 meta server (Zhang *et al*., 2018) and EggNOG v. 5.0 database (Huerta-Cepas *et al*., 2018) were used to search for carbohydrate-active domains in each identified gene. The carbon, nitrogen, sulfur, oxygen, and hydrogen

cycles related genes were identified using the metabolisHMM tool (McDaniel *et al*., 2019). The clusters of genes associated with the secondary metabolism were found using the AntiSMASH tool v.5.0 (Blin *et al*., 2019). The metabolic modeling of the consortia was done using the EnrichM pipeline (https://github.com/geronimp/enrichM).

3.9. Phylogenetic Analysis of the identified MAGs

According to analyses based on the Roary pipeline, the phylogenetic tree was made using the Maximum-Likelihood approach, using the homologous sequences present in most MAGs (Page *et al*., 2015). We adopted the parameters as: i) at least 80% of the genomes should present the common sequence; and ii) 60% of minimum percentage of identity for the blastp step. The sequences were aligned using MAFFT version 7 (Katoh *et al*., 2019). The best-of-fit model for each alignment and Maximum likelihood analyses were performed using IQ-Tree software (Trifinopoulos *et al*., 2016), with 1,000 ultrafast bootstrap resampling (Hoang *et al.*, 2018).

4. RESULTS

4.1 Scanning electron microscopy shows bacteria attached to sugarcane fibers suggesting their role in the deconstruction of lignocellulosic biomass

We evaluated our bacterial consortium's potential to break down sugarcane fibers by scanning electron microscopy approach. Four different circumstances were compared: only the fiber-rich sterile medium alone (i) and kept under shaking (ii), and the fiber-rich medium

with the community and kept under shaking during the 2nd (iii) and 20th (iv) weeks of cultivation (Figure 6). A flat and compact structure was observed without bagasse fiber peels, indicating that the autoclaving process did not interfere with the sugarcane bagasse fiber structure (Figure 6A and B). The material starts to crumble when it remains ten days in mechanical agitation in the sterile cultivation medium. However, it still maintains a compact structure and little peeling (Figure 6C and D). The 2nd and 20th weeks of cultivation (Figure 6E, F, G, and H) showed a deconstruction of the planar and compact structure of the bagasse, the presence of cracks and peeling, and the adhesion of various bacterial types on their surface. This observation indicates that the structure of sugarcane bagasse was modified by cultivation with the consortium, leading to partial fiber disruption, exposing the fibers, and facilitating bacteria's adhesion to hydrolyze the lignocellulosic fractions. Interestingly, distinct bacterial morphological types are observed attached to the sugarcane fibers, suggesting that lignocellulosic deconstruction occurs through different microorganisms. These results strongly suggest that the bacterial consortium might be changing the lignocellulose fiber structure to use it as a carbon source.

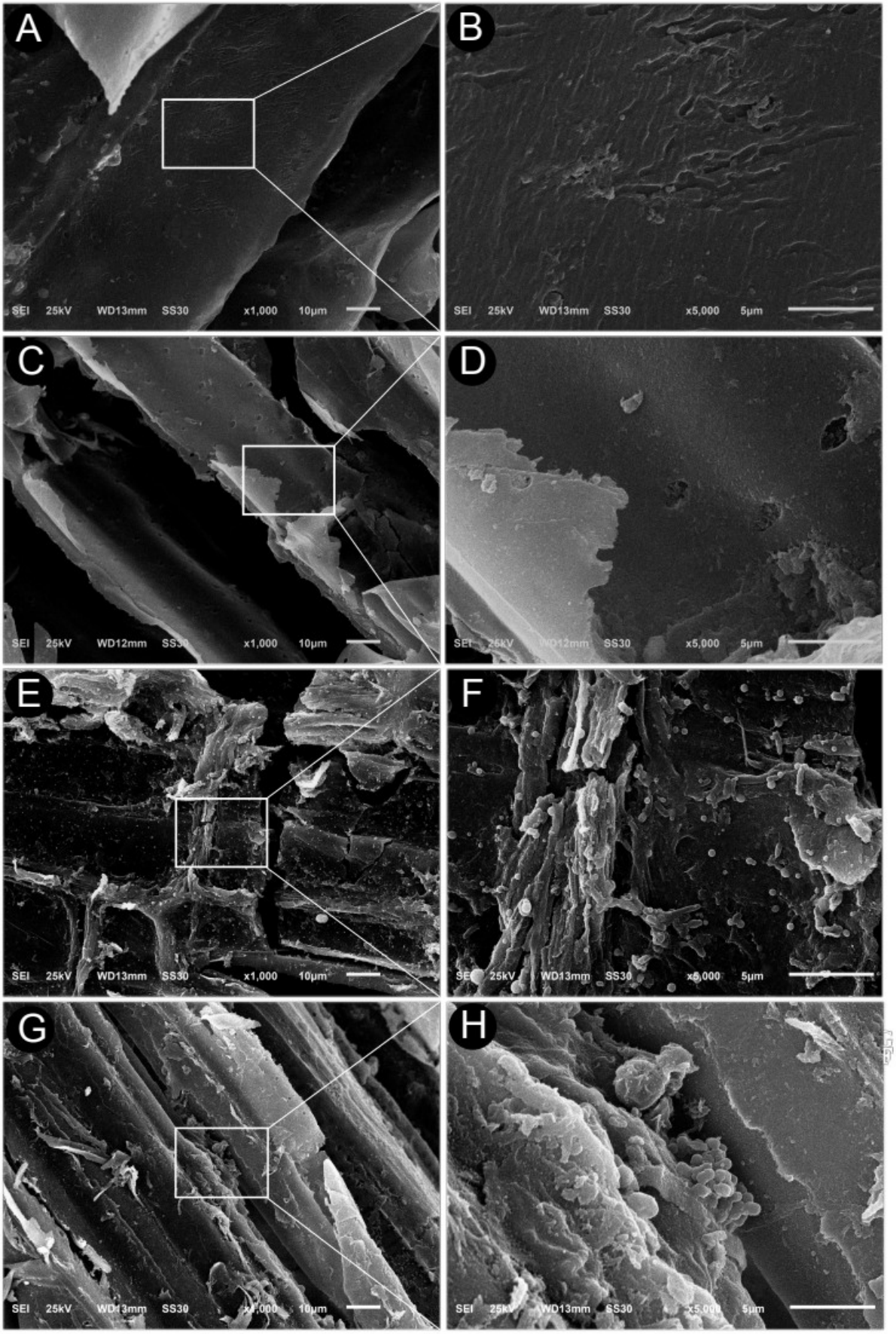
A
SEI 25kV WD13mm SS30 x1,000 10µm
B
SEI 25kV WD13mm SS30 x5,000 5µm
C
SEI 25kV WD12mm SS30 x1,000 10µm
D
SEI 25kV WD12mm SS30 x5,000 5µm
E
SEI 25kV WD13mm SS30 x1,000 10µm
F
SEI 25kV WD13mm SS30 x5,000 5µm
G
SEI 25kV WD13mm SS30 x1,000 10µm
H
SEI 25kV WD13mm SS30 x5,000 5µm

Figure 6. Scanning electron microscopy of the medium containing sugarcane fibers as sole carbon source. **A** and **B** show the sugarcane fibers in sterile culture. **C** and **D** show the sugarcane fibers in agitated and sterile culture for ten days. **E** and **F** show the sugarcane fibers in a culture media from the 2nd-week consortium growing under agitation. **G** and **H** show the s sugarcane fibers in a culture media from the 20th-week consortium growing under agitation. No alteration in the fibers' appearance was observed in **A** and **B**, while **C** and **D** show minimal alteration in the fibers' physical structure. **E**, **F**, **G**, and **H** show an evident alteration in the fibers' structure and colonization by different consortium members. These findings indicate that the consortium is causing flaking, peeling, and overall deconstruction of the sugarcane fibers. **A**, **C**, **E**, and **G** show regions under 1,000x amplification, while **B**, **D**, **F**, and **H** are insets of **A**, **C**, **E**, and **G,** respectively, showing 5,000x amplification.

4.2. Quantification of cellulose, hemicellulose, Lignin, and glucose indicates a dynamic process of lignocellulosic biomass deconstruction.

We further verified the decomposition of lignocellulosic biomass and glucose consumption in the culture medium during the 2nd and 20th weeks of cultivation. Our results indicate that the bacterial consortia can degrade cellulose, hemicellulose, but not Lignin during the 2nd week of cultivation (Figure 7A). Conversely, it was observed the degradation of cellulose, but neither hemicellulose nor Lignin during the 20th week. These findings support the consortium's performance in degrading mainly cellulose. We also checked the glucose availability in the medium and observed that glucose availability during the 2nd week of cultivation is approximately 3.5x higher than in the 20th week (Figure 7B). Moreover, hydrolysis efficiency analysis shows a yield of glucose of 75.6% during the 2nd week and negative values (-36.7%) during the 20th week of cultivation. These results indicate a dynamic lignocellulosic decomposition process, suggesting that the consortium releases glucose more than consuming during the 2nd week and consuming more than releasing from the 20th week.

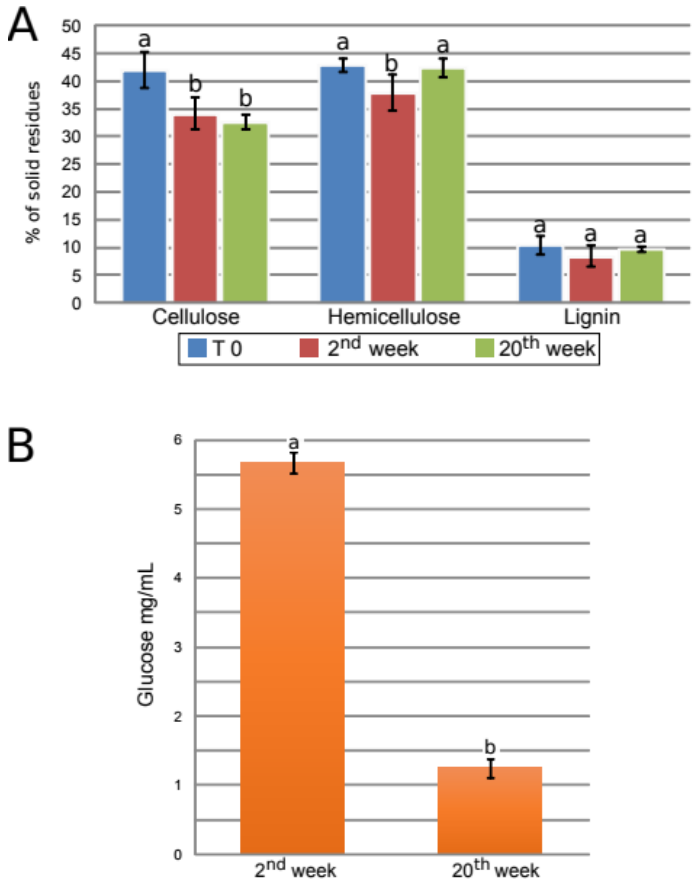

Figure 7. Bar chart comparing the quantities of cellulose, hemicellulose, Lignin, and glucose. **A.** Shows the comparison of cellulose, hemicellulose, and Lignin assayed before (T 0) and after the 2nd and the 20th week of cultivation. Error bars indicate the standard error of six independent biological replicates. **B.** Shows the quantity of glucose assayed at the 2nd and the 20th week of cultivation. Error bars indicate three independent biological replicates' standard error.

4.3. Lignocellulose-decomposing bacterial consortium metagenome sequencing, assembly, and taxonomic profile

On average, 360 Gb of high-quality paired-end reads were generated and assembled for each week of cultivation and respective fractions (Free- and Attached-fractions) (Table 3). In general, each sample was assembled into ~200 Gb contained in more than 130,000 scaffolds and showing an average N50 of 5kb. The average nucleotide identity (ANI) between

the fractions (free and attached) across the 2nd and 20th weeks is ~98.75%, corroborating a near-identical taxonomic composition between the samples.

To evaluate the bacterial consortia's entire taxonomic profile, all samples were concatenated and assembled in 374 Mb with an average GC content of 60%, representing the total bacterial consortia metagenome (Table 3). At least 11 orders derived from 4 phyla were identified in the bacterial consortia metagenome (Bacteroidetes, Proteobacteria, Firmicutes, and Actinobacteria) (Figure 8). Interestingly, slight variations were observed among the 2nd and 20th weeks of cultivation and respective fractions, indicating differential bacterial abundance. For instance, the main differences are related to the significant presence of the Firmicutes phylum (Bacillales order) in the Free-fraction of the 20th week and the prevalence of Actinobacteria (mostly from Micrococcales and Chitinophagales orders) on the 2nd week of cultivation (Figure 8 A and B). It is also worth mentioning that most non-annotated and unknown sequences shown in Figure 8A were derived from contaminants from the sugarcane fibers themselves (~130 Mb). Low coverage assembled contigs (<1,000 bp), representing low abundance and possibly non-essential microorganisms in the consortium, were not considered in our further analyses. Moreover, functional predictions reported incomplete and non-essential pathways related to biomass deconstruction among the unbinned sequences, and thus they were not considered for further analyses.

Table 3. Sequencing and assembly of the lignocellulose-decomposing bacterial community.

	Raw Reads	Reads after trimming	Assembled size (bp)	N50 (Kb) / L50	Longest scaffold (bp)	Scaffolds > 300bp
Free: 2nd week	53,311,191	36,795,258	216,621,114	6,524 / 3,706	1,251,422	141,395
Attached: 2nd week	50,362,156	40,708,747	205,005,671	7,330 / 2,355	1,557,200	133,028
Free: 20th week	50,849,621	35,056,982	198,735,922	3,227 / 7,520	1,211,013	156,975
Attached: 20th week	58,663,806	36,039,348	207,655,023	3,510 / 7,599	479,265	161,807
All (megahit)	213,186,774	148,600,335	374,211,453	12,854 / 3,318	1,904,463	127,034

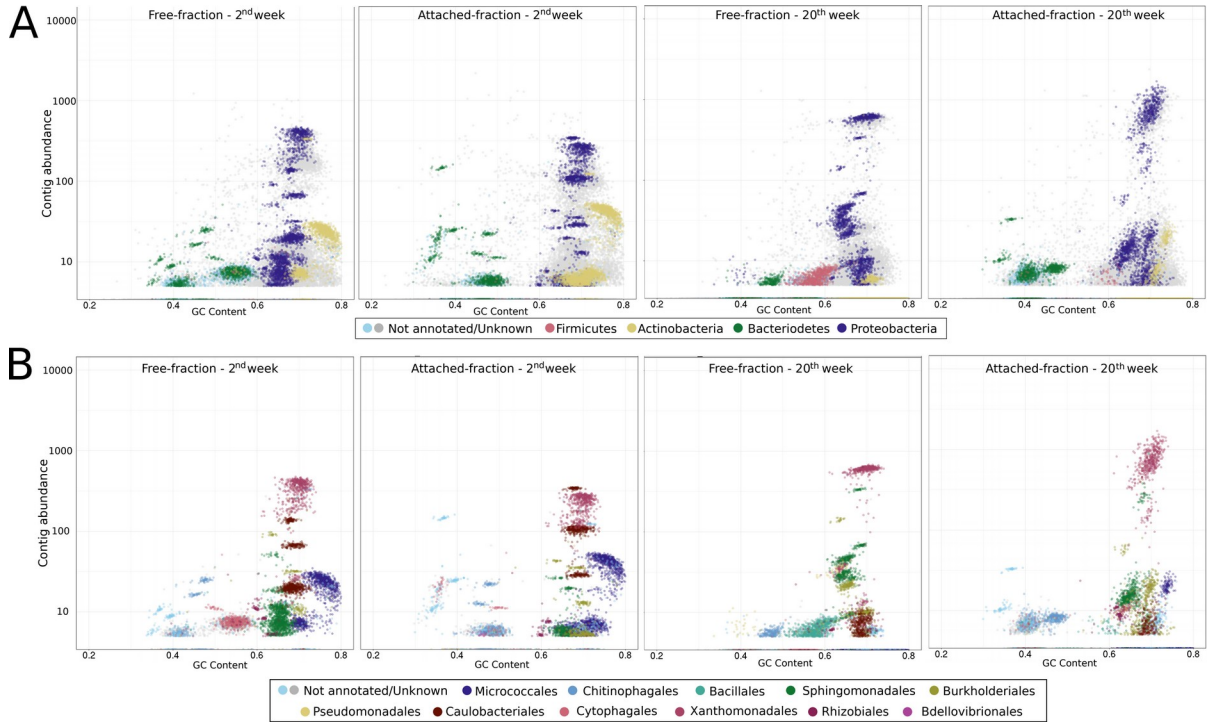

Figure 8. GC vs. abundance blobplots associated with the taxonomy assignment among the 2nd and 20th weeks of cultivation and their associated fractions (free and attached). **A.** Phyllum blob plot, **B.** Order blob plot. All unbinned sequences from the Order blob plot were removed for clarity.

4.4. Total bacterial consortium metagenome binning revealed 52 different species from four main Phyla.

We recovered a total of 52 Metagenome-Assembled Genomes (MAGs), resulting in 240,626 Mbp of total genomic attribution of the metagenomic assembly (mean of 4.63 Mpb for each MAG). The unbinned sequences corresponding to ~130Mb are mostly related to low-quality bins (completeness below 50%, contamination above 20%) and eukaryotic contamination (mainly derived from sugarcane fibers). Moreover, functional predictions report incomplete and non-essential pathways related to biomass deconstruction among the unbinned sequences. Thus for the objectives of surveying the metabolic potential for each MAG, the unbinned sequences were not considered for further analyses.

The MAGs were taxonomically assigned to four main phyla (Actinobacteria [n = 8], Bacteroidetes [n = 14], Firmicutes [n = 2], and Proteobacteria [n = 28]), 8 classes (Actinobacteria [n = 7], Alphaproteobacteria [n = 17], Bacilli [n = 2], Bacteroidia [n = 11], Cytophagia [n = 3], , Gammaproteobacteria [n = 10], Oligoflexia/Bdelovibrionia [n = 1], Thermoleophilia [n = 1] (Figure 9 and Table 4).

Four MAGs from the complete set of 52 MAGs showed an estimated 100% completeness and less than 5% contamination. One of such highly complete and lowly contaminated MAGs, *Chryseobacterium* sp. Bin7 also showed no contamination. Considering the criteria established by Parks *et al*. (2015), we found that 31 (59.6%) of the 52 MAGs in the consortium could be classified as near-complete (over 90% complete), and 16 (30.6%) as substantially complete (between 70% and 90% complete).

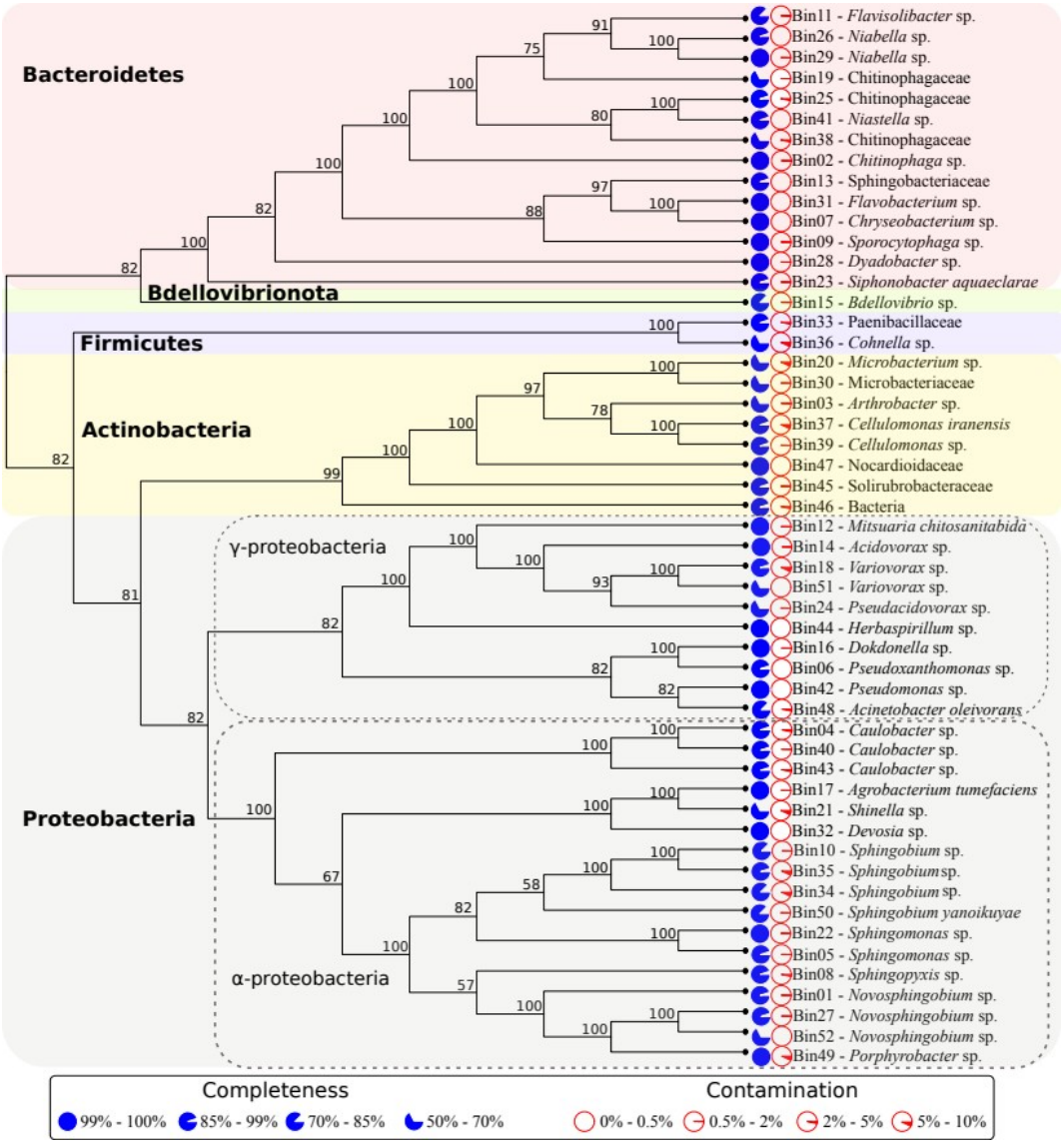

Figure 9. Phylogenetic multilocus Maximum-Likelihood tree showing the MAGs diversity found in the consortium (as total binned metagenome). Names (from GTDBtk analysis) on the branch tips are followed by a blue circle indicating the completeness level and a red circle indicating the contamination level estimated by CheckM.

Table 4. Taxonomic classification and genome completeness of each MAG obtained from the metagenome binning procedure. (continue)

Bin	Phylum	Class	*Closest taxon (GTDBtk)*	*Closest taxon (Kraken)*	Kraken (%)	Compl. (%)	Cont. (%)	Scaff.	N50	Size
1	Proteobacteria	Alphaproteobacteria	*Novosphingobium sp.*	*Novosphingobium sp.*	75	91.26	1.62	44	215778	4841739
2	Bacteroidetes	Bacteroidia	*Chitinophaga sp.*	*Chitinophaga sp.*	60.87	100	1.15	23	817786	7940272
3	Actinobacteria	Actinobacteria	*Arthrobacter sp.*	Micrococcaceae	56.22	61.76	1.16	852	3299	2297904
4	Proteobacteria	Alphaproteobacteria	*Caulobacter sp.*	*Caulobacter sp.*	56.29	93.69	4.33	167	213375	4928801
5	Proteobacteria	Alphaproteobacteria	*Sphingomonas sp.*	*Sphingomonas sp.*	100	96.4	0.85	38	223715	4163301
6	Proteobacteria	Gammaproteobacteria	*Pseudoxanthomonas sp.*	*Pseudoxanthomonas sp.*	63.82	88.26	0.34	246	49426	3613034
7	Bacteroidetes	Bacteroidia	*Chryseobacterium sp.*	*Chryseobacterium sp.*	90.91	100	0	22	770808	5162373
8	Proteobacteria	Alphaproteobacteria	*Sphingopyxis sp.*	*Sphingopyxis sp.*	66.67	93.38	3.18	141	88516	4252535
9	Bacteroidetes	Cytophagia	*Sporocytophaga sp.*	Cytophagia sp.	10.94	100	1.19	64	334426	6632112
10	Proteobacteria	Alphaproteobacteria	*Sphingobium sp.*	*Sphingobium sp.*	56.91	73.61	0.09	123	57144	3014415
11	Bacteroidetes	Bacteroidia	*Flavisolibacter sp.*	Bacteroidia sp.	54.23	79.5	0.57	721	7812	3946228
12	Proteobacteria	Gammaproteobacteria	*Mitsuaria chitosanitabida*	*Mitsuaria sp.*	90	99.22	1.23	40	201288	5853077
13	Bacteroidetes	Bacteroidia	Sphingobacteriaceae	Sphingobacteriaceae	87.1	97.61	0.51	31	471170	5831202
14	Proteobacteria	Gammaproteobacteria	*Acidovorax sp.*	*Acidovorax sp.*	91.3	99.81	1.24	23	411193	4761690
15	Bdellovibrionota	Bdellovibrionia	*Bdellovibrio sp.*	*Bdellovibrio bacteriovorus*	60.2	83.06	0.9	505	10261	3540359
16	Proteobacteria	Gammaproteobacteria	*Dokdonella sp.*	*Dokdonella koreensis*	83.33	99.37	1.99	18	559764	4843315
17	Proteobacteria	Alphaproteobacteria	*Agrobacterium tumefaciens*	*Agrobacterium tumefaciens*	100	99.53	0.55	26	576092	4941267
18	Proteobacteria	Gammaproteobacteria	*Variovorax sp.*	*Variovorax sp.*	99.15	89.72	6.49	118	92476	6007252
19	Bacteroidetes	Bacteroidia	Chitinophagaceae	Bacteroidetes	36.69	64.93	1.27	1390	3694	4340376
20	Actinobacteria	Actinobacteria	*Microbacterium sp.*	*Microbacterium sp.*	84.11	76.7	2.48	214	23019	3141805
21	Proteobacteria	Alphaproteobacteria	*Shinella sp.*	Rhizobiaceae	91.48	52.59	2.59	352	14994	3686717
22	Proteobacteria	Alphaproteobacteria	*Sphingomonas sp.*	*Sphingomonas sp.*	100	99.17	1.11	8	1058682	3922504
23	Bacteroidetes	Cytophagia	*Siphonobacter aquaeclarae*	FCB Group sp.	36.48	92.15	1.05	1028	7330	5469544
24	Proteobacteria	Gammaproteobacteria	*Pseudacidovorax sp.*	Burkholderiales	97.69	70.36	1.81	606	13771	5489212
25	Bacteroidetes	Bacteroidia	Chitinophagaceae	Chitinophagaceae	42.48	87.67	2.11	638	13788	5986324
26	Bacteroidetes	Bacteroidia	*Niabella sp.*	*Niabella soli*	87.55	95.23	0.5	241	31344	4572513
27	Proteobacteria	Alphaproteobacteria	*Novosphingobium A sp.*	Sphingomonadales	100	97.69	0.97	18	322993	3529383

Table 4. Taxonomic classification and genome completeness of each MAG obtained from the metagenome binning procedure. (end)

Bin	Phylum	Class	Order	Family	*Closest taxon (GTDBtk)*	*Closest taxon (Kraken)*	Kraken %	Compl. (%)	Cont. (%)	Scaff.	N50	Size
28	Bacteroidetes	Cytophagia	Cytophagales	Spirosomaceae	*Dyadobacter sp.*	*Dyadobacter fermentans*	88.24	99.7	0.6	17	972536	7826826
29	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	*Niabella sp.*	FCB Group sp.	57.89	100	0.66	76	356462	6316793
30	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	73-13 sp.	Microbacteriaceae	81.66	71.44	1.77	289	9615	2059590
31	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	*Flavobacterium sp.*	*Flavobacterium sp.*	83.33	99.65	0.4	18	1285462	5514676
32	Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosiaceae	*Devosia sp.*	*Devosia sp.*	92.86	99.19	0.2	14	365123	4412235
33	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Paenibacillaceae	*Paenibacillus sp.*	56.62	92.74	2.5	461	28044	8180064
34	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingobium sp.*	Sphingomonadaceae	56.06	78.37	9.73	66	300094	3137218
35	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingobium sp.*	*Sphingobium sp.*	53.49	88.59	5.27	86	146196	3371965
36	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	*Cohnella sp.*	Paenibacillaceae	52.6	83.46	8.9	1114	5418	4468037
37	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	*Cellulomonas iranensis*	*Cellulomonas sp.*	78.71	94.7	0.96	202	42736	3729899
38	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	UBA2791 sp.	Bacteroidia sp.	39.17	74.88	4.93	1103	3533	3330359
39	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	*Cellulomonas sp.*	*Cellulomonas sp.*	64.03	93.11	0.58	139	120147	4301452
40	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*Caulobacter sp.*	*Caulobacter sp.*	79.66	87.85	0.92	290	35555	4300618
41	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	*Niastella sp.*	Chitinophagaceae	89.95	94.95	0.49	172	72074	7632723
42	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas E sp.*	*Pseudomonas putida*	98.65	99.61	0.13	74	124470	5639801
43	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*Caulobacter sp.*	*Caulobacter sp.*	50	96.19	2.82	170	269066	5344511
44	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*Herbaspirillum sp.*	*Herbaspirillum robiniae*	100	99.07	0.1	12	634595	4320850
45	Actinobacteria	Thermoleophilia	Solirubrobacterales	Solirubrobacteraceae	Solirubrobacteraceae	Actinobacteria	67.14	93.85	1	140	52223	4104691
46	Cyanobacteria	Sericytochromatia	S15B-MN24	UBA4093	UBA4093 sp.	Bacteria sp.	88.24	92.31	3.42	34	459146	4281017
47	Actinobacteria	Actinobacteria	Propionibacteriales	Nocardioidaceae	Nocardioidaceae	*Nocardioides sp.*	80	99.48	0	15	1110375	4757382
48	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter oleivorans*	*Acinetobacter oleivorans*	68.92	81.77	3.11	785	5218	3027271
49	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Porphyrobacter A sp.*	Sphingomonadales	94.12	99.42	7.61	17	424016	3651235
50	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingobium yanoikuyae*	*Sphingobium yanoikuyae*	83.22	73.57	1.71	429	13754	3761729
51	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*Variovorax sp.*	*Variovorax sp.*	86.92	50.5	0	107	69134	4085608
52	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Novosphingobium A sp.*	Sphingomonadaceae	74.42	62.7	0	43	170560	2360576

4.5. Species relative abundances and abundances change across the 2nd and 20th weeks of cultivation indicate a dynamic community degrading the lignocellulosic biomass.

The relative abundance of each MAG was estimated based on the reads mapping assignment methods in standard metaWRAP pipeline, and was calculated for each time and fraction (2nd and 20th weeks, Free and Attached). A normalized comparison shows some differences in MAGs abundances between the Free and Attached fractions along the cultivated weeks, thus indicating that our approach was sufficient to separate the Free and Attached fractions and reveal that most of the MAGs were found in both fractions (Figure 10). However, as expected, the results also show considerable changes in MAGs abundances between the 2nd and the 20th week of cultivation. For instance, three MAGs found in the 2nd week in both fractions were absent in both fractions of the 20th week of cultivation: *Arthrobacter* sp. Bin 3, *Cellulomonas iranensis* Bin 37, and Chitinophagaceae Bin 38.

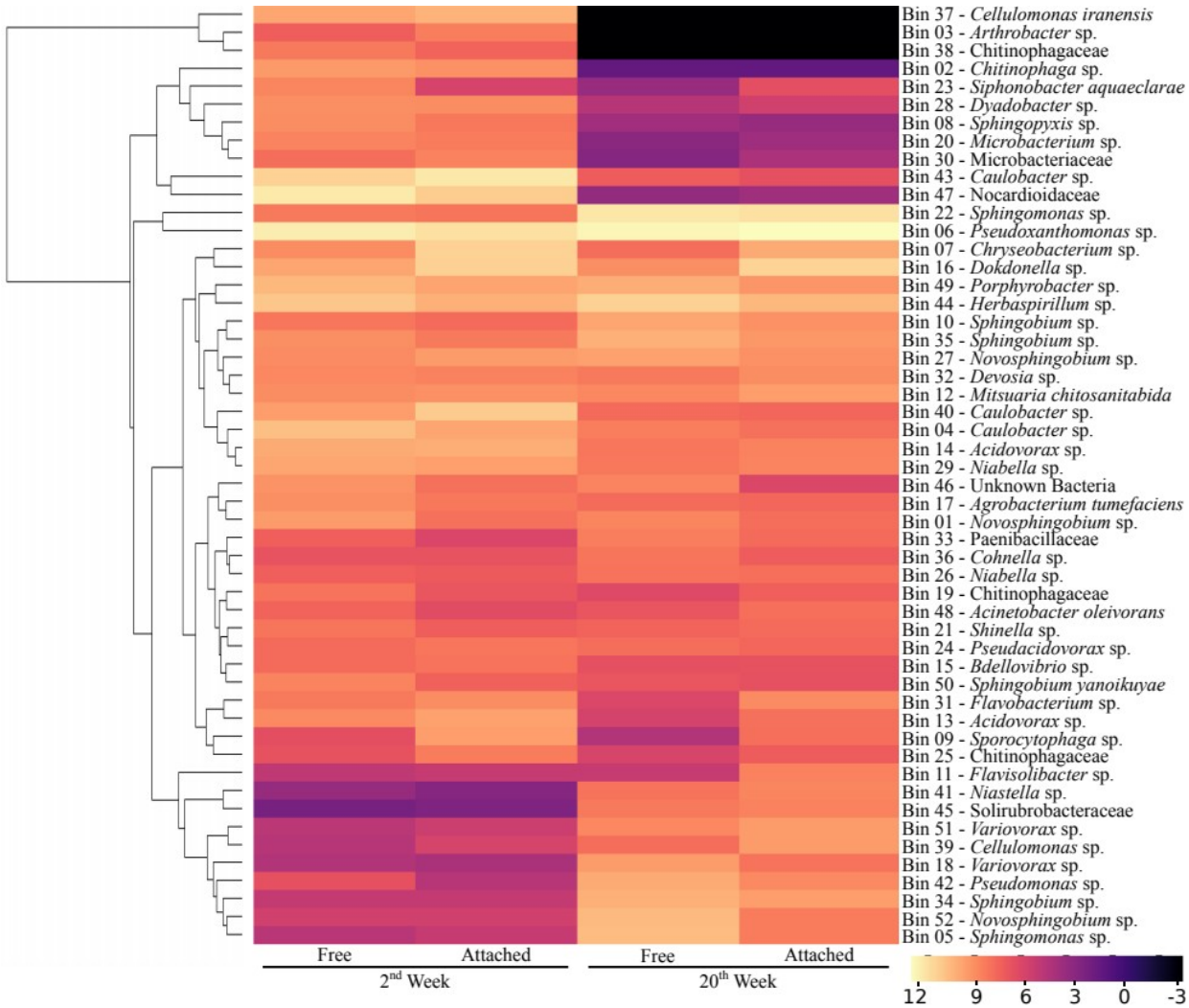

Figure 10. Global abundance heatmap of each MAGs/Bins across the 2nd and 20th weeks of cultivation and their associated fractions (Free and Attached). Lighter colors indicate higher relative abundance concerning the metagenome.

Furthermore, reduction in relative abundance was drastic for *Caulobacter* sp. Bin43 and Nocardioidaceae Bin47. Nevertheless, reduction in abundance was observable but less intense in *Chryseobacterium* sp. Bin7, *Caulobacter* sp. Bin40, *Caulobacter* sp. Bin4, *Acidovorax* sp. Bin 14, *Niabella* sp. Bin29, *Dokdonella* sp. Bin16, *Acidovorax* sp. Bin13, *Chitinophaga* sp. Bin2, *Porphyrobacter* sp. Bin49, *Dyadobacter*

sp. Bin28, *Sporocytophaga* sp. Bin9, *Sphingopyxis* sp. Bin8 and *Microbacterium* sp. Bin20, respectively in decreasing order (Figure 10). Figure 10 and Figure 11 show the same data, although with different perspectives: Figure 10 emphasizing the change in each MAG, while Figure 11 shows the change in each MAG emphasizing the Phylum to which the MAGs belong.

An increase in relative abundance was conspicuous to *Pseudoxanthomonas* sp. Bin6 and *Sphingomonas* sp. Bin22. Some growth was also observed in *Sphingomonas* sp. Bin5, *Novosphingobium* sp. Bin52, *Sphingobium* sp. Bin34, *Herbaspirillum* sp. Bin 44, *Pseudomonas* sp. Bin42, *Sphingobium* sp. Bin35, *Sphingobium* sp. Bin 10, *Variovorax* sp. Bin51, *Variovorax* sp. Bin18, *Cellulomonas* sp. Bin39, Solirubrobacteraceae Bin45, respectively, in decreasing order of magnitude (Figure 11).

The remaining 24 MAGs with species assignments found in these communities showed almost zero change in relative abundance between the 2nd and 20th week of sampling, supporting a potential role related to housekeeping and maintenance or growth as an opportunist in the consortium.

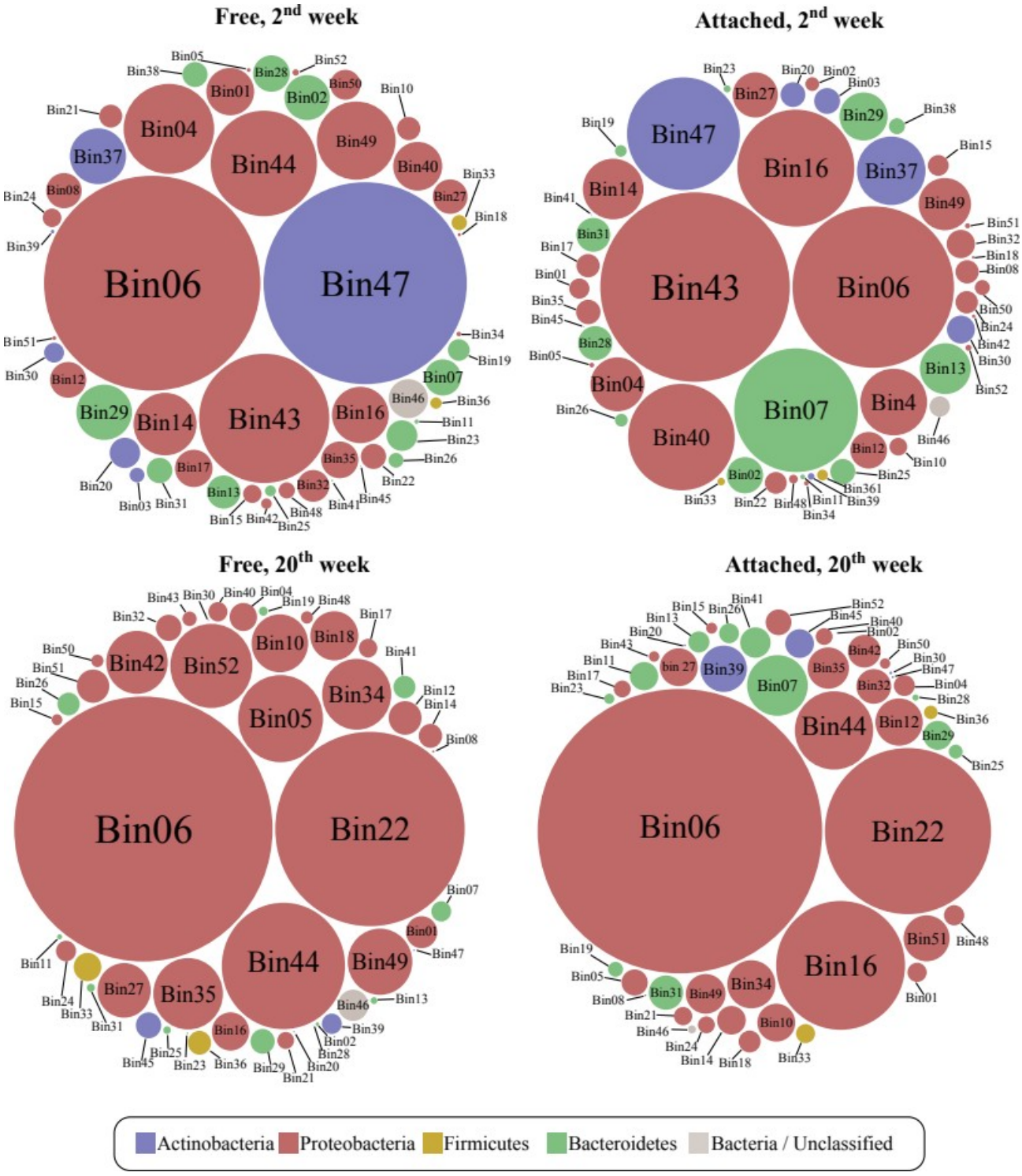
Free, 2nd week
Attached, 2nd week
Free, 20th week
Attached, 20th week
Actinobacteria
Proteobacteria
Firmicutes
Bacteroidetes
Bacteria / Unclassified

Figure 11. Abundance circle plot of each MAGs/Bins from the 2nd and 20th weeks of cultivation and their fractions (Free and Attached).

4.6. Global functional patterns of the lignocellulosic degrading community

A total of 235,594 ORFs were detected among the 52 MAGs. At least 97,018 (41%) were functionally annotated using the KEGG database (Figure 12). Phylum Firmicutes comprised only 2 MAGs and showed a more substantial proportion of KEGG's functionally annotated ORFs related to Environmental information processing (23.06%). Oligoflexia/Bdellovibrionia showed a more significant proportion of category Genetic Information processing (26.64%) in the group's total. However, this group is represented by only one MAG (*Bdellovibrio* sp. Bin15). Excluding these two cases, each category showed approximately similar proportions to the total found in each phylum, suggesting that there is no clear relation between the global metabolic activities indicated by the KEGG annotation by each group at this taxonomic rank.

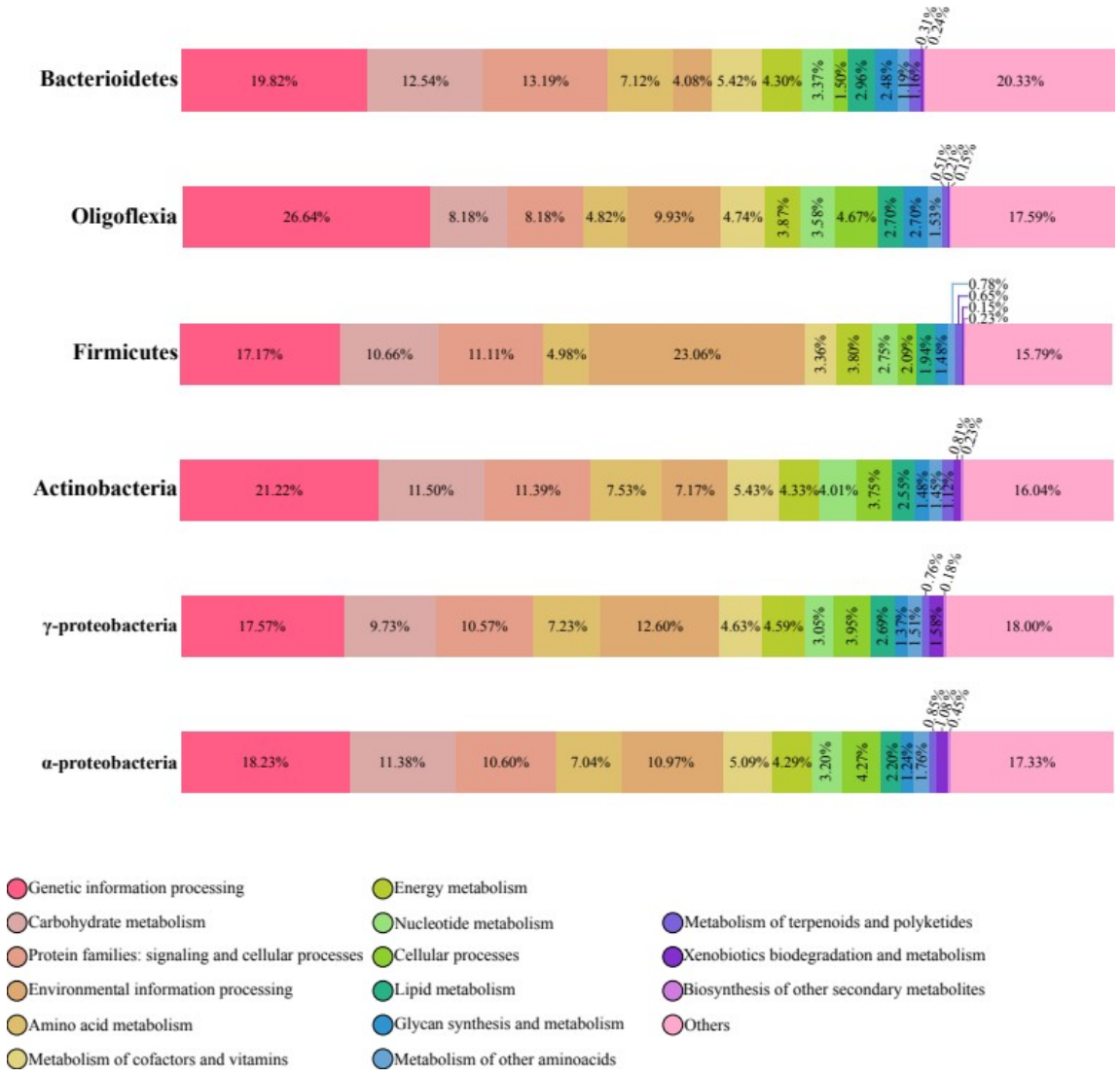

Figure 12. KEGG Annotation of the total metagenome using the GhostKOALA tool. Relative abundances of each KEGG functional category for each Phylum/Superclass identified on the consortia metagenome.

Further, KEGG analyses revealed 133 complete pathways modules in the consortium, including carbohydrate metabolism, energy metabolism (carbon fixation, methane, nitrogen metabolism, and sulfur metabolism, and ATP synthesis), glycan

metabolism, biosynthesis of terpenoids and polyketides, and xenobiotics biodegradation. Moreover, drug resistance modules were also identified (beta-Lactam resistance, *bla* system, multidrug resistance, and efflux pumps).

COG functions were searched over EggNOG annotation to improve knowledge of cellular metabolism of the total metagenome. We found that Transcription processes (K) and Amino acid transport and metabolism category (E) were present in a relatively high proportion (respectively 8% and 7.5%). In contrast, Unknown function (S) is the most frequent COG category (26.5%). Notably, the Carbohydrate transport and metabolism category (G) showed approximately 6.7% of all annotated sequences. (Figure 13).

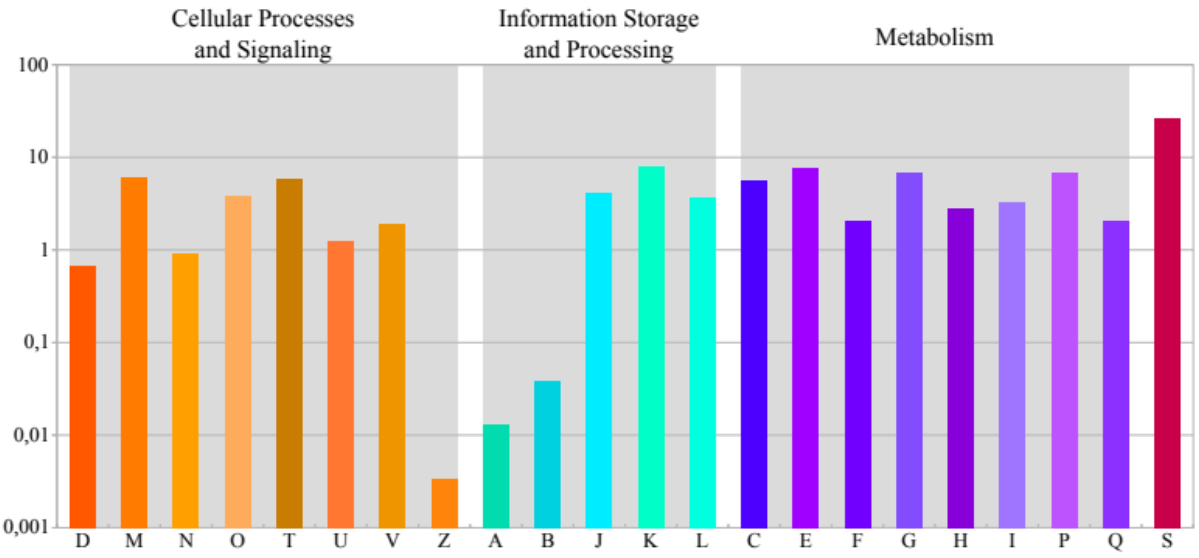

Figure 13. Histogram showing the COG (Clusters of Orthologous Groups) categories found in the total metagenome. The scale is a logarithmic percentage relative to the overall COG found in the total metagenome. Categories are Cellular Processes and Signaling: (D) Cell cycle control, cell division, chromosome partitioning, (M) Cell wall/membrane/envelope biogenesis, (N) Cell motility, (O) Post-translational modification, protein turnover, and chaperones, (T) Signal transduction mechanisms, (U)

Intracellular trafficking, secretion, and vesicular transport, (V) Defense mechanisms; Information Storage and Processing: (Z) Cytoskeleton, (A) RNA processing and modification, (B) Chromatin structure and dynamics, (J) Translation, ribosomal structure and biogenesis, (K) Transcription, (L) Replication, recombination and repair; Metabolism: (C) Energy production and conversion, (E) Amino acid transport and metabolism, (F) Nucleotide transport and metabolism, (G) Carbohydrate transport and metabolism, (H) Coenzyme transport and metabolism, (I) Lipid transport and metabolism, (P) Inorganic ion transport and metabolism, (Q) Secondary metabolites biosynthesis, transport, and catabolism; (S) Function unknown.

4.7. A wide variety of gene clusters related to secondary metabolism suggest plurality in the consortium's ecophysiological interactions.

A total of 223 different genes related to secondary metabolism were found in the 52 MAGs that compose the consortium. Bacteroides class (77 gene clusters and 11 MAGs) and Cytophagia class (30 gene clusters and 3 MAGs), from Bacteroidetes species, were the classes presenting the most of such clusters, 107 clusters in 14 MAGs in total. They were followed by Gammaproteobacteria class (43 gene clusters, 10 MAGs) and Alphaproteobacteria class (40 gene clusters, 17 MAGs), from Proteobacteria species 83 gene clusters 27 MAGs in total. Proteobacteria also includes Oligoflexia/Bdellovibrionia class (3 gene clusters and 1 MAG), increasing Proteobacteria counting up to 86 groups and 28 MAGs. Actinobacteria class (15 gene clusters and 7 MAGs) and Thermoleophilia class (5 gene clusters and 1 MAG), from Actinobacteria phylum, presented 20 clusters in 8 MAGs. Bacilli class (10 gene clusters and 2 MAGs), the only class of Firmicutes species found in the consortium (Table 4, Figure 15).

Table 5. The number of sequences by category of secondary metabolite found in the MAGs by antiSMASH platform for each Phylum and Class. T1PKS and T3PKS are Type I PKS and Type III PKS, respectively.

Phyla	**Actinobacteria**		**Bacteroidetes**		**Firmicutes**	**Proteobacteria**			
Classes	**Actinobac.**	**Thermoleo.**	**Bacteroides**	**Cytophagia**	**Bacilli**	**Oligoflexia**	**α-Proteob.**	**γ-Proteob.**	**Total**
MAGs	7	1	11	3	2	1	17	10	52
NRPS	5	0	17	2	2	0	7	17	50
Arylpolyene	2	0	18	7	0	0	0	9	36
NRPS-Like	3	1	6	3	1	0	5	7	26
Terpene	0	1	6	0	2	0	15	0	24
T1PKS	1	0	5	2	0	0	6	5	19
T3PKS	2	1	7	3	1	1	3	0	18
Siderophore	1	2	5	2	0	2	0	1	13
Bacteriocin	0	0	4	1	2	0	1	1	9
Resorcinol	0	0	5	2	1	0	0	1	9
Trans-AT-PKS-Like	0	0	0	6	0	0	0	0	6
Lanthipeptide	0	0	3	0	1	0	2	0	6
Indole	0	0	1	0	0	0	1	0	2
Aminoglycoside	1	0	0	0	0	0	0	0	1
Protein	0	0	0	0	0	0	0	1	1
Acylamino Acids	0	0	0	0	0	0	0	1	1
PKS-Like	0	0	0	1	0	0	0	0	1
Other	0	0	0	1	0	0	0	0	1
Total sequences	15	5	77	30	10	3	40	43	223

Seventeen different types of gene clusters related to the secondary metabolism were found in the 52 MAGs. Nonribosomal peptide (NRP) was seen as the most abundant (50 clusters, 22.42%), followed by Arylpolyene (36 clusters, 16.14%) and NRPS-Like (26 clusters, 11.66%).

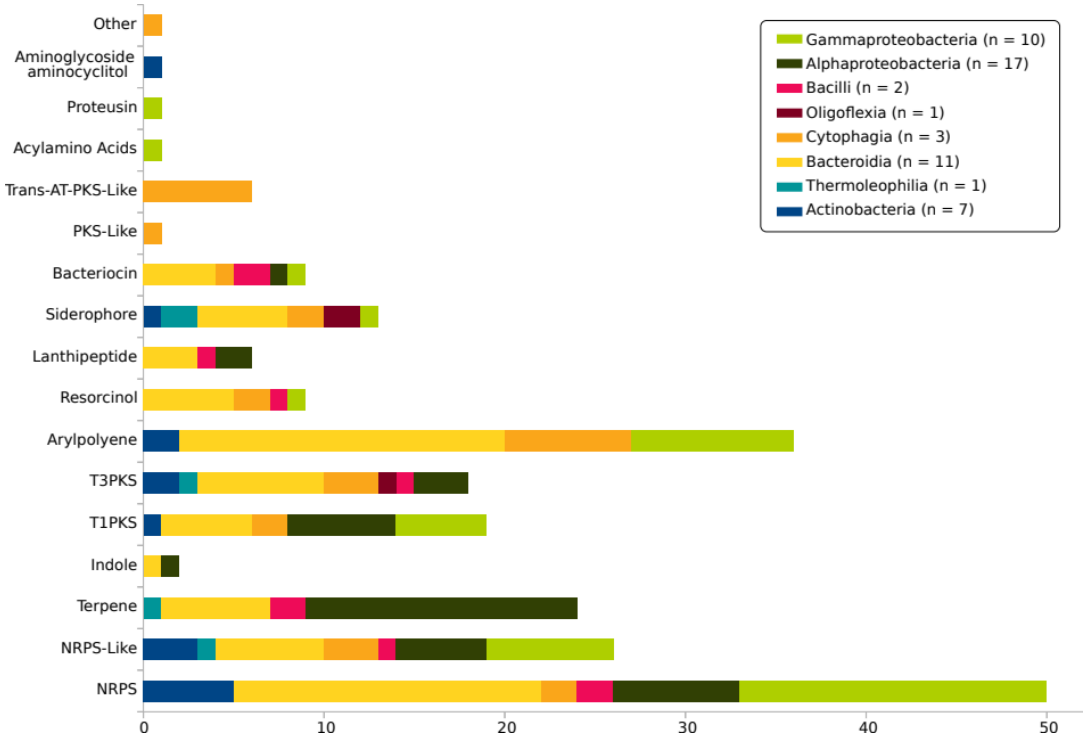

Figure 14. Bar chart showing the abundance of each type of secondary metabolite gene cluster found for each class in the community. Numbers in brackets are quantities of MAGs for each class.

4.8. CAZY enzymes abundance and distribution indicates a synergistic action of each MAG to degrade the lignocellulosic mass

At least 236 different CAZY enzymes families or subfamilies with the potential to participate in the deconstruction of the lignocellulosic biomass were identified in the consortium metagenome (8 AAs, 14 CBMs, 13 CEs, 35 GTs, 146 GHs, and 20 PLs) (Figure 16, 17 and 18).

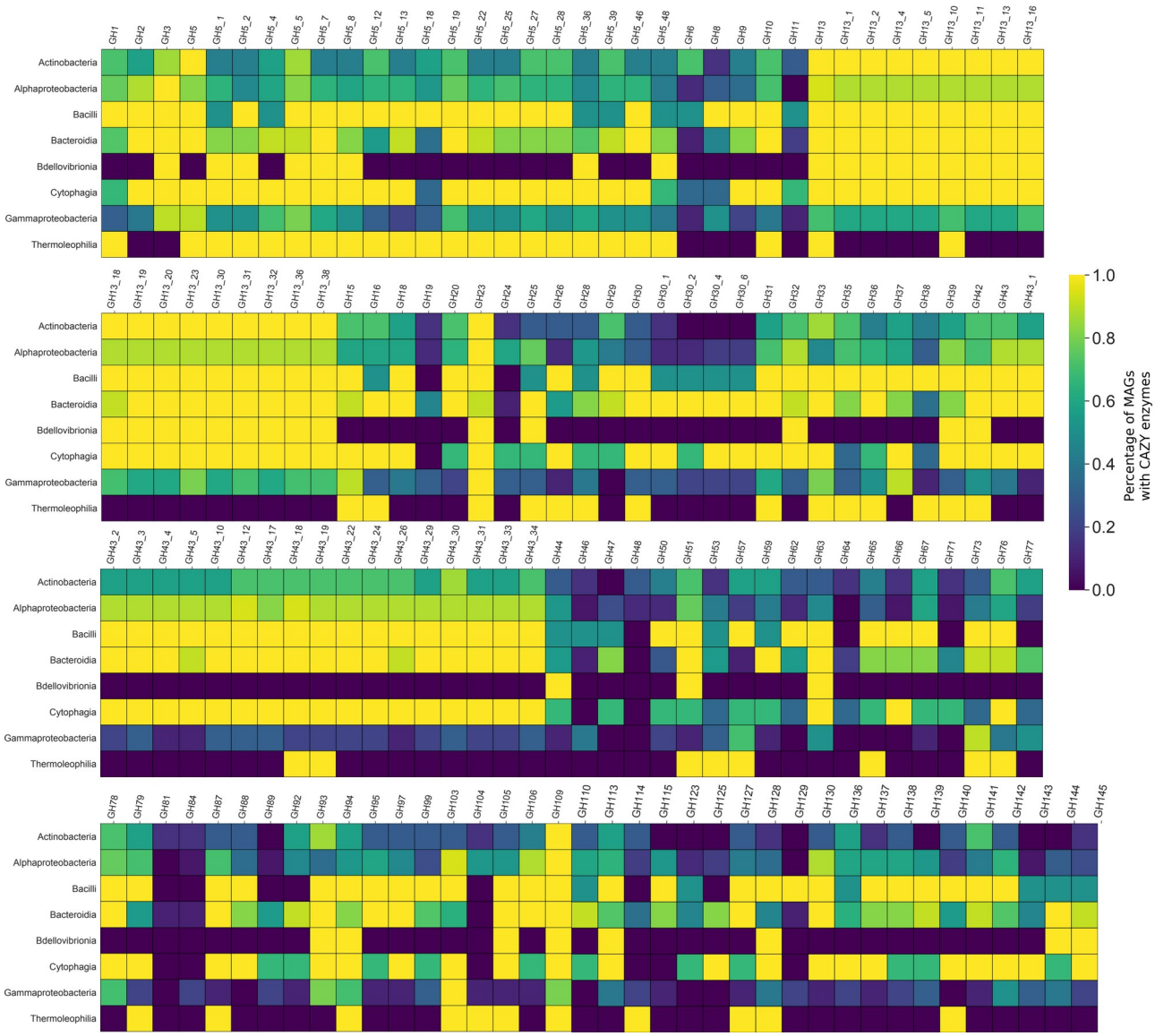

Figure 15. Heatmap showing the abundance of GHs identified in the consortia metagenome in proportion to the total of each category and each taxonomic class.

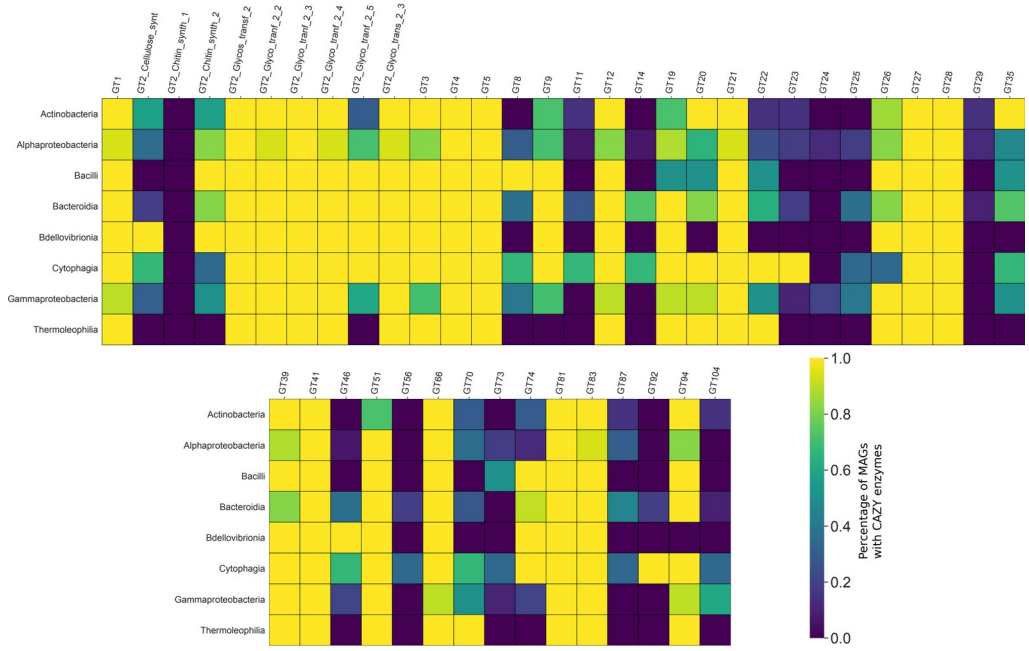

Figure 16. Heatmap showing the abundance of GTs identified in the consortia metagenome in proportion to the total of each category and each taxonomic class.

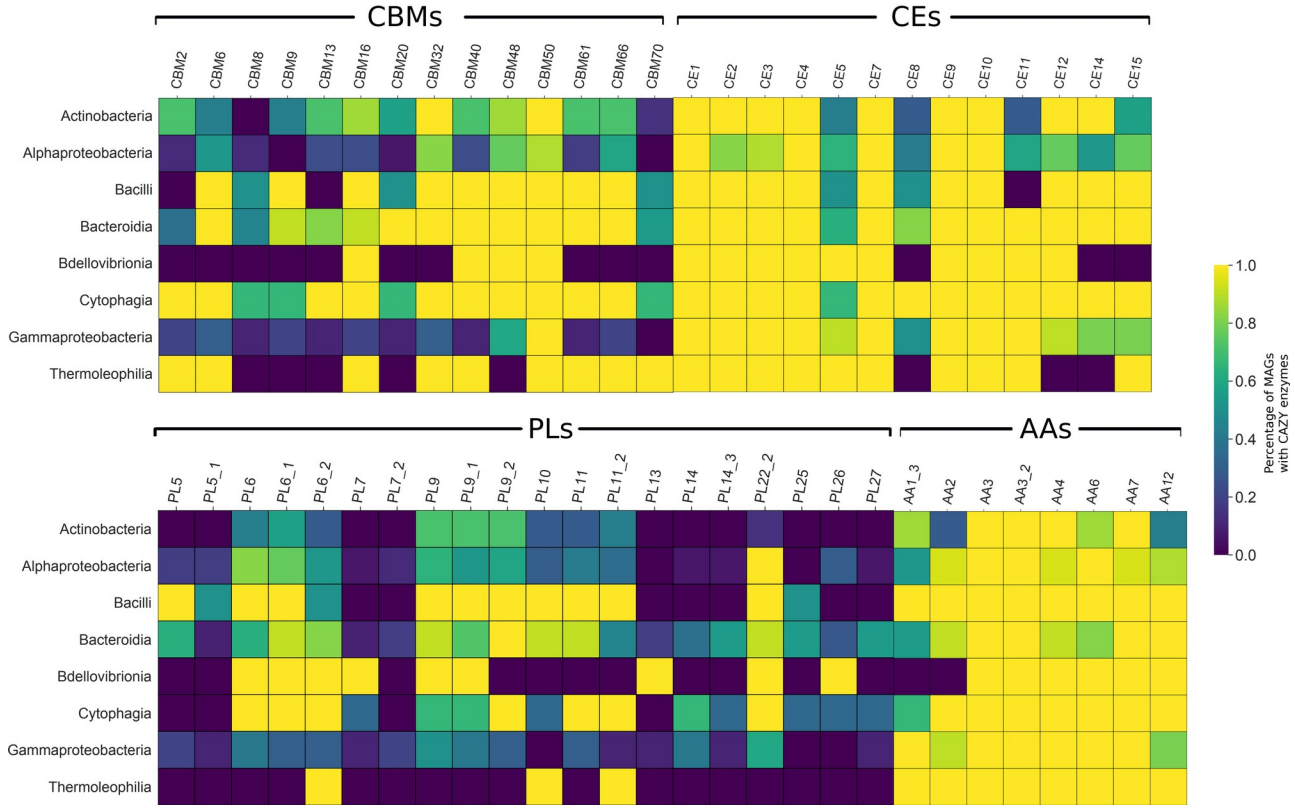

Figure 17. Heatmap showing the abundance of CBMs, CEs, PLs, and AAs identified in the consortia metagenome in proportion to the total of each category and each taxonomic class.

PCA Analyses were performed to clarify the quantitative relationship between many sequences related to the deconstruction of lignocellulose and taxonomy. These analyses indicate differentiation between taxonomic groups and the number of CAZyme sequences and taxonomic groups, and the number of KEGG EC-number sequences (Figures 19A and B, respectively). These results suggest some degree of specificity can be observed in the plotting of the PCAs, which tend to form groups related to the abundance of CAZyme/KEGG EC-number sequences and taxonomy.

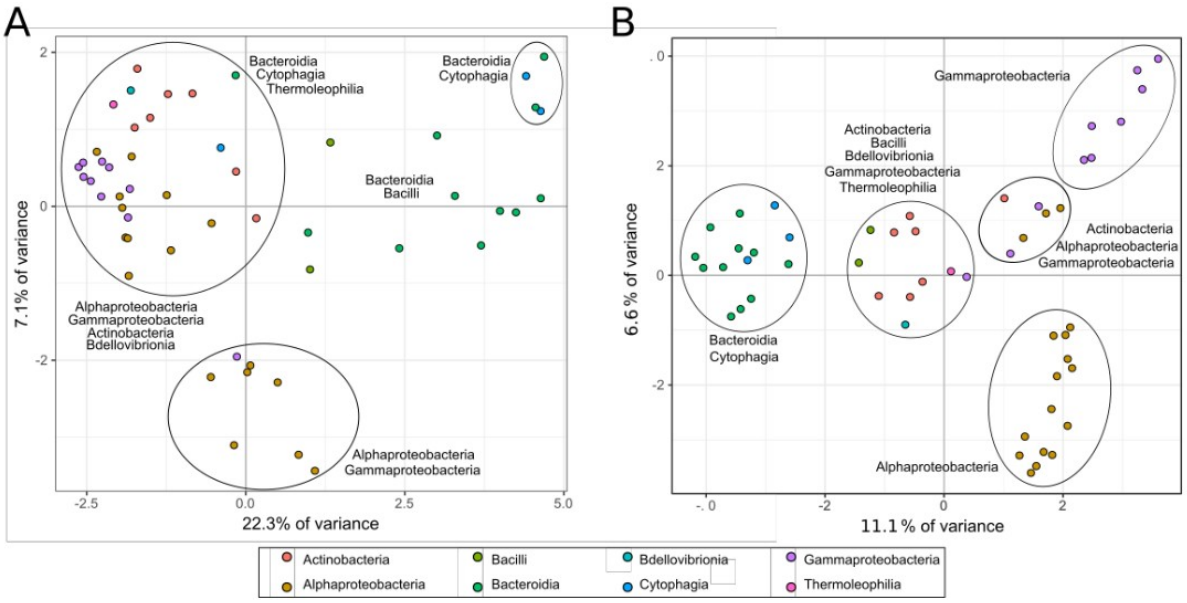

Figure 18. PCA plot showing the relationship between taxonomy (colors) and number variation of sequences identified as indicative of lignocellulose deconstruction. The quantities utilized in the construction of the PCAs are the quantity of sequences found in each genome. Grouping delineated visualy. **A.** CAZymes sequence numbers, **B.** KEGG EC-numbers sequence quantity.

5. DISCUSSION

In this work, we showed that the cultivated consortium can deconstruct lignocellulose. However, the lignin content showed no modification. We individualized 52 genomes with quality high enough to build a metabolic model for each, allowing species' qualitative comparison regarding its capacity to deconstruct saccharidic and lignin polymers, and other aspects of their metabolism. We also observed that the consortium is mostly stable in its dynamics in between the 2nd and the 20th weeks of cultivation, indicating appropriateness for its exploitation in biotechnological and industrial

applications. The overall metabolism of the groups presents a high level of redundancy. This, taken together with the potential of a division of biochemical labor indicates that this consortium is apt for engineering and synthetic biology efforts.

Through scanning electron microscopy imagery, we were able to observe that the consortium can alter the organization and conformation of the fiber, suggesting a possible deconstruction of the lignocellulose. It is inappropriate to affirm that the process of deconstruction is effective only by the use of one measurement (i.e. the images of physical alteration of the fibers), as there is no single physical or chemical characteristic of the lignocellulose that can be used to indicate the effectiveness of enzymatic hydrolysis (Gupta *et al*., 2013). To establish another indication of deconstruction, we proceeded to chemical quantification of the polymers and glucose. We chose glucose as a marker as it is the main monomer of interest for biofuel production (Gutierrez-Rivera *et al*., 2011). We verified that glucose increased in the medium when the bagasse was exposed to the consortium. Also, the reduction of cellulose and hemicellulose content between the 2nd and the 20th weeks of cultivation supports this interpretation. This indicated that not only fibers conformation change but also the chemical composition. Therefore, the consortium was able to deconstruct the lignocellulosic biomass concerning the saccharidic polymers.

In fact, at the 2nd week of cultivation, the community showed a substantial increase in glucose and reduced cellulose and hemicellulose. However, at the 20th week of cultivation, the amount of glucose available was reduced *ca* 3.5 times, when

compared to the 2nd week, and deconstruction of hemicellulose was reduced. This indicate that an intricate system of interactions was under scrutiny.

There was no chemical evidence of deconstruction of lignin, even though most organisms found in this consortium showed properties enabling them to act as ligninolytic. Sequences related to ligninases were found in most genomes comprising the consortium's metagenome. Among these sequences, a resorcinol synthesis gene cluster was found in some genomes, which may be associated with the metabolism of deconstruction of lignin (Brink *et al*., 2019). We speculate that the lack of elicitors, or some other ecophysiological characteristic of the *in vitro* environment, may be the cause of reduced or absent deconstruction of lignin, as previously observed (Xu et al., 2018). Potentially, the time in which the consortium was exposed to the lignocellulosic biomass may not have being long enough to induce the activation of ligninolytic metabolism. Many of the species found (and, among these, some of the most abundant in this consortium) are phylogenetically related to genera known to accomplish this particular process (e.g., *Pseudomonas* sp., *Sphingomonas* sp., *Sphingobium* sp., *Acinetobacter* sp., *Variovorax* sp., *Paenibacillus* sp., *Pseudoxanthomonas* sp., *Chryseobacterium* sp., and others) (Ventorino *et al*., 2015; Beckham *et al*., 2016; Carlos *et al*., 2018; Puentes-Tellez *et al*., 2018; Brink *et al*., 2019; Puentes-Tellez *et al*., 2020).

This work's procedural objective was to construct a metabolic model expressing the differential potential for each organism found in this consortium. For this, we decided to separate (to "bin") each genome. We proceeded to interpret each species metabolic potential, building the metabolic model from this perspective - a Systems Biology's

bottom-up modeling approach (Çakır *et al*., 2014; Henry *et al*., 2016). To achieve a precise observation of such properties, we attempted to accomplish the best bioinformatic procedure balancing: i) the best genome quality assembly for the safest metabolic modeling; ii) the best designation of metagenomic sequencing reads to each genome for the best usage of information available (also improving the safety of prediction through an increase of coverage for assembly).

We extracted 52 genomes from the total metagenome of this consortium. From these 52 MAGs, we recovered one genome estimated to be 100% complete and 0% contaminated - *Chryseobacterium* sp. Bin7 (Bacteroidetes), notably abundant in the 2nd week Attached fraction. It is interesting to note that *Chryseobacterium* genera were already found to deconstruct lignin in other works (Puentes-Tellez *et al*., 2018; Carlos *et al*., 2018). Also, we recovered 31 near-complete and 16 substantially complete genomes, all with estimated low to none contamination, from the total metagenome. These results allowed us to safely infer each genome's metabolic potential. Such a strategy is proposed in Oh et al. (2007) and Henry et al. (2016).

ANI analyses (98.75% of similarity), overall MAG identification, and relative abundance estimations allowed us to observe only marginal differences between the Free and Attached fractions. For instance, 24 MAGs (46.15% of all metagenome's MAGs) showed no relevant modification in relative abundance, thus indicating stability in the consortia during the 18 weeks of cultivated interval (Tzamali *et al*., 2011). Nevertheless, the differences between the 2nd and 20th week were more apparent regarding the species' abundances than species' richness, indicating that the number of

species found in the 20th week was almost the same as the number of species found in the 2nd week. Also, although the community's species showed a mostly stable abundance, some species showed a small change in abundance. *Arthrobacter* sp. Bin3, *Cellulomonas* sp. Bin37 and Chitinophagaceae sp. Bin38 are absent in the 20th week of cultivation, while *Caulobacter* sp Bin43 and Nocardioidaceae sp. Bin47 showed a conspicuous reduction in abundance. *Pseudoxanthomonas* sp. Bin6 and *Sphingomonas* sp. Bin22, both among the most abundant species in all fractions and weeks, showed a substantial increase in abundance. This suggests that stochastic processes may also be relevant in the consortium dynamics - i.e., the most abundant species were kept highly abundant in the consortium, mostly due to their original high relative abundance. Although it is challenging to undeniably circunscribe the influence of stochastic processes in community dynamics such as this, it is broadly recognized that this phenomenon is frequent in bacterial communities and may not be ignored (Jimenez *et al*., 2017).

The Classes found in the consortium presented a highly redundant overall metabolic potential. This may help this consortium's engineering efforts when aiming to improve biotechnological interest (Brenner *et al*., 2008; Tzamali *et al*., 2011; Puentes-Tellez *et al*., 2018). We inquired if this aspect had implications on the consortium's lignocellulolytic aspect as well, and in which taxonomy level it was more relevant. Two PC analyses, one contrasting taxonomic classification, quantities, types of KEGG E.C. Number sequences, and other contrasting taxonomic classification and amounts and types of CAZymes sequences present in each MAG showed a clustering pattern. These

results also suggested a taxonomic specialization of potential metabolic capacities, supporting a taxonomic DoBL concerning the lignocellulosic biomass's deconstruction (Brenner *et al*., 2008).

PCA allowed speculation that some taxons may show more in-group similar potential lignocellulosic deconstruction capacities in the consortium if considering variation in such sequences of various types as indicative of specificity to the whole process of deconstruction. Alphaproteobacteria, Gammaproteobacteria, and Oligoflexia/Bdelovibrionia classes tend to group (all belonging to Proteobacteria Phylum). In the same fashion, Cytophagia and Bacteroidia classes (belonging to Bacteroides Phylum) tend to group, while separated from Proteobacteria Phylum groups – and also it is the case for Actinobacteria and Thermoleophilia classes (Actinobacteria Phylum).

On the other hand, it is relevant to consider that the relatively low eigenvalues of both PCAs indicates a weak dimension representation, possibly due to our data's high dimensionality (quantities of hundreds of different types of sequences, for each of many genomes). This may suggest that the decomposition of this data into each family of CAZymes and grouping the MAGs into Classes is stringent to this analysis. In other words, we lose information regarded to the Division of Biochemical Labor (DoBL) when dividing the shared quantities of sequences, as many sequences indicate the same (or very close) biochemical activity potential over the lignocellulose polymers. This is corrected when we pool the sequences and compare each MAG data in overall activities (CAZymes related to Ligninases and CAZymes related to Hemicellulases and

Cellulases). However, it points to each group's specificity through the clustering – the taxonomic groups are more similar between themselves in their capacities to deconstruct lignocellulose than to other groups. It is also expected to observe some overlapping between groups, considering that these enzyme gene sequences' classification schemes are not comprehensive (in opposition to a taxonomic classification of sequences, e.g.). The relationship between the number of KEGG EC-number and taxonomy results in more clearly defined groups than the relationship between the CAZyme sequences and taxonomy. Taken together, these results clearly indicate that the consortium shows a taxonomy-defined DoBL to achieve the deconstruction of the lignocellulosic biomass, even though many reactions are shared between many groups.

Although a broad spectrum of action, GH is a family of enzymes that may catalyze the glycosidic bond's hydrolysis between carbohydrates (Henrissat *et al*., 1991). Family GH13 was found with very high frequency in almost all classes, except for Gammaproteobacteria and Thermoleophilia (less than 60%). Alphaproteobacteria also showed a slightly lower rate than most groups, albeit still high frequency of this gene family in its constituting MAGs (around 80%). GH13 is a significant family coding for enzymes that act over substrates presenting α-glycoside linkages (Svensson *et al*., 1994). This indicates the broad spectrum of activities found in this family and justifies the various subfamilies that constitute GH13; many found in this consortium in high frequency in the classes, and potentially include actions such as amylases, pullulanase, neopullulanases, glucosidases, and going as far as amino acid transport (Glycoside

Hydrolase Family 13, http://www.cazypedia.org/). Considering such aspects of this family and the pattern of abundance for each class, we speculate a frequent potentiality for the bacterial groups to act over α-glycoside linkages found in cellulose and hemicellulose polymers of the substrate. However, the specific potential activities are challenging to determine *in-silico* without a detailed experimental analysis of each sequence considering the wide variety of substrates that can be acted upon by the enzymes found in this family.

GH23 family was found as a high-frequency enzyme family among all classes of this consortium, encompassing lytic transglycosylases and cleaving the β-1,4-linkage N-acetylmuramyl and N-acetylglucosaminyl on peptidoglycan substrate (Blackburn *et al*., 2001). GH43 enzyme family shows a wide diversity of structures, and their activities are known to encompass α-L-arabinofuranosidases, endo-α-L-arabinanases, and β-D-xylosidases (Flipphi *et al*., 1993). Such a broad range of specificities justifies the various divisions of this enzymatic family. In the consortium, we observed a high frequency of this family occurrence in most classes, although absent in the only MAG of the Oligoflexia/Bdellovibrionia class (Bdellovibrio sp. Bin15). GH43 enzyme family showed as being in rare occurrence frequency in the Gammaproteobacteria class (less than 50%) and somewhat frequent in the Alphaproteobacteria class (around 80%). This indicates that most classes found in this consortium showed at least some potential capacity to deconstruct hemicellulose, except Oligoflexia/Bdellovibrionia. Also, the Proteobacteria group showed a less pronounced potential for this activity when compared to the other groups. GH109 is an enzyme family that encompasses the α-N-

acetylgalactosaminidase activity and is observed to be found often (> 80%) in all classes of this consortium (Liu *et al*., 2007). Although this family may present activities over glucose-derived substrates, substrates other than α-N-acetylgalactosamine are still unknown but only predicted (Liu *et al*., 2007).

GTs (Glycosyl Transferases) are enzymes that catalyze the transfer of saccharide moieties products in oligosaccharides and polysaccharides in mechanisms of retention or inversion of the substrate (Sinnott *et al*., 1990). GT1 is an enzyme family with a broad spectrum of activities, evidenced by the large quantity of EC variations found in this family (EC 2.4.1.-, encompassing .17, .40, .45, .47, .91, .115, .120, and many others) (http://www.cazy.org/GT1.html). This family comprises essential enzymes responsible for the first step in deconstructing branched glycan polymers (http://www.cazy.org/GT1.html). It is also found in very high frequency in all classes found in the consortium. GT2 family is composed of very well-known enzymes like cellulose synthase (EC 2.4.1.12) and chitin synthase (EC 2.4.1.16), but also less critical enzymes, as N-acetylglucosaminyltransferase (EC 2.4.1.-) and others (http://www.cazy.org/GT2.html). The family is characterized by showing an inverting transfer of saccharide moieties (Tarbouriech *et al.*, 2001). In the context of this consortium, many subfamilies (GT2-2, GT2-2_2, GT2-2_3, GT2-2_4) of this family were found in high frequency in all classes, suggesting an important role, potentially related to the deconstruction of the biomass through the transference of moieties.

GT3, GT4, and GT5 were families found in high frequency in all classes. GT3 (EC 2.4.1.21) is a bacterial glycogen synthase, an enzyme relevant to the bacterial carbon

metabolism, showing potential relevance to the deconstruction of lignocellulose (https://enzyme.expasy.org/EC/2.4.1.21). GT4 is a family that offers a wide diversity of activities in the carbon metabolism in bacteria and other groups. As such, it may be indirectly relevant to the deconstruction of lignocellulose (http://www.cazy.org/GT4.html). GT5 also shows a potentially indirect relevance to the biomass's deconstruction. It may act as glycogen synthase or a few other activities related to that mechanism of transference of saccharide moieties (http://www.cazy.org/GT5.html). In the same fashion, GT51 (EC 2.4.1.129, murein polymerase), GT83 (EC 2.4.2.43), GT94 (EC 2.4.1.251) are enzymatic GT families found in high frequency in all classes and are related to the bacterial metabolism, and thus most possibly associated indirectly to the deconstruction of the biomass substrate (http://www.cazy.org/GT51.html, https://www.enzyme-database.org/query.php?ec=2.4.1.129).

CBMs (Carbohydrate-Binding Modules) are amino acid sequences within an enzyme sequence harboring a well-defined carbohydrate-binding activity (Baraston *et al*., 2005). CBM2 is a modular enzymatic family found in all Cytophagia and Thermoleophilia. It is in high frequency in Actinobacteria classes, but less often found in Alphaproteobacteria and Gammaproteobacteria classes, and absent in Bacilli and Oligoflexia/Bdellovibrionnia classes in this consortium. This family of enzymes participates in the solubilization of cellulose and, in less frequent cases, hemicellulose, assisting the effectiveness of other catalytic regions of the same peptide (Gilkes *et al*., 1988; Black *et al*., 1996). This may indicate the relevance of Cytophagia, Thermoleophilia, and Actinobacteria in the deconstruction of cellulose by this

consortium. In most cases, CBMs may promote the effectiveness of the other accompanying catalytic regions on the peptide, synergistically deconstructing the substrate to which the CBM shows affinity (Cantarel *et al*., 2009). As a counterexample, CBM50 is a family found in high frequency in all classes of this consortium, showing binding activity to bacterial cell walls, particularly to N-acetylglucosamine residues.

Alphaproteobacteria and Gammaproteobacteria classes showed an overall low frequency of all CBMs found in this consortium (except for CBM50). In opposition, Bacteroidia, Bacilli, and Cytophagia classes showed an intermediate to a high frequency of CBM family sequences in their compounding MAGs, which indicates their substantial relevance to the potential deconstructions of the biomass through CBM deconstruction processes.

CEs (Carbohydrate Esterases) catalyze the acylation of substituted saccharides (Cantarel *et al*., 2009). As CEs act over acylated moieties of polysaccharides, the enzymes promote the deconstruction of the lignocellulosic polymers by allowing the access of GHs (Christov *et al*., 1993). Overall, these families' sequences were found in high frequency in all classes that comprise the consortium, with few exceptions. In particular, CE5 (rate of 50%-100%) shows the activity of hydrolysis of acetylated moieties in polymeric xylan, acetylated xylan, and glucose, a potentially relevant process to the deconstruction of lignocellulose in this consortium. On the other hand, we found CE8 (frequency of 0%-100%), a pectin methylesterase (EC 3.1.1.11), connected only to the deconstruction of the pectin fraction of the biomass (Sarmiento *et al.*, 2017). Pectin is absent, or almost absent, in Sugarcane lignocellulosic biomass. Thus, CE8 seems to

have little relevance to the global deconstruction in our in-vitro cultivation, although this indicates some community' metabolic versatility.

PLs (Polysaccharide Lyases) are enzymes that cleave polymers containing uronic acid, resulting in an hexenuronic acid residue and a reducing end (Chakraborty *et al*., 2017). PLs were found sparsely in the consortium's groups, being PL6 and PL22_2, the two most abundant ones (varying from 0% to 100% depending on the class), suggesting the indirect or partial relevance of this enzymatic family to the shared metabolism of the consortium. Bacilli class showed a comparatively elevated frequency. Bacteroides class also showed a higher frequency. Gammaproteobacteria and Thermoleophilia showed very low to no sequence of this type of enzyme.

AAs (Auxiliary Activities) are redox-active enzymes that may be involved in lignin deconstruction, allowing the GH, PL, and CE families of enzymes to reach the saccharidic polymers in the biomass (Levasseur *et al.*, 2013). The AA enzymatic family was found in comparatively high frequency in all consortium classes, suggesting that most MAGs can participate in the deconstruction of lignin. This enzymatic family encompasses enzymes that can oxidize phenolic substrates, internally or externally, to the cell (Levasseur *et al.*, 2013). The AA1 family is also known as Laccase (EC 1.10.3.2), which indicates the ability to degrade lignin. In the consortium, this enzymatic family is found in 100% of Bacilli, Gammaproteobacteria, and Thermoleophilia MAGs, and is absent in Oligoflexia/Bdellovibrionia.

Considering the generalities required when using superior taxonomic classification (i.e., Phyla, Superclasses, Classes, Orders), we decided to express our

qualitative metabolic model at the genome level (i.e., Species assumed from the MAGs separated from the metagenome). It seems relevant to consider that even strains of the same species vary genome structure, gene content, and metabolic potentialities (Kettler *et al*., 2007; Willis *et al.*, 2011). This offered another difficulty, consisting of a considerable amount of information to be condensed into the model – many CAZyme families with widely varying amounts of knowledge about substrate specificity, subcategories of each, and modes of action. Considering this problem, we choose to pool sequences of enzymes into two categories: i) sequences of enzymes related to the deconstruction of saccharidic polymers and oligomers (Hemi/Cellulases), and ii) sequences of enzymes associated with the deconstruction of other aromatic polymers – Ligninases, in particular.

The pooling of all the CAZyme families showed differences in the relevance of each Class found in the total metagenome. Also indicated further particularity of each species of the consortium on the DoBL related to the deconstruction of lignocellulose, as is evident in the model of metabolic potential proposed. The central premise adopted in the metabolic model proposed is that the quantity of sequences related to each polymer's deconstruction found in the bagasse is proportional to the species' relevance to the effectuation of such reaction in the process of deconstructing lignocellulose – i.e., more sequences, more relevance. This knowledge can help inform the manipulation of this consortium in respect of selecting species or groups more relevant to any particular step in the process of deconstruction of lignocellulose, potentially improving the consortium’s efficacy and efficiency (Brenner *et al*., 2008; Tzamali *et al*., 2011).

To improve and clarify the understanding of this consortium's potential activities regarding the deconstruction of the lignocellulosic biomass, we presented an overall model of its components. The proposed model aims for a detailed classification of each MAG found in the consortium as a participant of the DoBL, evidenced in previously presented results, and ordered by taxonomic classification (Figure 19). The CAZyme families were pooled as Ligninases or Hemicellulases and Cellulases, and counted. These sums were used as markers of each MAG and Class's potential participation in the DoBL. Cellulases and Hemicellulases (Hemi/Cellulases) were pooled together, as many oligomers and monomers found in the cellulose polymer are also found in hemicellulose polymer deconstruction. So the enzymes can act over both polymers.

Bacteroidia, Cytophagia (both Bacteroidia Phylum), and Bacilli (Firmicutes Phylum) are the classes in which we can find most MAGs with the higher potential to deconstruct Hemicellulose and Cellulose. *Niastella* sp. Bin 41 (Bacteroidia) shows a notably high potential to deconstruct hemicellulose and cellulose, presenting more than 1,800 domain sequences linked to this process, followed by Paenibacillaceae sp. Bin 33 (Bacilli, 1,476 sequences) and *Dyadobacter* sp. Bin 28 (Cytophagia, 1,446 sequences). *Niabella* sp. Bin 29 (Bacteroidia), *Flavobacterium* sp. Bin 31 (Bacteroidia), Sphingobacteriaceae sp. Bin 13 (Bacteroidia), Chitinophagaceae sp. Bin 19 (Bacteroidia), *Chitinophaga* sp. Bin 2 (Bacteroidia), *Niabella* sp. Bin 26 (Bacteroidia), and *Siphonobacter* sp. Bin 23 (Cytophagia) shows more than 1,200 domain sequences related to Hemi/Cellulases, representing their relevance to this process.

Gammaproteobacteria is the class with most MAGs showing higher potential to deconstruct lignin. The 3 MAGs with higher potential in the consortium belong to this class: *Variovorax* sp. Bin 18, *Pseudoacidovorax* sp. Bin 24, and *Pseudomonas* sp. Bin 42, all were presenting more than 115 domain sequences. Also, in this class, *Variovorax* sp. Bin 51 showed 99 domain sequences. In Class Actinobacteria, Nocardioidaceae sp. Bin 47 showed notably high potential, reaching 105 domain sequences. Only four more MAGs showed more than 80 domain sequences, namely *Agrobacterium tumefaciens* Bin 17 (Alphaproteobacteria, 89 sequences), *Shinella* sp. Bin 21 (Alphaproteobacteria, 83 sequences), *Dyadobacter* sp. Bin 28 (Cytophagia, 83 sequences), and Unknown sp. Bin 46 (Unknown, 80 sequences).

The model shows that although there is some above-species grouping concerning potential participation in DoBL, this potential is also relevant to species level. Thus, DoBL seems to be species-specific. Although many steps in the deconstruction of the lignocellulosic biomass may be shared between species or above-species groups, at least some steps depend on fewer species for each polymer type in the lignocellulosic biomass.

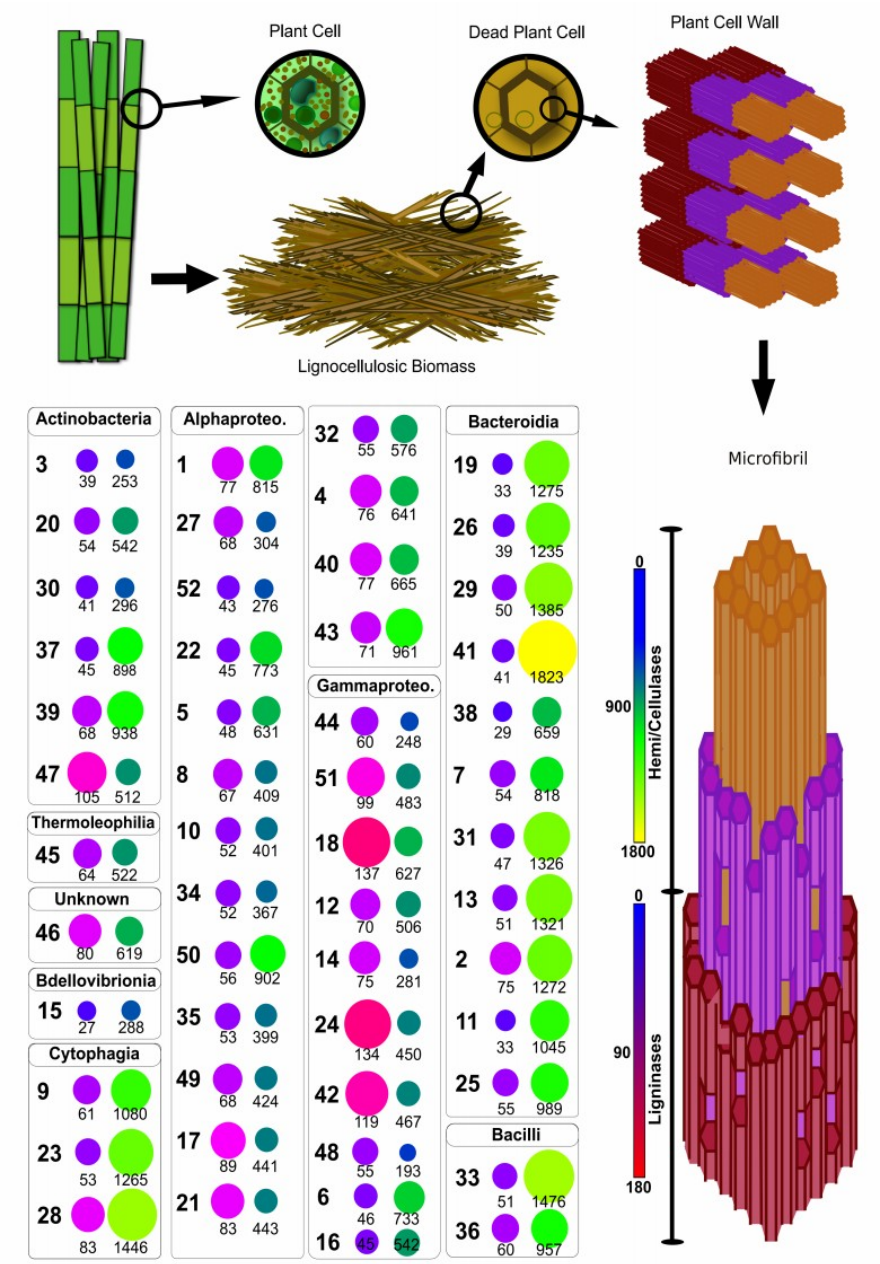

Figure 19. Model of the potential participation of each species, found in the consortium's metagenome, in the process of deconstruction of lignocellulosic biomass. Species with higher amounts of sequences related to this process are considered more relevant to the process. The MAGs were sorted based on their Class. Circles on the left show ligninases, circles on the right show hemicellulases and cellulases together.

It was not evident through our genomic-based metabolic qualitative modeling, the reasons lignin was not being subject to deconstruction as cellulose and hemicellulose were. Also, some community dynamics occurred during the 18 weeks of the sampling interval, such as the severe decrease in Nocardioidaceae sp. Bin47 abundance, a species phylogenetically related to species known to be effective in lignocellulose deconstruction. Thus, the secondary metabolite metabolism was inspected, improving the metabolic characterization of the consortium. In this respect, the consortium showed more than 200 genetic clusters related to secondary metabolite synthesis, indicating the relevant importance of ecophysiological phenomena controlled by this mean. Seventeen different types of gene clusters are found in this community, revealing a rich diversity of processes that may contribute to its dynamics.

Most genetic clusters related to secondary metabolites metabolism are of uncertain activity in the ecological processes governing bacterial communities in general (Tyc *et al.*, 2017). NRPs are a peptide-based secondary metabolite synthesized by mega-enzymatic Synthetase complexes (NRPS), thus not requiring the cell's ribosomal machinery as is usual for other types of peptides (Evans *et al*., 2011). NRPs show an very broad range of bioactivities, requiring caution to generalize the ecophysiological meaning of the abundance of this type of sequence for the suggested interactions in this consortium. Often, metabolites produced by NRPSs presents activity related to competitive interference, interacting to reduce populations of competing species or

groups (as is the case of antibiotics and cytotoxins). Also, some metabolites produced by NRPSs show antioxidant properties (as pigments).

NRPS-like is a category of sequence that shows sequence similarity to NRPS. Still, it does not show similarity (and/or identity) enough to indicate to which type of NRPS that sequence may be designated. Frequently, this may result from a broken NRPS cluster sequence (i.e., fragments of the same gene cluster in different contigs), so representing a purely technical insufficiency of the genome sequencing methodology. Thus, although this was the third most abundant secondary metabolite sequence category found by antiSMASH in this consortium, it was not considered in depth. Arylpolyene metabolites are pigments shown to be taxonomically widespread in Bacteria and protect the bacterial cell from reactive oxygen species, acting as antioxidants in similar ways as carotenoids (Shöner *et al*., 2015). These three categories (NRPS, NRPS-like, and Arylpolyene) summed represent little over 50% of all gene clusters related to the secondary metabolism found in the MAGs.

As for NRPs and Arylpolyene gene clusters, Bacteroidia class (11 MAGs) were the group of the highest counting of such sequences - 17 and 18 gene clusters for NRPS and Arylpolyene. The Gammaproteobacteria class (10 MAGs) showed 17 gene clusters of NRPS, and Cytophagia class (from the Bacteroidetes phylum, as Bacteroidia) showed 7 Arylpolyene sequences in 3 MAGs. These distributions may indicate relevant ecological processes of competitive interference between species or groups in the community through NRP metabolites. The widespread presence of the Arylpolyenes and

NRPs antioxidant protective pigments in some groups suggests mild to an intense oxidative environment in which the consortium was cultivated.

Bacterial terpene metabolites were recently shown in various prokaryotes, mainly (but not restricted to) Burkholderiales, Sphingobacteriales, and Pseudomonadales Orders of the Firmicutes Phylum (Yamada *et al*., 2015), all found in this community. Bacterial terpenes are volatile molecules, thought to be notably useful in soil communities as this property allows a rapid diffusion rate in soil particulate and pores. However, such molecules' ecophysiological significance is still not well defined (Tyc *et al.*, 2016). In the gene clusters found in this consortium's MAGs, 24 (10.7%) terpene gene clusters were identified. Polyketide is a class of bacterial secondary metabolites widespread in the tree of life that is synthesized by polyketide synthases (PKSs), a multidomain or multimeric enzymatic complex. It also does not require the cell's ribosomal machinery, as NRPSs presented before. However, the biochemical process related to this complex is more akin to a fatty acid synthesis than ribosomal peptide synthesis (Tyc *et al*., 2016). Polyketides often show a soluble cytotoxic or antibiotic activity (Tyc *et al*., 2016). In this metagenome, the MAGs 19 (8.5%) and 18 (8%) gene clusters of Type I PKS and Type 3 PKS, respectively.

Siderophores are low-weight molecules that present a high affinity to iron, helping to solubilize and carry scarce ferric ions to the cell's interior. They can be synthesized as an NRP, but some other biosynthesizing processes are known. Siderophores often act as antagonistic mechanisms, limiting the availability of relevant iron resources (Eberl *et al*., 2009). In the MAGs found in this metagenome, 13 (5.8%) gene clusters were

identified as siderophore sequences. Bacteriocins are antibiotic molecules synthesized by bacteria as a means to control competing groups. Bacteriocins can be effective against particular species (narrow spectrum) or genera (broad spectrum) and are estimated to be found in the majority of bacterial species (Tyc *et al*., 2016). In the MAGs found in this metagenome, 9 (4%) gene clusters of secondary metabolite sequences were identified as bacteriocin sequences. Resorcinol is a category of molecules that presents antagonistic activities against bacterial and eukaryotic cells and may be related to lignocellulose deconstruction reactions (Calderon *et al*., 2014; Brink *et al*., 2019). In the gene clusters related to secondary metabolites found in the consortium's MAGs, 9 (4%) gene clusters were identified as resorcinol sequences. Taken together, terpene, polyketide, siderophores, bacteriocins, and resorcinol categories represent little over 41% of all gene clusters sequences found in the MAGs.

Terpene gene cluster sequences were found substantially more frequent in the Alphaproteobacteria class (15 clusters, 17 MAGs) and Bacteroides (6 clusters, 11 MAGs). This result suggests that Alphaproteobacteria and Bacteroidia classes may adopt ecophysiological strategies of fast diffusion in this community. Polyketide gene clusters of type I were found more often in Alphaproteobacteria (6 clusters, 17 MAGs), Gammaproteobacteria (5 clusters, 10 MAGs), and Bacteroidia (5 clusters, 11 MAGs). In contrast, Polyketide Type III was not found only in Gammaproteobacteria, and Bacteroidia class showed more gene clusters of this category (7 clusters). All groups presented at least one gene cluster of this category, indicating that all groups can act antagonistically to some degree through PKS-driven processes. This indicates the

relevance of such secondary metabolites to the ecophysiological interactions of this consortium.

Siderophores were not found in Alphaproteobacteria (17 MAGs) and Firmicutes (2 MAGs), suggesting that they do not adopt this strategy to collect iron resources from the medium. These sequences were more often found in Bacteroidia class (5 sequences, 11 MAGs), indicating that this class most possibly adopt a siderophore strategy to harvest iron resources from the environment compared to the other groups. More bacteriocin sequences were found in the Bacteroidetes phylum (5 sequences, 14 MAGs), indicating that this group may present this antagonistic close-quarters strategy more often than the other groups. Also, only in Firmicutes (2 sequences, 2 MAGs) and in Proteobacteria phylum (2 sequences, 28 MAGs), these sequences were found, indicating that this strategy is not a frequent option for the groups found in this community except Bacteroidetes phylum. Resorcinol sequences were found in Bacteroidetes (7 sequences, 14 MAGs), Firmicutes (1 sequence, 2 MAGs) phyla, and Gammaproteobacteria class (1 sequence, 10 MAGs). This result shows more lignin deconstruction potential through resorcinol molecules in Bacteroidetes than the other groups found in this metagenome.

Overall, the secondary metabolites analyses supported a potentially diverse multitude of ecophysiological strategies of interaction in this consortium through a plethora of secondary metabolites. Such an approach may regulate the species/groups' abundance change found in previous results through antagonisms as antibiotic and cytotoxicity, both in physical proximity and long-distance interaction. Thus, such

speculated communications indicate a potential interference in the community's shared metabolism, modifying the processes of deconstruction of the lignocellulosic biomass by adjusting abundances and competition for resources between the taxonomic groups (Jimenez *et al*., 2016).

It is noteworthy that the distribution of such genetic clusters among the taxonomic groups is uneven, indicating that each taxonomic group has different relevance to this community's ecophysiological processes. We also assume that more sequences relevant to these metabolic aspects indicates more influence in the consortium's dynamics through this mean. In general, Phylum Bacteroidetes is particularly relevant in this aspect. Beyond the direct participation of each species in the consortium over the deconstruction of lignocellulose, knowledge about the potential involvement in other ecophysiological processes can contribute to the engineering and Synthetic Biology efforts towards a biotechnologically efficient consortium, or controlled steps involved in this process (Perez-Garcia *et al*., 2016; Fang *et al*., 2020).

6. CONCLUSIONS

We showed the cultivated consortium was able to deconstruct lignocellulosic biomass. We could not find a definitive answer to the chemical conservation of lignin under the exposure of the lignocellulose fibers to the consortium, particularly considering that sequences related to this process were found in almost all genomes, and most of the highly abundant species found are phylogenetically related to already known

organisms able to execute this task. Separating the 52 genomes with high completeness and low contamination allowed us to build a metabolic model individually for each, then contrasting each species' capacity concerning the deconstruction of saccharidic and lignin polymers. We also observed that the consortium shows mostly stable dynamics on species richness and abundances, suggesting its potential use in biotechnological and industrial applications. The overall redundancy of the metabolism of the groups supports this proposition. Nevertheless, the division of biochemical labor indicated by the sequences related to the deconstruction of lignocellulose, and the ecophysiological potentialities characterized by the uneven distribution of gene clusters related to secondary metabolite metabolism, suggest that each species has its particular relevance in the structure of the consortium. Through the genomic information obtained for this community, we can suggest that 3 main processes control its structure and change in time, namely 1) stochastic processes, in which the most abundant species tended to keep its dominance; 2) division of biochemical labor, in which a concatenation of syntrophic relations pertaining the carbon source (lignocellulose) contributed to the maintenance of the consortium’s structure; and 3) taxonomically uneven but diverse ecophysiological regulation, such as antibiosis and competition, may interfere in the consortium’s structure though acting over species richness and abundance. Thus, this consortium is appropriate for efforts in engineering and synthetic biology and may contribute to the broadening of the knowledge about the myriad of biochemical processes involved in the deconstruction of lignocellulose and its stability under potential manipulation for applications of biotechnological efforts.

8.REFERENCES

Aittokallio, Tero, and Benno Schwikowski. "Graph-based methods for analysing networks in cell biology." **Briefings in bioinformatics** 7.3 (2006): 243-255.

Alvarenga, Danillo O., Marli F. Fiore, and Alessandro M. Varani. "A metagenomic approach to cyanobacterial genomics." **Frontiers in microbiology** 8 (2017): 809.

Amann, Rudolf I., Wolfgang Ludwig, and Karl-Heinz Schleifer. "Phylogenetic identification and in situ detection of individual microbial cells without cultivation." **Microbiological reviews** 59.1 (1995): 143-169.

Anantharaman, Karthik, et al. "Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system." **Nature communications** 7 (2016): 13219.

Aristidou, Aristos, and Merja Penttilä. "Metabolic engineering applications to renewable resource utilization." **Current opinion in biotechnology** 11.2 (2000): 187-198.

Baraston, A. B., et al. "A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrates-binding modules reveals a conserved mode of starch recognition." **J. Biol. Chem** 281 (2005): 587-598.

Basak, Bikram, et al. "Dark fermentative hydrogen production from pretreated lignocellulosic biomass: Effects of inhibitory byproducts and recent trends in mitigation strategies." **Renewable and Sustainable Energy Reviews** 133 (2020): 110338.

Bayer, E. A., Y. Shoham, and R. J. T. P. Lamed. "Lignocellulose-decomposing bacteria and their enzyme systems." **The prokaryotes** 4 (2013): 215-266.

Beckham, Gregg T., et al. "Opportunities and challenges in biological lignin valorization." **Current opinion in biotechnology** 42 (2016): 40-53.

Biggs, Matthew B., et al. "Metabolic network modeling of microbial communities." **Wiley Interdisciplinary Reviews: Systems Biology and Medicine** 7.5 (2015): 317-334.

Black, Gary W., et al. "Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates." **Biochemical Journal** 319.2 (1996): 515-520.

Blackburn, Neil T., and Anthony J. Clarke. "Identification of four families of peptidoglycan lytic transglycosylases." **Journal of Molecular Evolution** 52.1 (2001): 78-84.

Blin, Kai, et al. "antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline." **Nucleic acids research** 47.W1 (2019): W81-W87.

Boer, Wietse de, et al. "Living in a fungal world: impact of fungi on soil bacterial niche development." **FEMS microbiology reviews** 29.4 (2005): 795-811.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." **Bioinformatics** 30.15 (2014): 2114-2120.

Bond, Jesse Q., David Martin Alonso, and James A. Dumesic. "Catalytic strategies for converting lignocellulosic carbohydrates to fuels and chemicals." **Aqueous Pretreatment of Plant Biomass for Biological and Chemical Conversion to Fuels and Chemicals** (2013): 61-102.

Brenner, Katie, Lingchong You, and Frances H. Arnold. "Engineering microbial consortia: a new frontier in synthetic biology." **Trends in biotechnology** 26.9 (2008): 483-489.

Brethauer, Simone, and Michael H. Studer. "Biochemical conversion processes of lignocellulosic biomass to fuels and chemicals–a review." **CHIMIA International Journal for Chemistry** 69.10 (2015): 572-581.

Brink, Daniel P., et al. "Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database." **Applied microbiology and biotechnology** 103.10 (2019): 3979-4002.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. "Fast and sensitive protein alignment using DIAMOND." **Nature methods** 12.1 (2015): 59-60.

Bushnell, L. D., and H. F. Haas. "The utilization of certain hydrocarbons by microorganisms." **Journal of Bacteriology** 41.5 (1941): 653.

Busk, Peter K., et al. "Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function." **BMC bioinformatics** 18.1 (2017): 214.

Çakır, Tunahan, and Mohammad Jafar Khatibipour. "Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation." **Frontiers in Bioengineering and Biotechnology** 2 (2014): 62.

Calderón, Claudia E., Antonio de Vicente, and Francisco M. Cazorla. "Role of 2-hexyl, 5-propyl resorcinol production by Pseudomonas chlororaphis PCL1606 in the multitrophic interactions in the avocado rhizosphere during the biocontrol process." **FEMS microbiology ecology** 89.1 (2014): 20-31.

Canilha, Larissa, et al. "Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation." **Journal of Biomedicine and Biotechnology** 2012 (2012).

Cantarel, Brandi L., et al. "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." **Nucleic acids research** 37.suppl_1 (2009): D233-D238.

Cardona, Cesar, et al. "Network-based metabolic analysis and microbial community modeling." **Current Opinion in Microbiology** 31 (2016): 124-131.

Carlos, Camila, Huan Fan, and Cameron R. Currie. "Substrate shift reveals roles for members of bacterial consortia in degradation of plant cell wall polymers." **Frontiers in microbiology** 9 (2018): 364.

Carvalho, Ana Flávia Azevedo, et al. "Purification and Characterization of the α-glucosidase Produced by Thermophilic Fungus Thermoascus aurantiacus CBMAI 756." **The Journal of Microbiology** 48.4 (2010): 452-459.

Chakraborty, S., et al. "Polysaccharide lyases." **Current Developments in Biotechnology and Bioengineering**. Elsevier, 2017. 527-539.

Chaumeil, Pierre-Alain, et al. "GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database." (2020): 1925-1927.

Chistoserdova, Ludmila. "Recent progress and new challenges in metagenomics for biotechnology." **Biotechnology letters** 32.10 (2010): 1351-1359.

Christov, Lyudmil Pavlov, and Bernard Alexander Prior. "Esterases of xylan-degrading microorganisms: production, properties, and significance." **Enzyme and Microbial Technology** 15.6 (1993): 460-475.

Chylenski, Piotr, et al. "Lytic polysaccharide monooxygenases in enzymatic processing of lignocellulosic biomass." ACS **Catalysis** 9.6 (2019): 4970-4991.

Constancio, Milena Tavares Lima, et al. "Exploring the Potential of Two Bacterial Consortia to Degrade Cellulosic Biomass for Biotechnological Applications." **Current Microbiology** 77.10 (2020): 3114-3124.

Cook, Daniel J., and Jens Nielsen. "Genome-scale metabolic models applied to human health and disease." **Wiley Interdisciplinary Reviews: Systems Biology and Medicine** 9.6 (2017): e1393.

Cooper, Geoffrey M, and Robert E. Hausman. **The Cell: A Molecular Approach.** Washington, D.C: ASM Press, 2009. Print.

Dash, Satyakam, et al. "Thermodynamic analysis of the pathway for ethanol production from cellobiose in Clostridium thermocellum." **Metabolic engineering** 55 (2019): 161-169.

De Souza, Amanda P., et al. "Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production." **BioEnergy Research** 6.2 (2013): 564-579.

de Oliveira Bordonal, Ricardo, et al. "Sustainability of sugarcane production in Brazil. A review." **Agronomy for Sustainable Development** 38.2 (2018): 13.

dos Santos, Antonio Carlos, et al. "Lignin–enzyme interactions in the hydrolysis of lignocellulosic biomass." **Trends in biotechnology** 37.5 (2019): 518-531.

Eberl, Hermann J., and Shannon Collinson. "A modeling and simulation study of siderophore mediated antagonism in dual-species biofilms." **Theoretical Biology and Medical Modelling** 6.1 (2009): 30.

Evans, Bradley S., Sarah J. Robinson, and Neil L. Kelleher. "Surveys of non-ribosomal peptide and polyketide assembly lines in fungi and prospects for their analysis in vitro and in vivo." **Fungal genetics and biology** 48.1 (2011): 49-61.

Fang, Xin, Colton J. Lloyd, and Bernhard O. Palsson. "Reconstructing organisms in silico: genome-scale models and their emerging applications." **Nature Reviews Microbiology** 18.12 (2020): 731-743.

Fang, Zhen, Richard L. Smith, and Xiao-Fei Tian, eds. **Production of Materials from Sustainable Biomass Resources**. Vol. 9. Springer, 2019.

Festucci-Buselli, Reginaldo A., Otoni, Wagner C., Joshi, Chandrashekhar P. “Structure, organization, and functions of cellulose synthase complexes in higher plants.” **J. Plant Physiology** 19.1 (2007): 13.

Fierer, Noah, Albert Barberán, and Daniel C. Laughlin. "Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities." **Frontiers in microbiology** 5 (2014): 614.

Flipphi, Michel JA, et al. "Cloning of the Aspergillus niger gene encoding α-L-arabinofuranosidase A." **Applied microbiology and biotechnology** 39.3 (1993): 335-340.

French, Alfred D. "Glucose, not cellobiose, is the repeating unit of cellulose and why that is important." **Cellulose** 24.11 (2017): 4605-4609.

Francke, Christof, Roland J. Siezen, and Bas Teusink. "Reconstructing the metabolic network of a bacterium from its genome." **Trends in microbiology** 13.11 (2005): 550-558.

Gilkes, N. R., et al. "Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis." **Journal of Biological Chemistry** 263.21 (1988): 10401-10407.

Goering, Harold Keith, and Peter J. Van Soest. Forage fiber analyses: apparatus, reagents, procedures, and some applications. No. 379. **Agricultural Research Service**, US Department of Agriculture, 1970.

Greenblum, Sharon, Peter J. Turnbaugh, and Elhanan Borenstein. "Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease." **Proceedings of the National Academy of Sciences** 109.2 (2012): 594-599.

Gupta, Vijai Kumar, and Maria G. Tuohy. **"Biofuel technologies." Recent Developments**. Editorial Springer (2013).

Gutiérrez-Rivera, Beatriz, et al. "Conversion efficiency of glucose/xylose mixtures for ethanol production using Saccharomyces cerevisiae ITV01 and Pichia stipitis NRRL Y-7124." **Journal of Chemical Technology & Biotechnology** 87.2 (2012): 263-270.

Handelsman, Jo. "Metagenomics: application of genomics to uncultured microorganisms." **Microbiology and molecular biology reviews** 68.4 (2004): 669-685.

Hedlund, Brian P., et al. "Impact of single-cell genomics and metagenomics on the emerging view of extremophile “microbial dark matter”." **Extremophiles** 18.5 (2014): 865-875.

Henrissat, Bernard. "A classification of glycosyl hydrolases based on amino acid sequence similarities." **Biochemical journal** 280.2 (1991): 309-316.

Henry, Christopher S., et al. "High-throughput generation, optimization and analysis of genome-scale metabolic models." **Nature biotechnology** 28.9 (2010): 977-982.

Henry, Christopher S., et al. "Microbial community metabolic modeling: a community data-driven network reconstruction." **Journal of cellular physiology** 231.11 (2016): 2339-2345.

Hill, Jason, et al. "Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels." **Proceedings of the National Academy of sciences** 103.30 (2006): 11206-11210.

Hoang, Diep Thi, et al. "MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation." **BMC evolutionary biology** 18.1 (2018): 1-11.

Houfani, Aicha Asma, et al. "Insights from enzymatic degradation of cellulose and hemicellulose to fermentable sugars–a review." **Biomass and Bioenergy** 134 (2020): 105481.

How about lignin?. “Lignin structure”, 05 Jul. 2013, www.howaboutlignin.blogspot.com/2013/07/lignin-structure.html. Accessed 19 Jan. 2021.

Huerta-Cepas, Jaime, et al. "eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses." **Nucleic acids research** 47.D1 (2019): D309-D314.

Isikgor, Furkan H., and C. Remzi Becer. "Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers." **Polymer Chemistry** 6.25 (2015): 4497-4559.

Kamimura, Naofumi, et al. "Advances in microbial lignin degradation and its applications." **Current opinion in biotechnology** 56 (2019): 179-186.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. "BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences." **Journal of molecular biology** 428.4 (2016): 726-731.

Kang, Qian, et al. "Bioethanol from lignocellulosic biomass: current findings determine research priorities." **The Scientific World Journal 2014** (2014).

Katoh, Kazutaka, John Rozewicki, and Kazunori D. Yamada. "MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization." **Briefings in bioinformatics** 20.4 (2019): 1160-1166.

Kettler, Gregory C., et al. "Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*." **PLoS Genet** 3.12 (2007): e231.

Kulkarni Vishakha, S., D. Butte Kishor, and S. Rathod Sudha. "Natural polymers–A comprehensive review." **International journal of research in pharmaceutical and biomedical sciences** 3.4 (2012): 1597-1613.

Levasseur, Anthony, et al. "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes." **Biotechnology for biofuels** 6.1 (2013): 41.

Li, Dinghua, et al. "MEGAHIT v1. 0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices." **Methods** 102 (2016): 3-11.

Liu, Qiyong P., et al. "Bacterial glycosidases for the production of universal red blood cells." **Nature biotechnology** 25.4 (2007): 454-464.

Limayem, Alya, and Steven C. Ricke. "Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects." **Progress in energy and combustion science** 38.4 (2012): 449-467.

Lombard, Vincent, et al. "The carbohydrate-active enzymes database (CAZy) in 2013." **Nucleic acids research** 42.D1 (2014): D490-D495.

Lynd, Lee R., et al. "Microbial cellulose utilization: fundamentals and biotechnology." **Microbiology and molecular biology reviews** 66.3 (2002): 506-577.

Lynd, Lee R., et al. "How biotech can transform biofuels." **Nature biotechnology** 26.2 (2008): 169-172.

Marcy, Yann, et al. "Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." **Proceedings of the National Academy of Sciences** 104.29 (2007): 11889-11894.

McDaniel, Elizabeth A., Karthik Anantharaman, and K. D. McMahon. "metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways." **BioRxiv** (2019).

Merino, Sandra T., and Joel Cherry. "Progress and challenges in enzyme development for biomass utilization." **Biofuels**. Springer, Berlin, Heidelberg, 2007. 95-120.

Milanez, Artur Yabe, et al. "De promessa a realidade: como o etanol celulósico pode revolucionar a indústria da cana-de-açúcar: uma avaliação do potencial competitivo e sugestões de política pública." (2015).

Mood, Sohrab Haghighi, et al. "Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment." **Renewable and Sustainable Energy Reviews** 27 (2013): 77-93.

Nobu, Masaru K., et al. "Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor." **The ISME journal** 9.8 (2015): 1710-1722.

Oberhardt, Matthew A., Bernhard Ø. Palsson, and Jason A. Papin. "Applications of genome-scale metabolic reconstructions." **Molecular systems biology** 5.1 (2009): 320.

O'Brien, Edward J., Jonathan M. Monk, and Bernhard O. Palsson. "Using genome-scale models to predict biological capabilities." **Cell** 161.5 (2015): 971-987.

Ogeda, Thais Lucy, and Denise FS Petri. "Hidrólise enzimática de biomassa." **Química nova** 33.7 (2010): 1549-1558.

Oh, You-Kwan, et al. "Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data." **Journal of Biological Chemistry** 282.39 (2007): 28791-28799.

O'Leary, Nuala A., et al. "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation." **Nucleic acids research** 44.D1 (2016): D733-D745.

Page, Andrew J., et al. "Roary: rapid large-scale prokaryote pan genome analysis." **Bioinformatics** 31.22 (2015): 3691-3693.

Palin, Robert, and Anja Geitmann. "The role of pectin in plant morphogenesis." **Biosystems** 109.3 (2012): 397-402.

Parks, Donovan H., et al. "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." **Genome research** 25.7 (2015): 1043-1055.

Parks, Donovan H., et al. "Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life." **Nature microbiology** 2.11 (2017): 1533-1542.

Pérez, J., et al. "Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview." **International microbiology** 5.2 (2002): 53-63.

Perez-Garcia, Octavio, Gavin Lear, and Naresh Singhal. "Metabolic network modeling of microbial interactions in natural and engineered environmental systems." **Frontiers in microbiology** 7 (2016): 673.

Pippo, W. Alonso, et al. "Energy recovery from sugarcane-trash in the light of 2nd generation biofuels. Part 1: current situation and environmental aspects." **Waste and Biomass Valorization** 2.1 (2011): 1-16.

Puentes-Téllez, Pilar Eliana, and Joana Falcao Salles. "Dynamics of Abundant and Rare Bacteria During Degradation of Lignocellulose from Sugarcane Biomass." **Microbial ecology** 79.2 (2020): 312-325.

Ridley, Brent L., Malcolm A. O'Neill, and Debra Mohnen. "Pectins: structure, biosynthesis, and oligogalacturonide-related signaling." **Phytochemistry** 57.6 (2001): 929-967.

Rodrigues, Rita CLB, et al. "Response surface methodology for xylitol production from sugarcane bagasse hemicellulosic hydrolyzate using controlled vacuum evaporation process variables." **Process Biochemistry** 38.8 (2003): 1231-1237.

Rubin, Edward M. "Genomics of cellulosic biofuels." **Nature** 454.7206 (2008): 841-845.

Sander, Elizabeth L., J. Timothy Wootton, and Stefano Allesina. "What can interaction webs tell us about species roles?." **PLoS computational biology** 11.7 (2015): e1004330.

Santos, Filipe, Joost Boele, and Bas Teusink. "A practical guide to genome-scale metabolic models and their analysis." **Methods in enzymology**. Vol. 500. Academic Press, 2011. 509-532.

Schöner, Tim A., Darko Kresovic, and Helge B. Bode. "Biosynthesis and function of bacterial dialkylresorcinol compounds." **Applied microbiology and biotechnology** 99.20 (2015): 8323-8328.

Shi, Jian, Ratna R. Sharma-Shivappa, and Mari S. Chinn. "Microbial pretreatment of cotton stalks by submerged cultivation of Phanerochaete chrysosporium." **Bioresource technology** 100.19 (2009): 4388-4395.

Simon, Carola, and Rolf Daniel. "Construction of small-insert and large-insert metagenomic libraries." **Metagenomics**. Humana Press, Totowa, NJ, 2010. 39-50.

Sinnott, Michael L. "Catalytic mechanism of enzymic glycosyl transfer." **Chemical Reviews** 90.7 (1990): 1171-1202.

Steele, Helen L., et al. "Advances in recovery of novel biocatalysts from metagenomes." **Journal of molecular microbiology and biotechnology** 16.1-2 (2009): 25-37.

Streit, Wolfgang R., and Ruth A. Schmitz. "Metagenomics–the key to the uncultured microbes." **Current opinion in microbiology** 7.5 (2004): 492-498.

Sarmiento, Felipe, et al. "Bioprospection of Extremozymes for Conversion of Lignocellulosic Feedstocks to Bioethanol and Other Biochemicals." **Extremophilic Enzymatic Processing of Lignocellulosic Feedstocks to Bioenergy**. Springer, Cham, 2017. 271-297.

Svensson, David, Stefan Ulvenlund, and Patrick Adlercreutz. "Efficient synthesis of a long carbohydrate chain alkyl glycoside catalyzed by cyclodextrin glycosyltransferase (CGTase)." **Biotechnology and bioengineering** 104.5 (2009): 854-861.

Sweeney, Matt D., and Feng Xu. "Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments." **Catalysts 2.2** (2012): 244-263.

Tabita, F. Robert, et al. "Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships." **Journal of experimental botany** 59.7 (2008): 1515-1524.

Taherzadeh, Mohammad J., and Keikhosro Karimi. "Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review." **International journal of molecular sciences** 9.9 (2008): 1621-1651.

Tarbouriech, Nicolas, Simon J. Charnock, and Gideon J. Davies. "Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases." **Journal of molecular biology** 314.4 (2001): 655-661.

Terrapon, Nicolas, et al. "The CAZy database/the carbohydrate-active enzyme (CAZy) database: principles and usage guidelines." **A practical guide to using glycomics databases.** Springer, Tokyo, 2017. 117-131.

Thor, ShengShee, Joseph R. Peterson, and Zaida Luthey-Schulten. "Genome-scale metabolic modeling of archaea lends insight into diversity of metabolic function." **Archaea** 2017 (2017).

Trifinopoulos, Jana, et al. "W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis." **Nucleic acids research** 44.W1 (2016): W232-W235.

Tyc, Olaf, et al. "The ecological role of volatile and soluble secondary metabolites produced by soil bacteria." **Trends in microbiology** 25.4 (2017): 280-292.

Tzamali, Eleftheria, et al. "A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities." **BMC systems biology** 5.1 (2011): 167.

Ventorino, Valeria, et al. "Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application." **Scientific Reports** 5.8161: 1.

Von Mering, C., et al. "Quantitative phylogenetic assessment of microbial communities in diverse environments." **Science** 315.5815 (2007): 1126-1130.

Wang, Lu, et al. "Diverse bacteria with lignin degrading potentials isolated from two ranks of coal." **Frontiers in microbiology** 7 (2016): 1428.

Wheeler, Travis J., and Sean R. Eddy. "nhmmer: DNA homology search with profile HMMs." **Bioinformatics** 29.19 (2013): 2487-2489.

Wilhelm, Roland C., et al. "Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing." **The ISME Journal** 13.2 (2019): 413-429.

Willis, Anusuya, et al. "Genome variation in nine co-occurring toxic *Cylindrospermopsis raciborskii* strains." **Harmful algae** 73 (2018): 157-166.

Wongwilaiwalin, Sarunyou, et al. "Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system." **Enzyme and Microbial Technology** 47.6 (2010): 283-290.

Wood, Derrick E., and Steven L. Salzberg. "Kraken: ultrafast metagenomic sequence classification using exact alignments." **Genome Biology** 15.3 (2014): 1-12.

Xu, Rong, et al. "Lignin depolymerization and utilization by bacteria." **Bioresource Technology** 269 (2018): 557-566.

Yamada, Yuuki, et al. "Terpene synthases are widely distributed in bacteria." **Proceedings of the National Academy of Sciences** 112.3 (2015): 857-862.

Yin, Yanbin, et al. "dbCAN: a web resource for automated carbohydrate-active enzyme annotation." **Nucleic acids research** 40.W1 (2012): W445-W451.

Zabed, Hossain M., et al. "Recent advances in biological pretreatment of microalgae and lignocellulosic biomass for biofuel production." **Renewable and Sustainable Energy Reviews** 105 (2019): 105-128.

Zhang, Han, et al. "dbCAN2: a meta server for automated carbohydrate-active enzyme annotation." **Nucleic Acids Research** 46.W1 (2018): W95-W101.

Zhao, Xuebing, Lihua Zhang, and Dehua Liu. "Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose." **Biofuels, Bioproducts and Biorefining** 6.4 (2012): 465-482.

Zhou, Xiaowei, et al. "A critical review on hemicellulose pyrolysis." **Energy Technology** 5.1 (2017): 52-79.

Zhu, Yongming, et al. "Calculating sugar yields in high solids hydrolysis of biomass." **Bioresource technology** 102.3 (2011): 2897-2903.