Lignin degradation by microbial consortia from termite : Enrichment and characterization of bacterial communities by (bio-)chemical and metagenomics approaches

Louison Dumond

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse

Présentée et soutenue par

Louison DUMOND

Le 13 décembre 2021

Dégradation de la lignine par des consortia microbiens issus de termites : Enrichissement et caractérisation des communautés par des approches (bio-)chimiques et métagénomiques

Ecole doctorale : **SEVAB – Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche : **TBI – Toulouse Biotechnology Institute, Bio & Chemical Engineering**

Thèse dirigée par **Guillermina HERNANDEZ-RAQUET**

Jury

Mme Marie Françoise Gorwa-Grauslund, Professeure, Université de Lund, Suède, Rapporteure
M. Stéphane Chaillou, Directeur de recherche, INRAE MICALIS, France, Rapporteur
M. Gabriel Paës, Directeur de recherche, INRAE FARE, France, Examinateur
Mme Nathalie Vialaneix, Directrice de recherche, INRAE MIA-T, France, Présidente
Mme Claire Hoede, Ingénieure de recherche, INRAE Genotoul Bioinfo, France, Co-encadrante
Mme Guillermina Henandez-Raquet, Directrice de recherche, INRAE TBI, France, Directrice de thèse

Résumé

Construite sur le modèle de la raffinerie pétrochimique, la bioraffinerie doit représenter un changement de paradigme dans l'utilisation des ressources, un retour à l'utilisation de la biomasse lignocellulosique, renouvelable, plutôt que des sources de carbone fossiles. Cependant, la transformation de la lignocellulose représente un enjeu de taille. En particulier, la valorisation de la lignine, polymère aromatique, présente un fort potentiel mais reste un point d'achoppement important en raison de la difficulté de sa déconstruction.

La nature regorge pourtant d'organismes pour lesquels la biomasse lignocellulosique est à la base du régime alimentaire. C'est notamment le cas des insectes xylophages comme les termites dont les communautés microbiennes endosymbiotiques sont capables de convertir la lignocellulose en acétate utilisé par l'hôte comme source d'énergie. En revanche, notre connaissance actuelle de la dégradation de la lignine par des communautés microbiennes issues de termites reste insuffisante. Nous nous proposons donc, dans cette thèse, d'étudier des consortia microbiens issus de termite et leur implication potentielle dans la délignification.

Dans le cadre du projet européen ZELCOR ("Zero Waste Ligno-Cellulosic Biorefineries by Integrated Lignin Valorisation") fondé par le partenariat public-privé BBI (Bio-Based Industries), nous avons procédé à l'enrichissement en batchs successifs de consortia microbiens sur différents substrats et dans différentes conditions afin d'obtenir un consortium capable de dégrader la lignine puis l'avons cultivé en bioréacteurs. Afin de montrer le potentiel de consortia microbiens issus de termite pour la valorisation de la lignine, nous avons couplé la caractérisation structurelle des substrats utilisés au cours de la digestion par hydrolyse acide, 2D HSQC NMR et ^{13}C-IS py-GC-MS à la caractérisation des communautés microbiennes par métagénomique « shotgun ».

Cette approche couplée nous a permis de démontrer le potentiel lignolytique, aussi bien en milieu anaéobie qu'aérobie de communautés microbiennes issues de termites. Nous avons également pu enrichir et étudier un consortium lignocellulolytique. Grâce à la reconstruction de MAGs, nous avons pu identifier des bactéries et enzymes d'intérêt biotechnologique. Par ailleurs, cela nous a également permis d'élucider le fonctionnement de ces consortia en montrant l'importance des interactions synergiques entre MAGs pour la transformation de la lignocellulose en énergie.

***Mots clefs** : lignocellulose, lignine, métagénomique shotgun, metabarcoding, MAGs, microbiote, communautés microbiennes*

Abstract

Based on the petrochemical refinery model, biorefinery are to represent a paradigm shift in resource usage with a call back to lignocellulosic biomass instead of fossil carbon sources. However, lignocellulose transformation is challenging. The case of lignin, an aromatic polymer, is especially tricky. On the one hand its valorization presents a high potential but on the other hand the difficulty to achieve its deconstruction limits the applications.

However, there is in Nature a multitude of organisms for which lignocellulosic biomass constitute the main diet. It is most notably the case for phytophagous insects such as termites whose endosymbiotic microbial communities are able to transform lignocellulose into acetate which is used by the host for energy. However, our knowledge of lignin degradation by termites is still sparse if not contradictory. In this PhD project, we thus aims to study microbial consortia from termite and their potential implication in delignification of biomass.

Within the BBI (Bio-Based Industries) joint undertaking under the ZELCOR ("Zero Waste Ligno-Cellulosic Biorefineries by Integrated Lignin Valorisation") framework, we have tried various substrates and conditions to enrich a lignolytic consortium in batch which we then implemented in bioreactors. So that we could investigate the potential of termite-derived microbial consortia for lignin valorization, we coupled structural characterization of the substrates by acid hydrolysis, 2D HSQC NMR and ^{13}C-IS py-GC-MS with the characterization of the microbial communities by « shotgun » metagenomics.

Thanks to this coupled approach, we could demonstrate the lignolytic potential of termite derived microbial consortia in both aerobic and anaerobic conditions. This also allowed us to enrich and characterize a lignocellulolytic consortium. Metagenomics approaches resulted in the reconstruction of numerous MAGs which enabled the identification of both bacteria and enzymes of biotechnological interest. Furthermore, we were able to decipher some aspects of the communities functioning by highlighting the importance of synergistic interactions between MAGs to convert lignocellulose into energy.

Keywords: *lignocellulose, lignin, shotgun metagenomics, metabarcoding, MAGs, microbiota, micribial communities*

Publications

PUBLISHED

Louison Dumond, Pui Ying Lam, Gijs van Erven, Mirjam Kabel, Fabien Mounet, Jacqueline Grima-Pettenati, Yuki Tobimatsu, and Guillermina Hernandez-Raquet. Termite Gut Microbiota Contribution to Wheat Straw Delignification in Anaerobic Bioreactors. *ACS Sustainable Chemistry & Engineering* **2021** *9* (5), 2191-2202

IN PREPARATION

Louison Dumond, Claire Hoede, Vincent Lombard, Bernard Henrissat, Guillermina Hernandez-Raquet1. Metagenome-assembled genomes from a termite-derived lignocellulolytic microbial consortium provide new insights into the anaerobic degradation of biomass

Louison Dumond, Clémentine Duarte, Clément Salvagnac, L. Cézard, A. Majira, Betty Cottyn, David Sillam-Dussès, Edouard Miambi, Stéphanie Baumberger, Guillermina Hernandez-Raquet. Characterizing the lignin degradation ability of Nasutitermes ephratae gut microbiome and its derived microbial consortia enriched on wheat straw

Louison Dumond, Vincent Chassagnac, Laurent Cézard, Amel Majira, Betty Cottyn, Stéphanie Baumberger, David Sillam-Dussès, Edouard Miambi, Pui Ying Lam, Yuki Tobimatsu ,Gijs van Erven, Mirjam Kabel, Claire Hoede, Vincent Lombard, Bernard Henrissat, Guillermina Hernandez-Raquet. Characterization of a lignin-degrading microbial consortium derived from the gut of the termite *Nasutitermes ephratae*

Poster

L. Dumond, E. Flajollet, C. Salvagnac, L. Auer, A. Lazuka, G. Hernandez Raquet. Lignin degradation by microbial consortia derived from termite gut. Lignin Gordon Research Conference, 4-8th august 2018

Oral presentation

L.Dumond : Exploring the lignolytic potential in aerobic and anaerobic consitions of termite gut -derived microbial consortia

EAGS 2021, Tours, 27-29th October 2021

Remerciements

Car on ne parvient jamais à grand-chose tout seul, il convient de remercier tous ceux sans qui ce travail n'aurait pas pu se réaliser.

Je tiens tout d'abord à remericer Guillermina pour m'avoir donné l'opprtunité d'effectuer cette thèse. Merci pour m'avoir encadré pendant ces quatre ans et m'avoir permis de travailler dans les meilleures conditions possibles. Merci d'être toujours disponible et de m'avoir poussé à aller plus loin dans la compréhension des choses et de toujours viser plus haut. Merci également à Claire pour ton suivi et tes conseils avisés.

Je souhaite remercier tout particulièrement Emilie et Emeline qui n'ont pas partagé qu'un bureau mais aussi tous les instants de joie, de doute, d'excitation, d'échecs, de blind tests et de succès qui sont consubstantiels à la thèse au et en dehors du laboratoire. Merci aux stagiaires que j'ai pu encadrer, Clément, Clémentine, Vincent pour votre aide et tous les bons moments passés.

Merci à l'ensemble de l'équipe SYMBIOSE pour m'avoir accueilli avec bienveillance et bonne humeur. Merci à mes collègues et amis (Amr, Aldo, Béa, Erika, Irène, Léa, mathilde, Mattéo, Sidonie, Thomas…) pour avoir partagé ces quelques années. Merci aux permanents pour leur accueil. Merci tout particulièrement à l'équipe technique (Mansour, Simon, Evrard) sans qui les manips n'auraient jamais vu le jour.

Merci également à ceux avec qui j'ai pu collaborer, de près ou de loin sur cette thèse : les membres du projet Zelcor et notamment Stéphanie qui m'a lancé sur la lignine, les membres du PEPI, Vincent et Edouard pour leur aide sur les termites, Fabien Mounet, Gijs van Erven, Yuki Tobimatsu et Vincent Lombard.

Et comme la thèse n'est pas qu'un « sacerdoce », il me faut aussi remercier ceux qui ont été là en dehors du labo.

Merci d’abord à ma famille pour m’avoir soutenu au long de cette aventure, je garde une pensée particulière pour mes parents et mon frère, pour tous ceux qui sont arrivés ou nous ont quitté.

Merci à mes amis aussi, merci aux fermatosaures (Anaïs, Célia, Julie, Manon, Mélanie, Nils, Némo, Thibault). On a beau s’éloigner géographiquement vous n’êtes jamais bien loin.

Merci enfin à tous ceux avec qui j’ai pu partager ma passion de la musique, merci aux membres de l’OSET et plus particulièrement aux pupitres de cors qui se sont succédé. C’était une formidable aventure.

Table of content

INTRODUCTION

Au cours de la seconde moitié du XIXe siècle, l'industrie chimique a commencé à se développer. Rapidement, l'industrie s'est appuyée sur des ressources fossiles, d'abord le charbon puis le pétrole pour satisfaire les besoins de la société. Avec le développement de la pétrochimie, une multitude d'applications ont vu le jour. Les villes pouvaient désormais s'éclairer au kérosène, les premiers moteurs diesel présageaient d'une révolution dans les transports et les rues se pavaient d'asphalte pour accommoder les premières voitures. Un demi-siècle plus tard des médicaments tels l'aspirine étaient synthétisés et bientôt le plastique allait devenir la nouvelle panacée…

Plus d'un siècle et demi après que le pétrole jaillît du sol grâce aux premiers derricks, le pic d'extraction de pétrole n'est toujours pas atteint malgré la finitude des réserves. Cependant, les impacts environnementaux de l'utilisation de ressources fossiles telles que le pétrole, le gaz naturel ou le charbon ne peux plus aujourd'hui être dénié. Seule une transition de la pétrochimie vers des solutions plus durables pourra atténuer les effets du dérèglement climatique. Il devient alors nécessaire de trouver des solutions de remplacement.

C'est dans cette optique que la biomasse lignocellulosique s'impose comme une ressource de choix. C'est une source à la fois abondante et renouvelable. Elle peut être utilisée comme matériel, comme combustible, en substitution de produits de la pétrochimie ou pour des applications à haute valeur ajoutée. C'est en partant de ce principe que démarre le monde de la bioraffinerie. Par analogie avec les raffineries pétrochimiques, une bioraffinerie peut être définie comme une usine intégrée où de la biomasse végétale est transformée en un ensemble de produits variés.

La transformation de la lignocellulose présente cependant de nombreux challenges auxquels il convient de répondre afin d'améliorer les performances des bioraffineries. Non des moindre, la valorisation de la lignine, qui représente près d'un tiers de la lignocellulose représente un défi de taille étant donné sa récalcitrance. Ce défi présente cependant un intérêt certain étant donné l'immense variété de molécules que sa valorisation permettrait d'obtenir.

Si la dégradation de la lignocellulose, et plus encore de la lignine, paraît complexe, il est des organismes qui en ont maîtrisé le fonctionnement depuis des

millions d'années. Champignons et bactéries possèdent tout l'équipement nécessaire à cette tâche. Pour cela elles disposent d'un arsenal enzymatique complexe leur permettant de transformer cellulose, hémicellulose ou lignine en énergie. Par ailleurs, des phénomènes d'interactions leur permettent d'interagir entre elles et avec leur environnement pour permettre une spécialisation. Dans la nature, les microorganismes fonctionnent en communauté. Si l'existence de ces communautés microbiennes a longtemps été ignorée, l'Homme en a pourtant tiré profit de dès l'Antiquité. Les produits fermentés tels que le fromage, vin, le pain, la bière ou le kimchi sont autant d'application de communautés microbiennes à notre avantage, longtemps avant que l'on soit en mesure d'étudier leur fonctionnement.

L'étude des communautés microbiennes, appelée écologie microbienne peut nous permettre de mieux comprendre le fonctionnement des écosystèmes spécifiques impliqués dans la dégradation de la lignocellulose. Depuis la découverte de la structure de l'ADN, les techniques de biologies moléculaires ont permis des avancées majeures et notamment le séquençage des génomes. Aujourd'hui, grâce aux méthodes de séquençage au débit, il est possible de séquencer l'ensemble des génomes bactériens d'un écosystème ou métagénome. Ce nouveau champ de recherche, appelé métagénomique devient un outil de taille pour la compréhension des communautés microbiennes et peut être appliqué pour découvrir de nouvelles bactéries, de nouvelles enzymes, de nouveaux systèmes capables de valoriser la lignine.

Pour cela, il convient de s'intéresser en particulier aux écosystèmes naturels connus pour leur capacité à dégrader la lignocellulose. Ainsi, l'étude des communautés microbiennes qui habitent le sol, le rumen de certains mammifères ou encore le système digestif d'insectes comme les termites représente une opportunité de découvertes pouvant aider à la concrétisation de bioraffineries intégrées.

Le projet Zelcor a pour objectif la valorisation de coproduits issus de la bioraffinerie de la biomasse lignocellulosique. Dans ce cadre, plusieurs équipes de recherche à travers l'Europe œuvrent à mettre en place des approches combinées pour transformer lignines et humines en molécules à haute valeur ajoutée. Pour cela une approche en cascade est mise en place qui associe la valorisation chimique,

enzymatique ou biologique des substrats initiaux. C'est dans ce cadre que s'est déroulé la thèse. Plus précisément, au sein d'un groupe de travail impliquant la société Ynsect, l'INRAE et l'université de PARIS XII, l'objectif était de produire des bioréacteurs utilisant des termites ou des consortia microbien issus de termites pour valoriser des substrats ligneux. Les objectifs initiaux ont donc été fixés en considérant l'intégration de ce travail de recherche dans un projet impliquant de multiples partenaires.

Le premier enjeu de ce projet était d'obtenir un consortium microbien dérivé de tractus digestifs de termite possédant des capacités lignolytiques. La seconde problématique consistait à enrichir ledit consortium en fonctionnalités d'intérêt, en l'occurrence la capacité à dégrader la lignine. Cela devait permettre la mise en culture à une plus grande échelle de ce consortium afin de le caractériser finement. il a été prévu d'effectuer à la fois une caractérisation chimique fine des substrats pour évaluer l'impact d'un consortium sélectionné sur un substrat ligneux et une caractérisation métagénomique. Cette dernière devait permettre de comprendre le fonctionnement d'une communauté microbienne lignolytique issue d'intestin de termite, ainsi que d'identifier et de reconstruire reconstruire les génomes d'espèces bactériennes d'intérêt.

Afin de répondre à ces problématiques, le présent manuscrit est divisé en six chapitres. Dans un premier chapitre (Chapitre I), est présentée une revue de la littérature. La première partie présente la structure de la biomasse lignocellulosique et son intérêt dans le cadre de bioraffineries. La deuxième partie décrit les acteurs et mécanismes de la dégradation de la lignocellulose en milieu naturel et comment l'on peut en tirer avantage. Un soin particulier est apporté à l'étude de la dégradation de la lignocellulose dans le cas des termites. La troisième partie, enfin, discute des différentes approches métagénomiques applicables à l'étude des communautés microbiennes.

Le deuxième chapitre (Chapitre II) est dédié aux protocoles expérimentaux mis en œuvre tout au long de la thèse.

Les chapitres suivants (Chapitres III-VI) présentent les résultats obtenus et la discussion associée sous forme d'articles de recherche. Chacun d'entre eux se propose de répondre à une partie de la problématique. Le chapitre III (article publié) apporte de nouvelles preuves du rôle potentiel du microbiote intestinal de termites supérieurs dans la délignification de paille de blé. Le chapitre IV (en préparation) se propose d'utiliser la métagénomique « shotgun » pour caractériser la fonctionnalisation d'un consortium microbien dans la dégradation de lignocellulose. Le chapitre V (en préparation) décrit le processus de crible effectué afin de sélectionner un substrat permettant d'observer la dégradation de lignine puis l'évolution des communautés microbiennes lors de l'enrichissement par batch successifs. Le chapitre VI (en préparation) se concentre sur la caractérisation fine du processus de dégradation de paille de blé prétraitée par le consortium enrichi préalablement obtenu. Y sont développés à la fois l'analyse chimique du substrat et l'analyse par métagénomique « shotgun » du consortium.

Enfin, une conclusion générale discutant des principaux apports de ce travail de thèse et proposant des pistes de réflexion pour l'approfondissement des thématiques étudiées clôture se manuscrit.

Ce projet de recherché a été financé par le programme Horizon 2020 de l'union européenne à travers le consortium Bio-based industries dans le cadre du projet Zelcor (bourse No 720303). Il a été réalisé dans l'équipe SYMBIOSE au sein de Toulouse Biotechnology Institute (TBI). Cette thèse a été réalisée en collaboration avec l'Unité de Mathématique et Informatique Appliquées, plateforme GenoToul Bioinfo (INRAE), des équipes de l'institut de recherche et développement (IRD) à Bondy, de l'institut Jean Pierre Bourguin (INRAE), de l'Université de Wageningem et de l'institut de recherche sur l'humanosphère durable (Univeristé de Kyoto).

I. LITERATURE REVIEW

I.1. THE LIGNOCELLULOSIC BIOMASS AND ITS POTENTIAL VALORISATION IN A BIOREFINERY

Biorefinery is not something new. Vegetable oils, paper, fermented beverage are all examples of biorefineries. But the industrialisation of biorefineries was closely linked with the food industry and the chosen feedstock were wheat, corn and potato for starch, sugarbeet and sugarcane for sugar and then ethanol production.(de Jong et Jungmeier 2015)But the competition between food and fuel is not sustainable on a planet where 9 billion people could leave in by 2050. This has led to numerous research programs and incentives aiming at using the lignocellulose to produce bioethanol without a too strong competition for land use (Directive (EU) 2015/1513).

I.1.1. Composition and structure of the lignocellulosic biomass

Lignocellulose is a polymer structuring plants, it is the most abundant biopolymer on earth. It has the crucial role of guarantying the mechanical properties of the plant cell wall. It provides both mechanical strength to allow the plant to grow while keeping plastic properties allowing for cell wall turgescence.

The lignocellulosic biomass consists mainly of three polymers namely cellulose, hemicellulose and lignin interwoven with each other to form a complex three-dimensional structure as shown in **Figure 1**.

Cellulose, hemicellulose and lignin compose the large majority of the lignocellulosic biomass but their relative content can greatly vary. Its composition usually comprise 30-50% cellulose, 10-30% hemicellulose and 10-30% lignin. Such composition is completed by ash, proteins and wax(Malherbe et Cloete 2002). A more in depth comparison of the composition of several biomasses of interest is presented in the **table 1.**

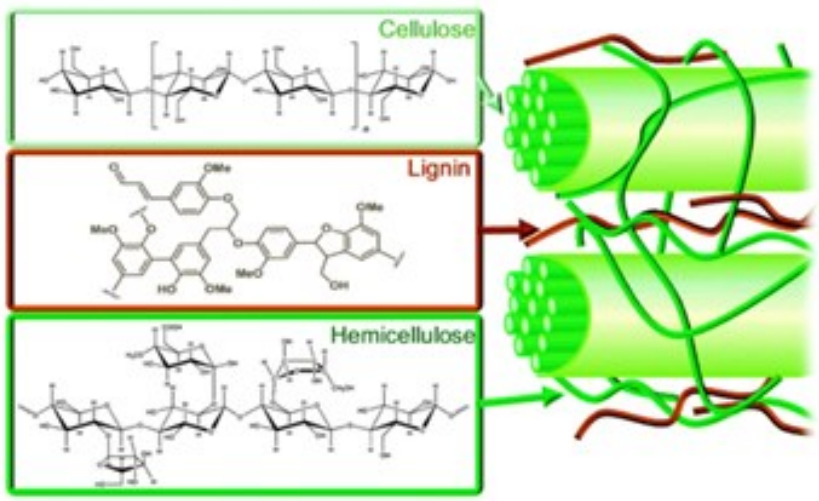


Figure 1: Structure of lignocellulosic biomass and relationship between the different polymers.(Alonso, Wettstein, et Dumesic 2012; Kobayashi et Fukuoka 2013)

Corn stover, wheat straw and rice straw share some particularities in term of polymer composition and lignin monomer content with higher proportion of guiacyl compared to syringyl units (also known as G/S ratio) and noticeable hydroxyphenyl (H) unit content (Table 1). Other herbaceous crops also have a high G/S ratio but with a higher cellulose/hemicellulose content. Softwood have a high cellulose and lignin content but low hemicellulose. Their lignin is characterized by the presence of almost only G units. Finally, hardwoods tend to have a high cellulose and lignin content. They have a G/S ratio close to 0.33 and no to very small amount of H.

Table 1: Composition in cellulose, hemicellulose and lignin of various biomass of interest (left side) and monomers recovered from the same biomass by thioacidolysis (right side).

Substrate	*Cellulose*	*Hemicellulose*	*Lignin*	*H*	*G*	*S*	*references*
Agricultural byproducts							
Corn stover	25-45	15-20	10 -20	5-16[a] 15-16[b]	33-43[a] 50-53[b]	48-54[a] 31-35[b]	(Berchem et al. 2017; Min et al. 2014)
Wheat straw	30-50	20-30	10-20	2-6	45-64	30-55	(del Río et al. 2012; Ai et al. 2016; H. Isikgor et Remzi Becer 2015; Lapierre, Pollet, et Rolando 1995; Herbaut et al. 2018)
Sugarcane bagasse	35-45	20-30	15-25	2-3	37-38	60	(Godin et al. 2010; Rezende et al. 2011; Maryana et al. 2014; del Río et al. 2012)
Rice straw	30-40	15-25	10-20	13	37	50	(Aita et Kim 2010; Godin et al. 2010; Lapierre, Pollet, et Rolando 1995)
Dedicated herbaceous crops							

Miscanthus	45-55	25-50	5-20	0-3	60-65	35-40	(Godin et al. 2010; van der Weijde et al. 2017; Herbaut et al. 2018; Bauer et al. 2012)
Bermuda grass	25-30	30-35	5-25	-	-	-	(Aita et Kim 2010; Xu, Wang, et Cheng 2011)
Switch gras	30-40	20-30	15-20	4	55	45	s(Aita et Kim 2010; Xu, Wang, et Cheng 2011; H. Isikgor et Remzi Becer 2015; Harman-Ware et al. 2016)
Wood							
Spruce	40-50	5-20	25-30	2	98	traces	(Aita et Kim 2010; Godin et al. 2010; H. Isikgor et Remzi Becer 2015; Lapierre, Pollet, et Rolando 1995) (softwood)
Douglas fir	45-55	3-20	25-30	0*	100*	0*	(Aita et Kim 2010; Godin et al. 2010; H. Isikgor et Remzi Becer 2015; Yoo et al.

							2016) (softwood)
Poplar	40-55	15-25	15-30	0-2	25-40	60-65	(Aita et Kim 2010; Godin et al. 2010; H. Isikgor et Remzi Becer 2015; Sannigrahi, Ragauskas, et Tuskan 2010; Herbaut et al. 2018; Harman-Ware et al. 2016) (hardwood)
Eucalyptus	45-55	10-20	20-30	0-2	33-56	42-66	(H. Isikgor et Remzi Becer 2015; Hamelinck, Hooijdonk, et Faaij 2005; Álvarez, Reyes-Sosa, et Díez 2016; Naron et al. 2017) (hardwood)

a=stem and cob, b=leaf, * data obtained by HSQC NMR

I.1.1.1. Cellulose

Cellulose is the most abundant polymer in plants, representing between 25 and 50% of the plant cell wall. It is a polymer of glucose linked by β(1-4) bonds which gives cellulose a structural role in the plant. This form enables intra and inter-chains hydrogen bonds. This makes cellulose a linear and rigid polymer with great

mechanical strength (**Figure 2**). In nature, cellulose can be found in an amorphous or a crystalline state.

Cellulose is assembled in glucose chains containing 2,000-25,000 glucose units. 36 of these chains are assembled in a microfibril which is stabilized by inter and intra glucose chain hydrogen bonds(Nishiyama et al., 2002). Microfibrils can also be assembled in macrofibrils. The degree of polymerization of cellulose can greatly vary depending on the origin, modifying the properties of cellulose.(Ioelovich 2008)

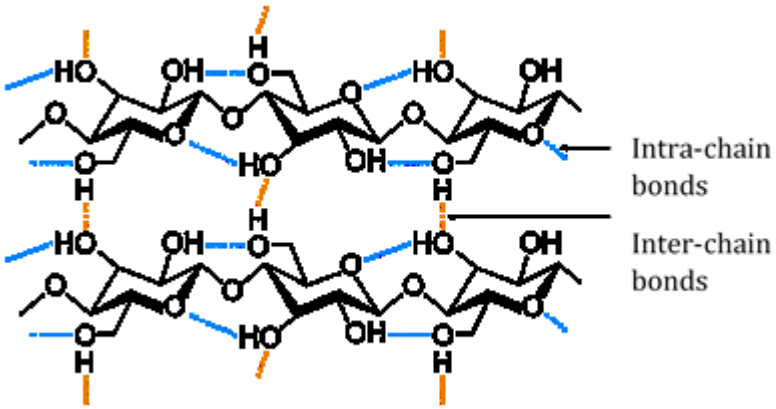


Figure 2: Structure of cellulose, a linear and rigid polymer

I.1.1.2. Hemicelluloses

Hemicelluloses are linear and branched polysaccharides that consist of xylose, glucose, mannose, galactose or arabinose with β(1-4) equatorial bonds. Their role in plant cell wall is to link cellulose with lignin and thus fortifying the structure while providing room for cell wall expansion. Several ramifications can occur on this linear backbone. Dicot cell wall mostly contains xyloglucan in the primary cell wall and glucuronoxylan is the main hemicellulose of secondary cell wall and thus the primary hemicellulose of hardwoods(Scheller et Ulvskov 2010). In both cases, the content of hemicelluloses can vary from 15 to 30 % (Table1). Grasses contain from 20 to 40% of hemicelluloses that are glucuronoarabinoxylans (GAX). Finally, softwoods are characterized by their galactoglucomannan content. The possible ramifications of hemicelluloses are related to the origin of the hemicellulose. For instance, grass hemicelluloses usually have ferulic acids that are bound to lignin. A typical wheat

hemicellulose is represented in **Figure 3**. Glucuronoarabinoxylans possess a xylan backbone containing acetyl, glucuronyl, and arabinosyl side groups. Its composition may differ between the primary and the secondary wall as most GAX have fewer decorations in the secondary cells that allows for a stronger interaction with cellulose and thus contribute to an increased resistance(Vogel 2008).

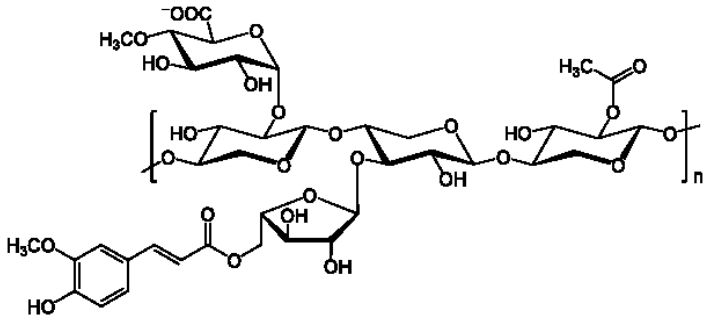


Figure 3: Structure of a wheat straw hemicellulose, a branched heteropolymer.

The biological role of hemicellulose mainly consists in helping the cell wall structuration, in conjunction with cellulose and lignin.

I.1.1.3. Lignin

Lignin is the third main component of plant cell wall. It is the most abundant source of aromatics on earth. Lignin is synthesized from three monolignols namely p-coumarylic, coniferylic and sinapylic alcohols. From these monolignols derived the phenylpropane units hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units (**Figure 4**), which are the elemental bulding blocks of lignin.

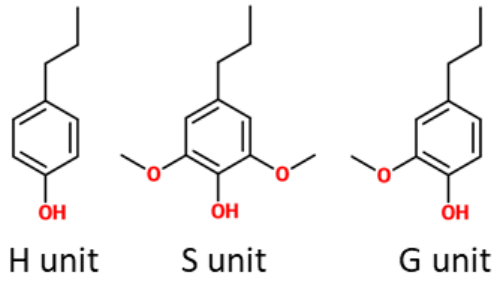


Figure 4: Phenylpropanoïd monomeric units of lignin: hydroxyphenyl (H), guaiacyl (G) and syringyl (S)

Although the mechanisms for lignin synthesis in plant cell wall have attracted lot of attention,(Ralph et al. 2004; Vanholme et al. 2010) they are not yet fully elucidated. Lignin remains a randomly assembled heteropolymer whose structure is highly dependent on its origin. For instance, softwood contains more than 90% of G units, up to 3% of H units and less than 1% of S units. Conversely, hardwood may contain from 50 to 75% of S units, 25 to 50% of G units and traces of H units(Constant et al. 2016) (**Table1**). The three phenylpropanoïd units are assembled by enzymes to form the complex polymeric lignin (F**igure 5)**. Different types of linkages can be found in lignin. The β(O-4) aryl ether linkage is usually the most abundant and the most fragile; biphenyl ether are also C-O linkages, C-C linkages include β-β resinol and β-5 phenylcoumaran.The proportion of different linkages is essential in comprehending lignin recalcitrance(Morozova et al. 2007). Hence, a more condensed lignin, with more C-C bonds, is more difficult to degrade compared to a lignin with more C-O bonds.

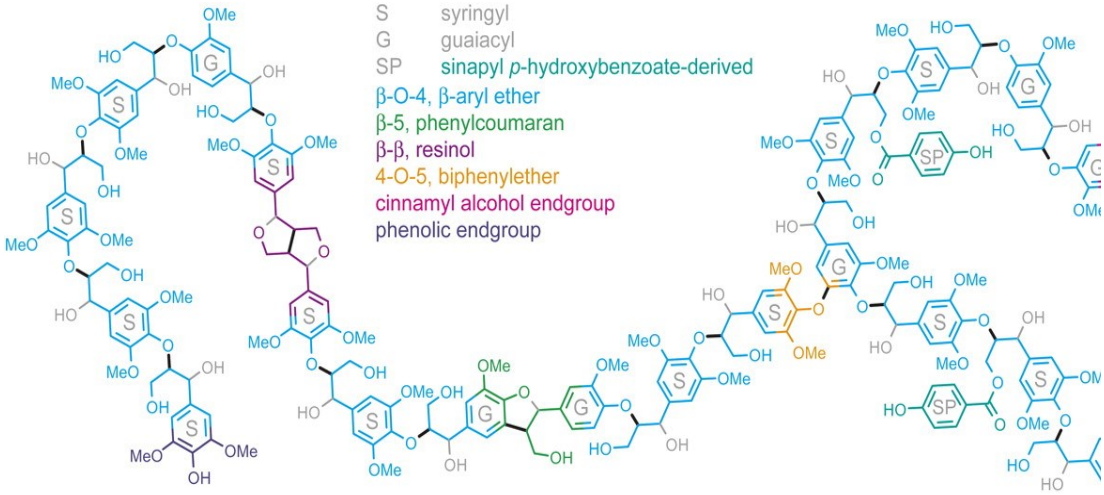


Figure 5: Structure of lignin showing the different units and linkages.(Vanholme et al. 2010)

Lignin also contains side chains involved in binding to other lignocellulose components **(Figure 6**). Such side chains notably include p-coumarate, ferulate and p-hydroxycinnamyl alcohols. Some grasses such as wheat or oat also contain a

significant amount of tricin, a flavonoid that has been assumed to serve as a nucleus for the polymerisation of lignin in grasses(Karlen et al. 2018).

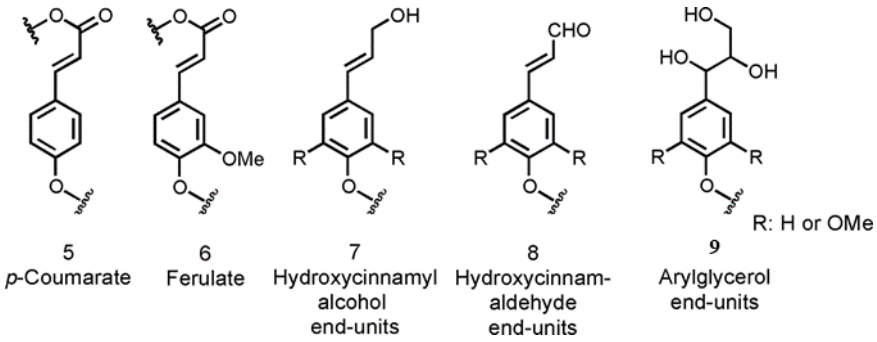


Figure 6 : Side chain structures in lignin end-units(Katahira, Elder, et Beckham 2018)

I.1.2. Lignocellulose biorefinery

I.1.2.1. The concept of biorefinery

The term of biorefinery was defined as "an overall concept of a processing plant where biomass feedstock is extracted and converted into a spectrum of valuable products". Its main objective is to transform lignocellulosic biomass into completely new valuable products or in products that can replace those typically obtained from petrochemicals. The development of the biorefinery looks for (*i*) guaranteeing energy security at a national or global level by countering the depletion of oil reserves, (*ii*) acting against climate change by limiting CO_2 emissions, and (*iii*) revitalising rural areas (Cherubini 2010). A biorefinery works in the same way as traditional refineries (Figure 7): The lignocellulosic raw material coming from agriculture, forestry or municipal waste, is fractionated into its building blocks, cellulose, hemicelluloses and lignin. These are then transformed into final products using a wide range of physical, chemical, biochemical or thermochemical processes.

I.1.2.2. Available substrates for a lignocellulosic biorefinery

In order to be used in a lignocellulosic biorefinery, the biomass used should be considered as renewable according to the definition given by the United Nations Framework Convention on Climate Change (UNFCCC) in 2006 (UNFCCC 2006).

The world's agricultural land represented 37.4% of the total of land areas in 2015.(« FAOSTAT » 2015) To be used in a biorefinery, agricultural products should not compete with food uses and must take into account their current functional use. Hence, various types of agricultural products can be considered such as wheat straw, corn stover, sugarcane bagasse or rice straw. Energy dedicated crops can also be included. They account for only 1% of arable lands and could reach up to 20% of arable lands in Europe by 2030 without harming the environment or threatening the equilibrium between food and energy (Holm-Nielsen et Ehimen 2016). Energy dedicated crops are high-yielding perennial C4 herbaceous crops. Unlike more traditional crops, they can be planted on marginal land with little need for inputs and provide ecosystem services, making them a good addition(Mitchell et al. 2016). Switchgrass (*Panicum virgatum L*), giant reed (*Arondo donax*) and miscanthus (*Miscanthus giganteum*) are the most expected. Sorghum, an annual crop, could also be a good solution on drier land(Rooney et al. 2007; Mitchell et al. 2016).

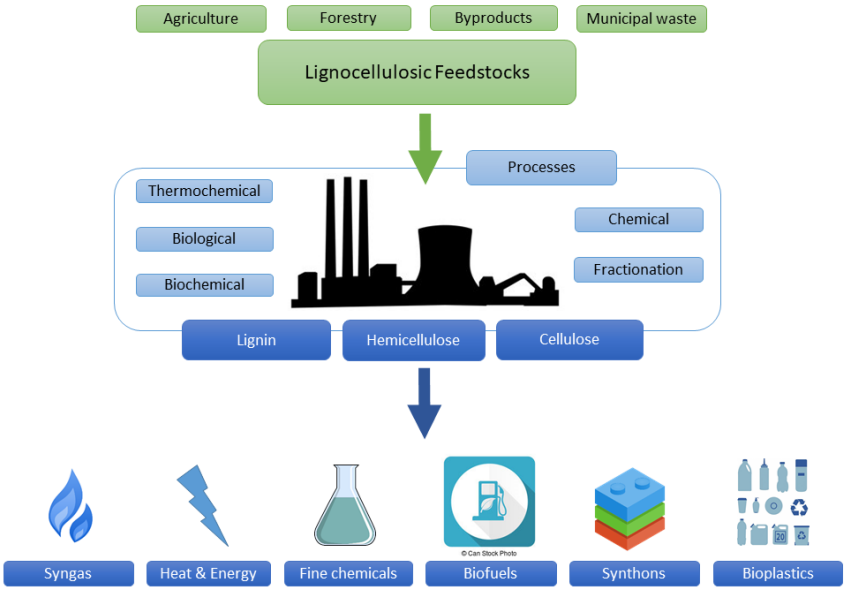


Figure 7: Schematic representation of a lignocellulosic biorefinery showing the potential feedstock and their transformation into various end-user products.

The potential contribution of forestry to biorefinery cannot be overlooked. Forest areas represent 30% of the lands. Furthermore, a biorefinery concept is completely suitable and complementary to a pulp paper factory. Some of the streams can be deviated to an integrated process. In this optic, lignin and hemicellulose streams that does not inter the pulping process are highly suitable(Moshkelani et al. 2013; Bajpai 2012). At the same time, forest residue could be used(Okello et al. 2013).

In addition, apart from perennial forests that are used for pulp and paper production, recent years have seen an increasing development of fast growing tree plantation, particularly short rotation coppice. Their use for biorefinery has shown its potential whether it is poplar, (Dou et al. 2017) willow (Guidi et al., 2013) or eucalyptus, (Gabrielle et al. 2012)they can also be planted in more traditional forests on marginal lands.

At last, animal manure (Chen et al. 2005) or solid municipal waste are also potential feedstock for lignocellulosic biorefinery (Sadhukhan et Martinez-Hernandez 2017).

I.1.2.3. The sugar platform

Today, sugars are the most easily and widely valorised fractions of biomass, used to produce ethanol or butanol.

In first generation ethanol biorefinery, sugars are obtained directly from sugarcane or sugarbeet or after an enzymatic pretreatment of wheat or corn. They are then fermented by genetically engineered *Saccharomyces cerevisiae* that are able to produce ethanol with a high yield(F. H. Lam et al. 2014). In more recent years, the focus has been on the development/isolation of efficient strains able of fermenting glucose and xylose(Matsushika et al. 2009).

In second generation bioethanol production, lignocellulosic biomass is used as substrate as not compete with food production. Sugars are less easily obtained, which reduce the overall yield of ethanol production. The production of sugars from lignocellulosic biomass is mainly achieved by physico-chemical pretreatment followed by enzymatic hydrolysis of biomass using an enzyme cocktail. The saccharification step is the most expensive step in biofuel production and is the less likely to be subject to cost reduction due to the price of enzymes and the high concentration required(Macrelli et al., 2012) .Therefore, many effort are being made to improve the pretreatment and fermentation steps.

Beside bioethanol, another biofuel has received a lot of interest over the last decade such as biobutanol that can be produced by *Clostridium acetobutylicum* and *Clostridium beijirenckii* through Acetone-Butanol-Ethanol (ABE) fermentation (Qureshi 2017).

I.1.2.4. The carboxylate platform

Enzymatic hydrolysis of lignocellulose is one of the limiting factors of the second generation biorefinery. It requires sterile conditions in order to avoid the loss of sugars in the process. To circumvent this step, it is possible to use microbial communities that directly degrade lignocellulosic biomass into short chain fatty acids (scFA) or carboxylates, which give a name to this platform (Agler et al. 2011). Carboxylates includes formate, acetate, propionate, butyrate, isobutyrate, valerate and

isovalerate, which could be transformed into solvents or bulk fuel after post processing.

I.1.2.5. Toward a lignin biorefinery

In most lignocellulose biorefinery scenarios discussed above, the lignin fraction of the biomass is only used for energy production, either by burning it or by converting it to syngas. Nevertheless, lignin has a very high potential for providing building blocks to the chemical industry (**Figure 8**). Therefore, a potential scenario for lignin-based biorefinery is to recover lignin from the conventional lignocellulose conversion process and direct it to another stream, as depicted in **Figure 8**.

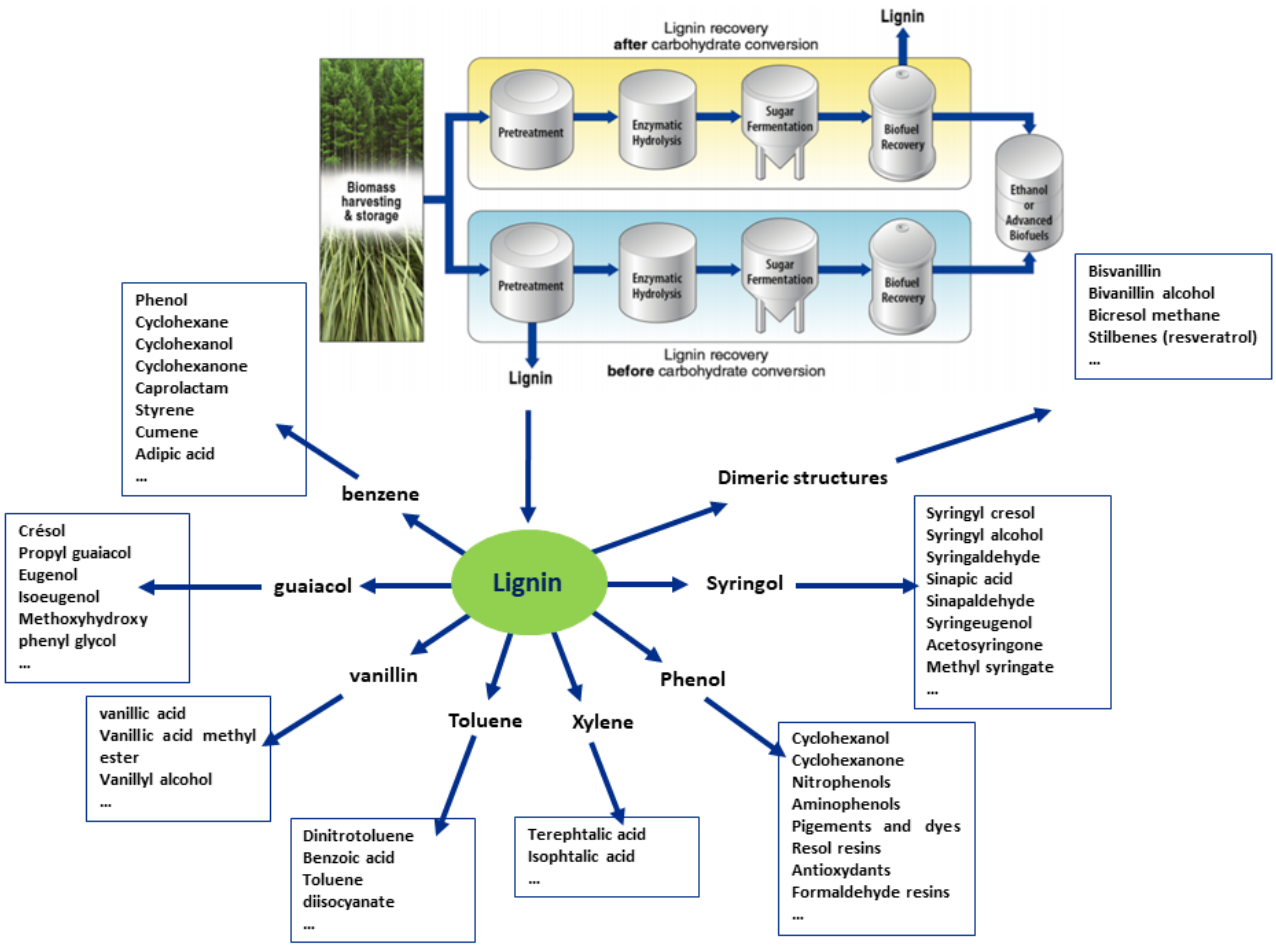


Figure 8: Schematic representation of the different steps of a lignocellulosic biorefinery. The focus has been on the recovery of lignin and the various chemical that can be obtained from it (adapted from Ragauskas et al, 2014 and Sadhukan et al, 2014).

Firstly, lignin can be upgraded into value-added materials such as carbon fibers, a lightweight composite material with multiple applications. However, lignin-based carbon fibers have shown poor mechanical properties until now(Ragauskas et al. 2014). Lignin can also be used in polymers as a stabilizer or to reinforce the structure(Polat et al. 2017). Lignin nanoparticles can also be used to make composite polymers and can even improve their properties(Hilburg et al. 2014). Lignin can also be blended with other polymers to tailor their properties. As an example, lignin has been used as an additive to impart antioxidant activity to packaging materials (Domenek et al. 2013) or as a sunscreen in cosmetics (Qian et al., 2015).

Secondly, lignin can be transformed into various chemicals (**Figure 8**) (Sadhukhan, et al., 2014). These include polymer chemistry building blocks such as terephtalic acid, caprolactam or styrene, dyes, antioxidants, and aromas such as vanillin, which is the leading aromatic molecule consumed in the world, or active ingredients for medicine and cosmetics (resveratrol).

With the necessary paradigm shift towards a lignocellulose-based biorefinery, with full valorisation of lignocellulose fractions, many efforts have been made in recent years to understand how macro- and microorganisms transform lignin in Nature.

I.2. LIGNOCELLULOSE DEGRADATION IN THE NATURAL WORLD

I.2.1. Lignocellulose degradation in natural ecosystems

In Nature, lignocellulose degradation is achieved by the collaborative works of a multitude of symbiotic microorganisms, each organism providing complementary enzymes that help break down the complex structure of biomass. Lignocellulose is transformed under highly variable conditions, from aerobic to anaerobic, over a wide range of temperatures, pHs or redox potentials. This highlights the diversity that can be found in Nature and how it affects lignocellulose degradation(Cragg et al. 2015).

Higher macroorganisms, including herbivores animals depend on the ability of microorganisms to degrade biomass. For instance, herbivorous mammals can access

lignocellulose polysaccharides through the complex microbial communities they harbor in their gut flora. The best example is ruminants, whose stomach is compartmentalized and whose rumen is dedicated to lignocellulose digestion. Lignocellulose is first degraded by rumination, which breaks down the lignocellulosic biomass into smaller particles that are then fermented in the rumen into volatile fatty acids; these are then metabolized by the animal. In the rumen, the fermentation of lignocellulose is also accompanied by the production of methane by the rumen archaea.

Phytophagous insects such as termites use similar methods (to be discussed in section 2.2). The lignocellulose is crushed by the termite mandibles, which are also involved in the first step of digestion by endogenous termite cellulases. Then, in the hindgut, the ground and predigested biomass is transformed into volatile fatty acids by anaerobic fermentation. This step is performed by microbial consortia living in the termite gut(Andreas Brune 2014).

This type of endosymbiotic behavior can also be found in marine wood-borers, which have a symbiotic relationship with *γ-Proteobacteria* that allow them to digest scraped wood particles from ships (O'Connor et al. 2014).

In addition to prokaryotes, eukaryotes are also able to degrade lignocellulose. This is especially clear in fungus-growing termites in which lignocellulose is predigested by fungi before been assimilated by termites (R. A. Johnson et al. 1981). It is also the case of lower termites where most of the digestion is carried out by flagellated protists.(Hongoh et al.,2003)

It is therefore thanks to the efficient methods of lignocellulose degradation developed by bacteria and fungi that most lignocellulosic carbon is recycled in Nature.

I.2.2. Lignocellulose degradation by bacteria and fungi

I.2.2.1. Cellulose and hemicellulose degradation

Bacteria and fungi are very efficient cellulose and hemicellulose degraders. To do so, they use a variety of different enzymes capable of degrading lignocellulose. The

degradation of cellulose and hemicellulose requires the action of a complex enzymatic cocktail containing at least three types of enzymes (**Figure 9 A)**: Cellobiohydrolases (I and II) (EC 3.2.1.91), which act on the reducting-end of the glucose chains and catalyze the hydrolysis of glucosidic bonds to release cellobiose; endo β(1-4) glucanase (EC 3.2.1.6), which catalyzes the hydrolysis of the glucosidic bonds inside the chain, releasing two shorter cellulose oligomers and, β-glucosidase (EC 3.2.1.21) which catalyses the hydrolysis of cellobiose into two glucose monomers.

Hemicelluloses are heterogeneous polymers, its breakdown requires the concerted action of several enzymes (F**igure 9 B)** including endo β(1-4) xylanase and β-xylosidase, plus specific enzymes to break down every type of bond existing between the different side chain monosaccharides and the xylan backbone. In the example presented here, α-D-glucronidases, α –L-arabinofuranosidases and α-galactosidases are required. As hemicelluloses can also be acetylated, acetyl esterases are also needed. Finally, ferruloyl esterases are needed to remove ferulate from the hemicellulose.

All these enzymes have been classified as carbohydrate active enzymes (CAZymes), which are enzymes that degrade, modify or create glycosic bonds(Cantarel et al. 2009).Cellulases and xylanases belong to the category of Glycoside Hydrolase (GH), which are listed in the appendix of this chapter (**Supplementary Table 1**). Acetyl- and ferruloyl-esterases belong to the category of Carbohydrate Esterases (CEs). CAZymes can also act on other polysaccharides that can be found in plants but are not technically part of lignocellulose, such as starch or pectin.

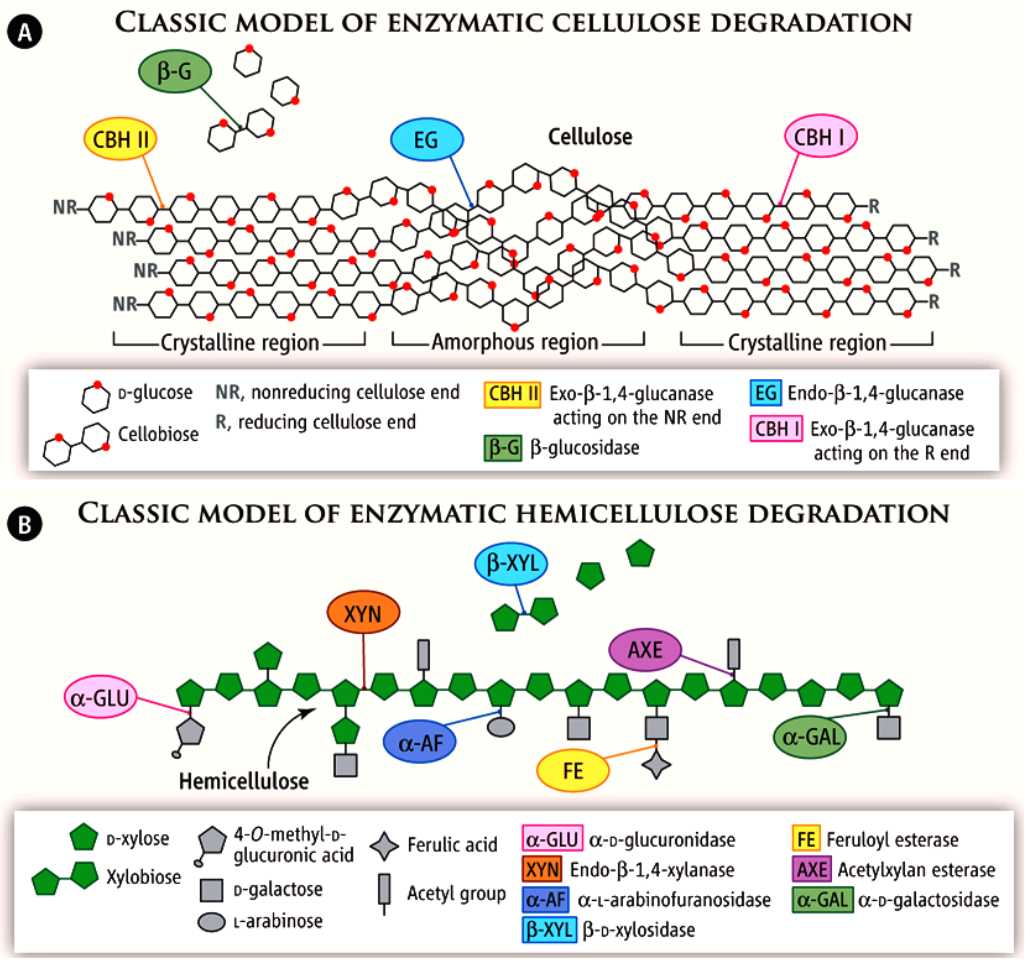


Figure 9: Schematic representation of the enzymatic hydrolysis of cellulose (A) and hemicellulose (B) with a focus on the different enzymes involved in these processes(Berlin 2013).

In order to optimize lignocellulose degradation, many bacteria such as *Clostridium cellulovorans* (Doi & Tamaru 2001) presents multiprotein complexes called cellulosomes where many catalytic sites are linked to a single scaffolding protein. This type of structures have been identified in a variety of anaerobic bacteria and fungi and confers an advantage for lignocellulose degradation(Artziet al., 2017; Haitjema et al. 2017).

Bacteroidetes are prime lignocellulose degrader that have also optimized their lignocellulolytic capacity by placing most of the genes dedicated to polysaccharide degradation on polysaccharide utilization loci (PULs), allowing them to efficiently release a large variety of enzymes(Lapébie et al. 2019).

Bacteroidetes, *Spirochaetes, Firmicutes* and *Fibrobacteres* have all been identified as key phyla in lignocellulose degradation in various environments and conditions.

I.2.2.2. Lignin degradation by fungi and bacteria

Lignin degradation by fungi has been studied for almost a century. It was first described by observation of wood decay. As early as the 1930's, using beech heartwood as substrate, Campbell proposed a theory in which an oxidation occurred on lignin, leading to decay decolouration of the rot(Campbell 1930). This lead to the now controversial classification of white and brown rot fungi(Riley et al. 2014). Wood-rotting fungi generally include fungi from the phyla *Basidiomycota* and *Ascomycota*. They are classified into brown rot fungi that leave the lignin largely intact, and white rot fungi that have the ability to degrade the lignin. White rot fungi include widely studied fungi such as *Phanerochaete chrysosporum, Trametes versicolor, Pleurota ostreatus* or *Tricoderma reesei*(Riley et al. 2014).

Fungi mainly degrade lignin by secreting lignolytic enzymes. Among them, lignin peroxidases (LiPs) (EC 1.11.1.14) catalyses the oxidation of a substrate using hydrogen peroxide as a co-substrate. They are produced by members of the *Basidiomycetes* phylum such as *Phanerochaete chrysporium*(Tien & Kirk 1984) or *Bjerkandera sp*. (ten Have Rimko et al. 1998) and by actinomyces bacteria such as *Streptomyces viridosporus*.(Ramachandraet al., 1988) The high redox potential of LiPs allows them to catalyse the oxidation of phenolic and non-phenolic compounds. Their optimum pH is generally around pH=3 (Pollegioni et al., 2015). Manganese peroxidases (MnPs) (EC 1.11.1.13) are structurally and functionally similar to LiPs. They are both heme peroxidases. They were discovered in *P. chrysosporium* (Glenn & Gold 1985). Unlike LiPs, organic acids, and especially oxalate are thought to act as chelating agents, necessary to stabilize the Mn^{3+} during catalysis(Perez & Jeffries 1993).Versatile peroxidases (VPs) (EC 1.11.1.16), which were discovered in *Pleurotus eryngii* (Hofrichter et al. 2010) can act in the same way as both LiPs and MnPs. However, their mode of action can change with pH and Mn^{2+} concentration, making them unpredictable and difficult to use(Knop et al. 2016). Another peroxidase found to be active on lignin, is dye-decolorizing peroxidases (DyPs) (EC 1.11.1.19).

They were discovered by Kim and Shoda, 1999 and are characterized by their ability to degrade dyes. This type of enzymes is thought to be involved in lignin degradation by fungi (Salvachúa et al. 2013) and bacteria (Ahmad et al. 2011).

Laccases (EC 1.10.3.2) are enzymes that have been studied since the end of the XIXe century(Morozova et al. 2007).They were found in plants where they are involved in lignin biosynthesis. Laccases were also found in several fungi known to degrade wood such as *Trametes versicolor* or *Pleurotus eryngii*. They are copper oxidases that catalyse the reaction described in equation [1]. Their ability to degrade lignin is still discussed as they can cleave lignin model compounds when a mediator is involved but fungi whose laccase were inhibited could still degrade lignin(Thurston 1994).

[1] $4\ benzenediol + O_2 = 4\ benzosemiquinone + 2H_2O$

Laccase and peroxidase have been the most studied enzymes involved in lignin degradation but other enzymes, while not degrading lignin directly, can act as auxiliary enzymes(Green 1977). proposed glucose oxidase as an enzyme that can collaborates with laccase in lignin degradation. This cooperation model with laccase (Szklarz et Leonowicz 1986) and LiPs (J. Lan et al. 2006) was confirmed subsequently. The role of this enzyme is to provide H_2O_2 that is used as a substrate by LiPs and participate in lignin degradation by producing hydroxyl radicals by Fenton chemistry. Fenton chemistry is thought to be involved in brown rot fungi(Bugg et al. 2011). Other oxidoreductases including veratryl alcohol oxidase (VAO) have also shown to participate in lignin depolymerization by *Pleurotus osteatus* by preventing repolymerization of lignin(Marzullo et al. 1995). At last, a dehydrolipoamide deshydrogenase from *Thermobifida fusca* has also shown a capacity to prevent lignin from repolymerizating (Rahmanpour et al. 2016).

I.2.3. Lignin degradation by termites

As mentioned earlier, lignocellulose degradation occur in very different organisms in Nature. In this section, we will focus on the description of lignocellulose degradation by the termite gut microbiome, which is the main topic of this thesis.

I.2.3.1. Termites

Termites belong to the order Blattodea. Termites are among the most common insects on Earth and are particularly common in tropical and sub-tropical regions of the world. They are known to feed on lignocelluosic biomass, but their diet can differ from species to species. The most common diets are those that feed on wood, litter, and grass(Mikaelyan et al. 2015).

Termites are social insects living in a cast system divided in reproducers, workers and soldiers. Reproducers are equipped with wings that allow them to fly to found a new nest far from their original nests, workers are tasked with various missions such as food gathering or nest construction and soldiers have to protect the colony.

Termites are divided into 7 families (**Figure 10**): *Rhonotermitidae*, *Serritermitidae*, *Kalotermitidae*, *Termopsidae, Hodotermitidae* and *Mastotermitidae* form the "lower" termites while *Termitidae* forms the "higher" termite family. Termites have adapted different strategies that shape their classification. For example, the separation between "lower" and "higher" termites is the loss of symbiotic protists in the gut of the latter.Zhange et al.) *Macrotermitinae* have separated from *Termitidae* by acquiring a mutualism with a *Termitomyces* fungus that they grow outside their body and which helps them degrade lignocellulose. Other *Termitidae* have also adapted by feeding on different substrates such as soil or litter.

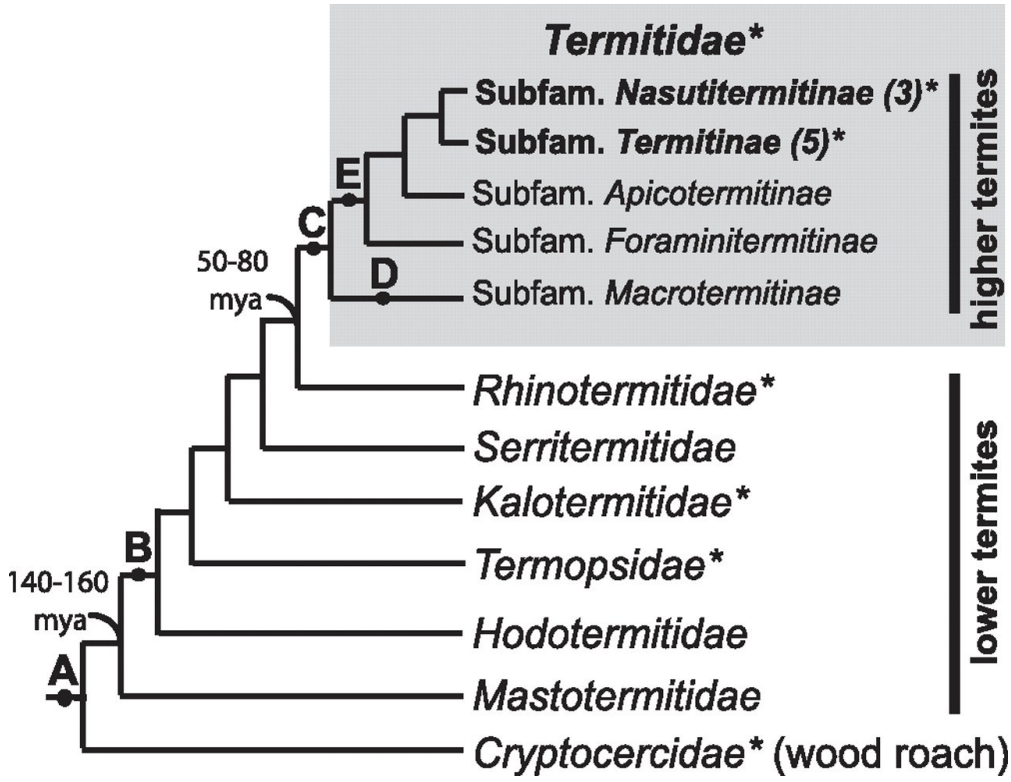


Figure 10: Cladogram of major termite families and Termitidae subfamilies(X. Zhang et Leadbetter, s. d.)

I.2.3.2. Digestion of lignocellulosic biomass in termites

In all termites, lignocellulose degradation comes from the symbiotic action of the termite and their gut microbiome. As mentioned earlier, anaerobic flagellates located in the hindgut are responsible for lignocellulose degradation in lower termite (**Figure 10**). These flagellates plays a specific role in providing enzymes to break down lignocellulose(Andreas Brune 2014). Some bacteria are also found in the gut of lower termites. This notably include *Spirochaetes,* which represent half of the total diversity in lower termite *Reticulitermes speratus.*(Hongoh et al.,2003) Other bacteria mainly consisted of *Termite group I* (endomicrobia associated with the flagellates), *Bacteroides* and *Clostridia.* Overall, the symbioses between protists, prokaryotes and the termite host is essential for lignocellulose degradation in lower termites.

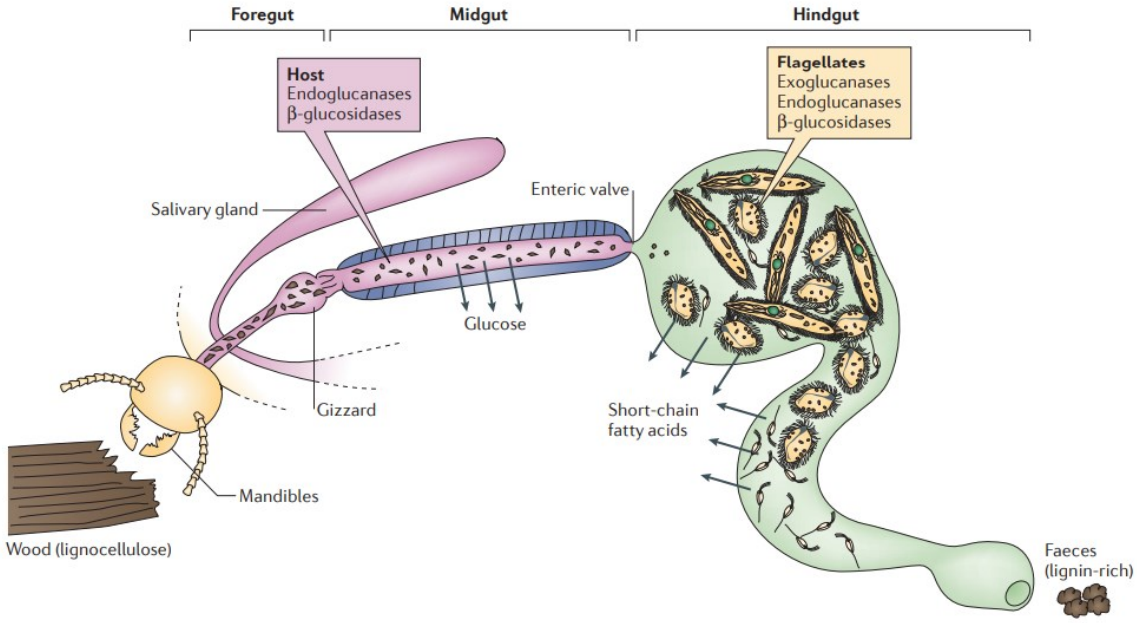


Figure 11: Schematic representation of lignocellulose digestion in lower termites(Andreas Brune 2014). *Lignocellulose is first ground into small particles by the mandibles then travel through the midgut where a first digestion is performed by the host enzymes and finally are phagocytated by the flagellates which convert them into volatile fatty acids that can be used by the host for energy.*

In higher termites other than *Macrotermitinae*, the flagellates have disappeared and the gut is much more clearly compartmentalised **(Figure 12**) and mainly harbours prokaryotes. As in lower termites, part of the digestion is achieved by the mechanical destruction by the mandible and endogenous cellulases (endoglucanases and β-glucosidases) produced by the host (Tokuda et al. 2018). Then, in the hindgut, the main stage of digestion is achieved by endosymbiotic bacteria. Depending on the substrates of the termites, the diversity of the microbial flora can change, with wood-feeding *Nasutitermes* having a microbiome dominated by *Spirochaetes* and *Fibrobacteres,* while soil-feeding *Termes* have a microbiome dominated by *Firmicutes, Bacteroidetes* and *Spirochaetes*(Andreas Brune 2014; Auer et al. 2017). The importance of *Spirochaetes* for hemicellulose degradation has been demonstrated in *Nasutitermes* (Tokuda et al. 2018) through their production of GH11 while *Bacteroidetes* (Yuki et al. 2014) are known to degrade cellulose and hemicellulose. The *Fibrobacteres* role

have been demonstrated in the degradation of cellulose and chitin(Abdul Rahman et al. 2016).

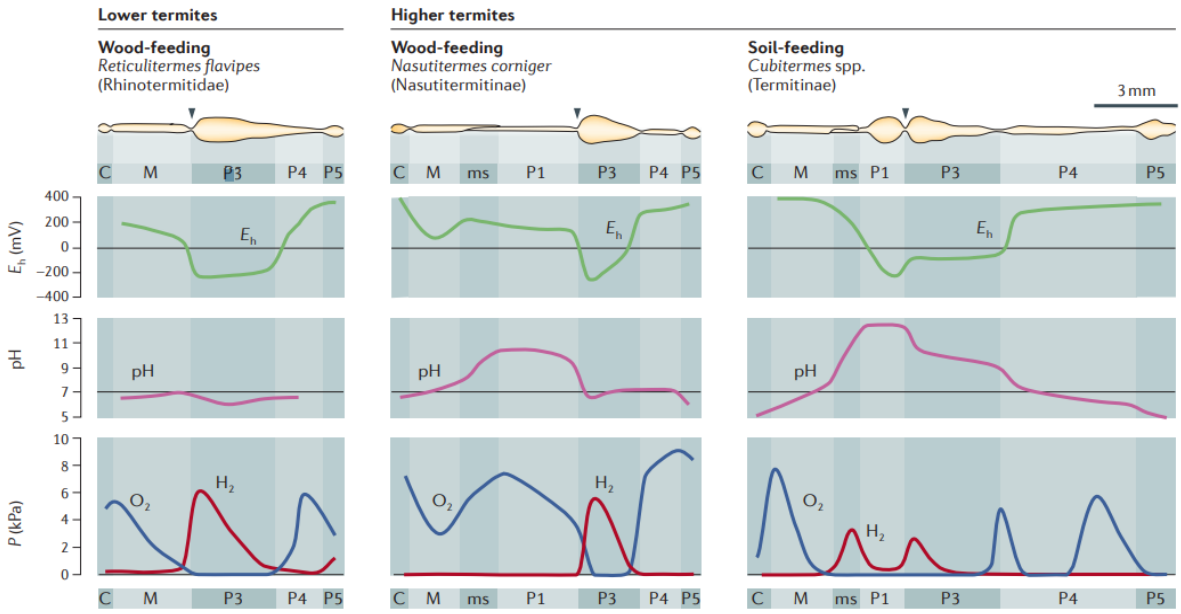


Figure 12: Gut compartmentation in termites and physico-chemical conditions of the gut(Andreas Brune 2014).

In both lower and higher termites, most lignocellulose degradation is done in anaerobic condition but the structure of their gut allow for a gradient of oxygenation and some part are in microaerobic or aerobic conditions (**Figure 12).**

Overall, lignocellulose degradation in termites is achieved by the collective action of the termite and its microbiome, whether it is in association to fungi or protists and prokaryotes or only prokaryotes. The microbial activity appears as the main driving force behind the lignocellulose degradation.

I.2.3.3. Lignin degradation in the termite gut: a century-old controversy

I.2.3.3.1. Evidence provided by experiments with ^{14}C-labelled substrates

Termites are well known for their capacity to degrade lignocellulose, especially in regions where termite infestation can cause significant damage to buildings. However, the available information on lignin degradation is poor or ambiguous.

It was initially thought that there was no lignin degradation in termites. Nevertheless, Butler and colleagues (Butler et Buckerfield 1979) have shown degradation of ^{14}C-labelled lignin and lignin model compounds by *Nasutitermes exitiosus*. The authors used ^{14}C-labelled natural or synthetic lignin and ^{14}C-labelled aromatics as feed for the termites. Labelling was carried out either on the aromatic ring, the methoxyl group or the β-carbon. Degradation varied greatly between the eight types of labelled substrates, ranging from 7% for ring-labelled sodium phenate to more than 60% for ring-labelled ferulic acid and methoxy group-labelled. Corn lignin was also labelled on the different positions, showing degradation of 15% to 32% depending on the location of the labelling. However, as remarked by Breznak and Brune,1994, these techniques are not without drawbacks. For instance, the pioneers Butler and Buckerfield, 1979 labelled lignocellulose without a quantifying the amount of non-lignin molecules labelled along with lignin. Thus, even when this quantification is done correctly, it is difficult to tell whether the ^{14}C-CO_2 respired is from high molecular weight lignin or from monomers. It has been suggested that monomers that were included in lignin during biosynthesis should be labelled in order to reduce errors in ^{14}C delivery. But even in this case, it remains difficult to know the exact fate of these precursors, which could just as easily be directed to carbohydrate metabolism, leading to an overestimation of the actual degradation of lignin. Synthetic lignin can help solve these problems, however, as close to the real structure as they are, dehydrogenation polymers (DHPs) cannot reproduce the heterogeneous structure of the original substrate that plays a decisive role in lignin recalcitrance. Therefore, it is not trivial to decide whether the degradation of DHPs is accurate evidence of lignin degradation.

I.2.3.3.2. Limits of termite lignin-feeding experiments

It was shown as early as the 1930's that during digestion, lignin is accumulated in termite faeces and is not degraded. Studies using these enriched faeces as substrate have shown little degradation, whether the assessment was made using ^{14}C-lignin (Cookson 1987) or more recently carbon NMR(Hopkins et al. 1998). Cookson and colleagues (Cookson 1987) investigated the degradation of three different ^{14}C-labelled lignins by *Nasutitermes exitiosus*, *Coptotermes acinaformis* and *Mastotermes darwinensis*. Lignin degradation was monitored by $^{14}CO_2$ production. It was found

that *N. exitios*us was able to degrade lignin up to 5% but negligible amount of $^{14}CO_2$ was released from the lower termites. This may be due to the fact that lower termites contain mainly protists. So far, no lignin degradation has been found in protists (Tarmadi et al. 2017; L. Xie et al. 2012) although they may have an activity toward crystalline cellulose(Yuki et al. 2015).

Lower termites, and in particular *Cryptotermes brevis* and *Coptotermes formosanus* have received much attention in the last decade. Tropical woods were used as a substrate to assess the fate of lignin when digested by *Cryptotermes brevis*(Katsumata et al. 2007). No degradation was found, although some modification of the lignin were observed with a slight increase of the S/G ratio. Some methoxyl groups were found outside of the Bjorkman lignin content, furthermore the Bjorkman lignin contained less methoxyl group suggesting polymerization of guaiacyl units through C-C bonds. Similar results were found on *Coptotermes formosanus* (Ke et al. 2011). A similar approach was also taken by Geib et al.,2008. The faeces of *Zootermopsis angusticollis* were collected after feeding on live wood. These faeces were analysed by pyrolysis in the presence of tetramethylammonium hydroxide (C^{13}-NMR) coupled to gas chromatography-mass spectrometry (Py-TMAH-GC/MS (C^{13}-TMAH-GC/MS)). However, contrary to previous studies significant lignin depolymerisation and demethylation was detected.

Lignin modification by *Coptotermes formosanus* was compared for different lignin feed sources (softwood, hardwood and poaceae). It was found that guaiacyl units were modified in softwood, guaiacyl-syringyl units in hardwoods and that the lignin of barley straw was not modified(Ke et al., 2013)

Recently, Tarmadi et al., (Tarmadi et al. 2017; 2018) found evidence against the general idea that lignin is degraded through side-chain oxidation of β–*O*–4 linkages. By feeding *Coptotermes formosanus* with lignin as the sole carbon source, they demonstrated that *C. formosanus* could not use lignin as a nutrient. The same termite species was also fed with Japanese beech, Japanese cedar or rice straw; NMR analysis of the faeces showed a substrate dependence(Keet al., 2013). However, the data did not support the hypothesis of a side chain oxidation mechanism. Indeed, an increase of β–*O*–4 over β–5 and β–β bonding was observed in the faeces, indicating that the

termite might preferentially degrade C-C bonds. Similar results have also been reported for the fungus-growing termite *Odontermes formosanus*(Hongjie Li et al. 2017). These evidence contrasts with the lignin degradation process that occurs in wood decomposing fungi.

I.2.4. Harnessing the power of microbial ecosystems for lignocellulose degradation

As microbial consortia hold great promises for lignocellulose conversion, they have been implemented in various ways to study them. These methods range from strain isolation, to growing entire ecosystems in controlled bioreactors, to using omics tools to characterize community members and enzymatic activities(Sethupathy et al. 2021). These methods, with their own advantages and disadvantages, can shed light on the functioning of lignolytic microbial ecosystems.

I.2.4.1. Cultivation of microbial consortia

One of the most common method of obtaining information on the specific activities of a microbial community is to culture it. In general, the first approach is to isolate bacteria with desired activity. Isolation has enabled the identification of many lignocellulolytic fungi and bacteria over the years. However, not all bacteria can be cultured (Wade 2002) which necessarily set limits to what can be studied in this way.

To overcome this problem, it is possible to grow microbial consortia in conditions as close as possible to their native environment. The use of lignocellulolytic consortia in bioreactors has been extensively studied as a means to best transcribing natural conditions into artificial environment.

I.2.4.2. Selection and enrichment of lignocellulolytic microbial consortia

Bioreactors are powerful tools for deciphering the mechanisms of lignocellulose degradation in microbial consortia. However, when seeking to use

microbial ecosystems for industrial applications, the selection of a stable and efficient consortium is necessary.

To overcome the complexity of lignocellulolytic communities, one strategy is to enrich the functionally relevant fraction of the microbial community by cultivation under defined conditions. This approach minimizes the complexity of the ecosystem and increases the fraction of the community expressing a particular function, which is essential to ensure the stability of a process(DeAngelis et al. 2010). Most of the time, the enrichment process is performed in sequential batch process, using a small portion of the previous batch to inoculate the next batch until the diversity and activity of the consortium is stabilized. To further simplify the consortium, dilution-to-extinction methods can also be applied. In this case, an environmental consortium (or a previously enriched consortium) undergoes serial dilutions and the characteristics of the consortium after each dilution can be assessed(Dingrong Kang et al. 2020). The most diluted consortium that still manages to exceed a required activity threshold can then be selected.

Approaches to enrich lignocellulolytic properties of a microbial consortium have been successfully applied to environments such as sugarcane plantation soil (Moraes et al. 2018), cow rumen supplemented with wheat straw (Lazuka et al. 2015; 2018) or soil supplemented with plant-based.. However, different factors must be considered when doing an enrichment. The number of cycle, for example, should be long enough for the consortium to be stable. To achieve this, some have opted for a large number of cycles. When enriching a community to reach high xylanase activity, Peng and colleagues performed 45 cycles of enrichment. However, they found that the diversity did not change significantly over the last 35 cycles(Guo et al. 2010). Other studies have shown that 5 to 7 culture cycles were sufficient to achieve consortia stability(Lazuka et al. 2015).

Culture conditions such as pH, temperature or the substrate used to perform the enrichment can also drastically alter the results. For instance, Lima de Brossi and colleagues (de Lima Brossi et al. 2016) have performed 9 cycles of enrichment of a soil-derived consortium using three different substrates, wheat straw, switchgrass and corn stover, at pH=7.2, and wheat straw at pH=7.2 and 9. Lignocellulose degradation

was better with switchgrass enriched consortia than with wheat straw or corn stover and, using wheat straw as the substrate, better results were obtained at pH=7.2. In addition to the performance of these consortia, microbial diversity diverged radically as a core of only 5 bacteria and 3 fungi were found in each line.

Finally, while consortia enrichment is an effective method to obtain lignocellulolytic consortia, it must be noted that the consortia obtained can by very different from its initial source. For example, when enriching a microbial consortium from termite gut microbiome to produce volatile fatty acids, Lazuka and colleagues (Lazuka et al. 2018) observed that after a single round of enrichment, the consortium dominated by Spirochaetes and Fibrobacteres became dominated by Bacteroidetes and Firmicutes.

I.2.4.3. Synthetic consortia

While the enrichment processes aim at selecting a consortium derived from its original source, the process of creating a synthetic consortium differs as it aims to create a rational consortium based on knowledge of the interactions/complementarities that occur in natural lignocellulolytic ecosystems.

Building synthetic microbial consortia emphases the importance of synergies in lignocellulose conversion. This type of synergy can be obtained with as few as two miroorganisms. For example, to create synthetic communities of various complexity, Hu and colleagues (Hu et al. 2017) combined four fungal strains, complementary to each other or to an enriched microbial community from soil. The enriched microbial consortium enhanced the performance of the fungi by providing positive interactions. It was particularly noted that the synthetic consortium with *T. reesei* or *A. tubigensis* and the microbial consortium highlighted the potential of such method to design synthetic consortia. Other studies used bacteria isolated from sugarcane soil to test multiple combinations that resulted in a 5-strains synthetic consortium (*Stenotrophomonas maltophilia, Paenibacillus sp., Microbacterium sp., Chryseobacterium taiwanense, and Brevundimonas sp.*) that was an effective lignocellulolytic synthetic community. This could help us find out which interactions are most important in such a community(Puentes-Téllez & Falcao Salles 2018).

Finally, it is possible to engineer synthetic communities by adopting a "division of labor" approach. This can only be done if the metabolic pathways are known and aims to rationalise what might be the best match to cover each microbial deficiency (Roell et al. 2019).

I.3. NOTIONS OF MICROBIAL ECOLOGY AND METAGENOMICS

I.3.1. Microbial ecology and biodiveristy

Microbial ecology is a science that studies the interactions between microorganisms and with their environment. This discipline uses tools adapted from both ecology and microbiology. Thus, it involves measurements and characterization of biodiversity by different approaches including meta-omics approaches (discussed below).

The characterization of the biodiversity is done by calculating diversity indices. The notion of α, β and γ-diversity was introduced by Whittaker 1960.

α-diversity defines the average number of different populations (species or OTUs) within a single sample. The simplest measure of α-diversity is to count the number of different species in a sample, which is called richness. However, this kind of index is depends on the size of the sample representing the entire ecosystem. In the case of metagenomics data, it depends on the sequencing depth. To overcome this problem, other indexes that take in account other aspects of the community have been proposed.

One of the most used is the Chao 1 index(Chao 1984). Let S_{obs} be the number of observed species, $f_1$the number of singletons and $f_2$the number of doubletons. The Chao 1 index S_{chao1} which is an estimation of the total richness, is defined in the equation [1] or [2] for the bias corrected version. However, this index can lead to an overestimation when a large number of singletons are accounted. As singleton are often removed from the sequencing data, an index based on singletons count can lead to errors and the count of singletons needs to be corrected (Chiu et Chao 2016).

[1]$S_{chao1} = S_{obs} \cdot \frac{{f_1}^2}{2f_2}$ [2] $S_{chao1} = S_{obs} \cdot \frac{f_1(f_1-1)}{2(f_2+1)}$

To give a lesser weight to unusual species while taking abundance into account, the Shannon index H' and the inverse Simpson index $\frac{1}{\lambda}$, are frequently used. These indexes are described in the equation 3 and 4, respectively, where R is the richness and p_i is the proportion of the species *i*.

[3] $H' = \sum_{i=1}^{R} p_i \ln(p_i)$

[4] $\frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{R} {p_i}^2}$;

These indexes are more comparable between studies but Shannon tends to give more weight to abundant species.

To compare diversity between samples, the β-diversity indexes is use. It informs on the similarity (or dissimilarity) in the community composition between different samples. β-diversity is based on a distance matrix between samples. In microbial ecology, three main distances are calculated: Jaccard, Bray Curtis (BCij) (equation [5]) and Unifrac(Lozupone et Knight 2005).

[5]$BC_{ij} = \frac{2\min(S_i,S_j)}{S_i+S_j}$ where $\min(S_i, S_j)$ is the minimum abundance of an OTU between 2 samples.

The Unifrac index differs from the Bray-Curtis index by the fact that the distance between two samples is based on the branch length of species that are present in either one of the samples. Thus, it is based on phylogenetic distances. It is possible to calculate a weighted Unifrac that takes into account the difference of branch length of the phylogenetic tree, weighted by their relative abundance. Since they use phylogenetic data, Unifrac distances are a good way to calculate the similarity between two samples that different taxa composition. These various methods can be synthesized by whether they use categories vs phylogeny, presence/absence vs abundance (**Figure 13**)

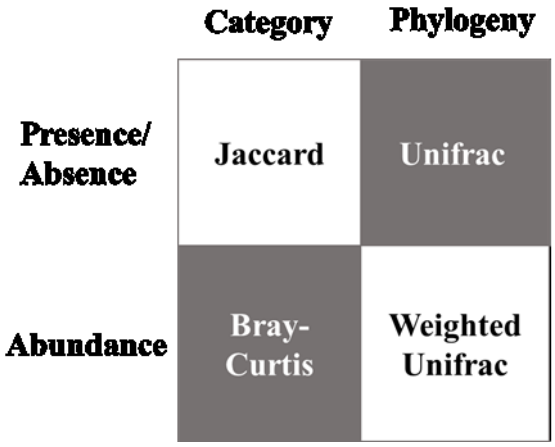


Figure 13: Main differences between measures of diversity distance**.**

The indexes previously discussed are good ways of comparing diversity in a simple way. However, one may want to see where the difference lies between samples. In order to do so, it is possible to make a map to discriminate the data. Principal component analysis (PCA) can be used to make a projection of the data and see how the variance is mostly explained. It is a cartography of the sample but may fail to discriminate. Partial Least Square-Determinant Analysis can be used to discriminate samples.

Some of the discriminant technics like PCoA or NMDS can also be used on a distance matrix.

I.3.2. Welcome to the world of "meta-omics" and NGS

Following the "genomics", which aim at studying the whole genome of single organisms, several other "omics" techniques have been born since the 1970's. For instance, "metagenomics" is interested in the metagenome, which is the whole pool of genes and genomes available in an environment. Many other examples can be noted: metatranscriptomics, metaproteomics, metabolomics, metainteractomics.

The metagenome can be described as the complete collection of microbial genes and genomes present in an ecosystem. Following the "omics" trend, "metagenomics" study the genome of an environment without isolating individual

strains. This approach has allowed for a reconsideration of the tree of life, shedding light on the overlooked diversity of microorganismes(Hug et al. 2016).

Investigating the metagenome requires a huge amount of sequencing power, and it is only when high throughput sequencing (Roche 454, 2005; Applied Biosystems SOLID, 2006; Illumina HiSeq, 2007) arrived on the market that research were granted access to this world. From then on, next generation sequencing (NGS) has never stopped to improve. In 2011, Illumina released MiSeq, the first lab table sequencer. More recently, in 2017 Oxford Nanopore brought the MinIon on the market, a sequencer of the size of an USB device. In recent years, a focus has been done on long read sequencing.

Currently, Illumina is the leading maker of short-read sequencers including MiSeq, HiSeq, NovaSeq and NextSeq series of instruments. (Goodwin, McPherson, et McCombie 2016; Bentley et al. 2008) BGI's MGISEQ and BGISEQ (Korostin et al. 2020)and Thermo Fisher's Ion Torrent sequencers (Lahens et al. 2017) provide alternatives. Short read sequencing allows the production of read up to 600 bases. However, these technologies struggle at resolving complex regions of a genome and fails on high GC and highly repeated regions (Pollard et al. 2018). To make up for these shortcomings, long read sequencing (Oxford Nanopore, PacBio and Chromium) have been improving quickly. While they were inhibited by high cost, high error rate and lack of IT tools available, the technologies are quickly improving and more and more software are developed to tackle the specificity of long read sequencing(Amarasinghe et al. 2020; Amarasinghe et al., 2021) .

All these technics allowed for a better study of microbial ecosystems.

I.3.3. Amplicon sequencing metabarcoding

The gene coding the16S rRNA unit has been the most used marker for the study of bacterial community since 1985(Lane et al. 1985). This is explained by the conserved nature of this housekeeping gene, that has not changed over time. It is also a gene that contains various conserved regions common to all or few taxonomic groups, and variable regions that are unique for each species.

16S rRNA gene is composed of 9 variable regions separated by conserved regions (**Figure 14).** Several studies have shown that most variability is contained in the V3-V6 regions(Su et al. 2017) making them the best target for species discrimination by sequencing. However, the length of this region is longer that the current NGS sequencing capacity (up to 300 bp). That is why the V3-V4 region has been preferentially sequenced for several years. *In silico* analysis of more than 440,000 16S rRNA sequences from the SILVA database has shown that it was possible to have a 40% sensitivity (true positivity rate) and 80% specificity (true false negative) with these regions(Martínez-Porchas et al., 2016). Universal primers have been proposed for sequencing these regions(Claesson et al. 2010).

However, choosing this region for diversity characterization can have some issues. Or instance, it has been shown that partial sequencing is not as reliable for taxonomy assignation than the sequencing of the complete 16S rRNA gene(M. Kim, Morrison, et Yu 2011). Indeed, the 16S rRNA gene is present as a multicopy gene with different ribotypes in a microorganism (Case et al. 2007). In addition, partial sequencing of the 16SrRNA gene is an efficient tool to identify genera but it not always allows to separate species.(J. S. Johnson et al. 2019),(Lan et al., 2016)

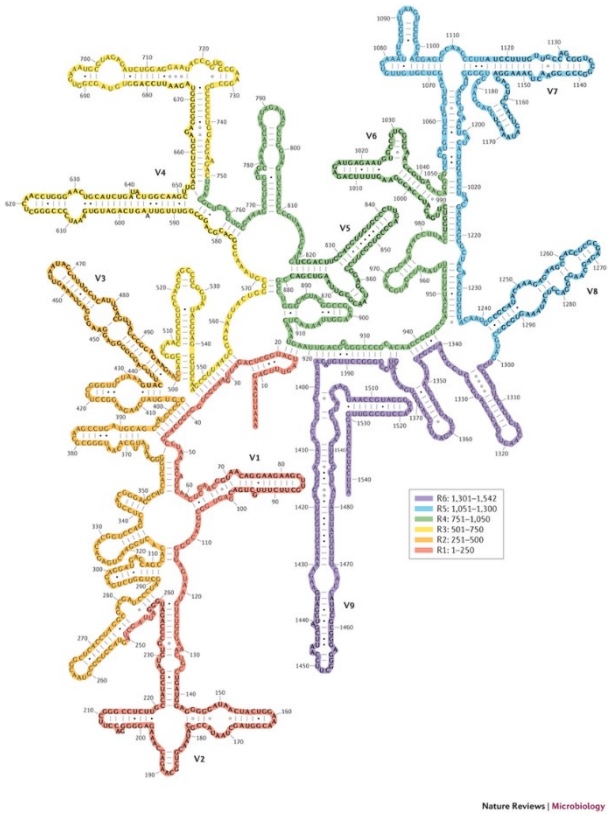


Figure 14: Secondary structure of the 16S RNA gene. The 9 different regions appear in a different colour(Yarza et al. 2014).

Other markers have been investigated : rpoB was shown to yield to the same diversity on Proteobacteria while having a better species cut-off(Vos et al. 2012) than 16S rRNA gene. Others genes such as *amoA, pmoA, nirS, nirK, nosZ,* and *pufM* have also been studied (Lan et al., 2016) allowing a better resolution of closely related species but they are less widely spread among prokaryotes. Other markers such as *gyrB* can be invaluable tools to resolve subspecies in some applications(Poirier et al. 2018).

Another noncoding genetic marker is the ITS (Internal transcribed spacer). This region is very variable between close related species but its use remains secondary; in contrast, it is very commonly used to decipher diversity in fungi(Raja et al. 2017).

Another approach can be done to find out marker genes that could replace 16S RNA. The use of complete genome sequencing has allowed the discovery of a broad range of universal marker which applications can take advantage of (AMPHORA, PhyloSift, AMPHORA2 , MetaPhyler , EMIRGE and PhylOTU)(Lan et al., 2016). The research of such markers takes advantage of whole genome sequencing to identify

genes that can be used as a marker for specific groups by assigning them different metrics such as universality (is it widely represented within the group?), evenness (is it present in the same number of copies within the group?), monophyly (is it specific of the group?), uniqueness (is it distinguishable for each member of the group?)Wu et al., 2013)

I.3.4. Functional metagenomics

Sequencing the 16S rRNA gene can efficiently answer to the question "Who is in my community?" However, in most of the cases it can bring no answer to the question "What do they do?" In order to tackle this problem, functional metagenomics can be applied. In this approach, DNA is extracted from the environment, large-size DNA fragments are selected and repaired before being ligated to a vector to construct fosmid or cosmid libraries (K. N. Lam et al. 2015). Such cured DNA fragment are then ligated to cos-based vector, packaged in a lambda phage and transducted to *Escherichia coli*. Alternatively, metagenomics libraries can be done with fosmids. When a large number of *E.coli* clones has been produced, they are screened for a specific activity, usually on chromogenic substrates. When it comes to lignocellulose conversion, screening is usually based on dye degradation. By implementing cell-sorting technologies such as microfluidics or fluorescence activated cell sorting (FACS), an important increment in throughput is expected(Taupp et al., 2011).

Several strategies can be used through functional metagenomics to discover new enzymes or enzymatic functions (**Figure 15**). The research of new functions and enzymes is crucial to better use a sequence-based research as many known domains annotated in databases are of unknown functions.(Ufarté et al., 2015) These technics can therefore be applied to the research of new lignin degrading enzymes. For instance, these methods have been successfully used to discover new oxidase from the rumen gut microbiota with an ability to degrade dyes and aromatics(Ufarté et al. 2018). By this approach, new transaminase active on monoaromatic compounds and capable of efficiently converting them into higher added value amines, have been successfully discovered(Pawar et al., 2018). However, this kind of technics have the drawback of

being limited by the chosen substrates and thus “only what was searched for can be found” (Heux et al. 2015).

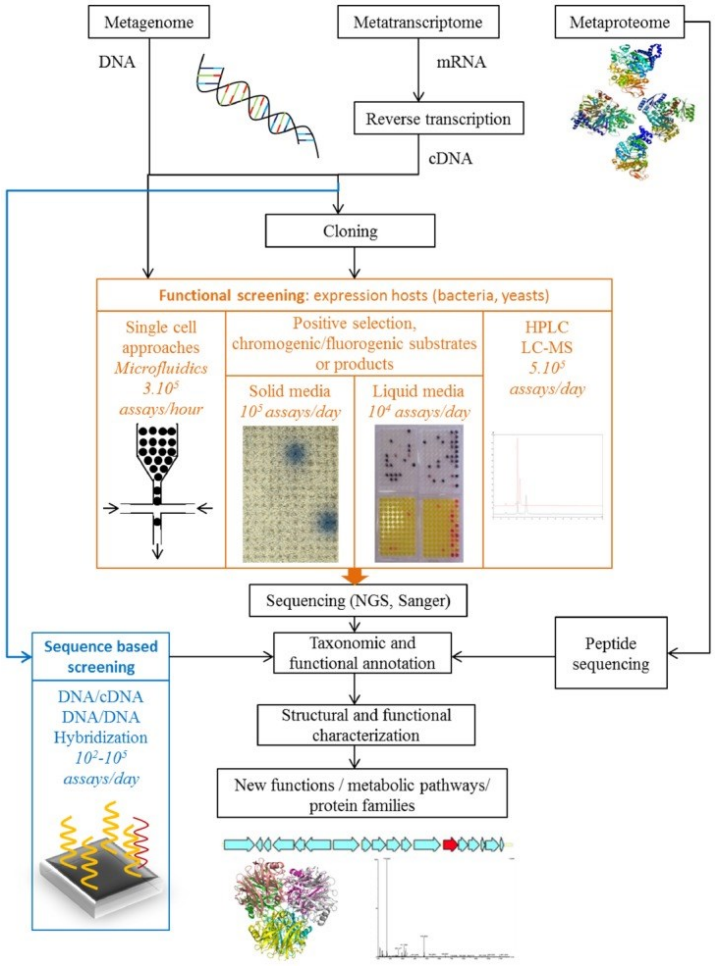


Figure 15: Strategies for the functional exploration of metagenomes, metatranscriptomes and metaproteomes to discover new functions and protein families(Ufarté et al., 2015)

I.3.5. “Shotgun” metagenomics

If functional metagenomics answer to the question “Can they do that?”, “shotgun” metagenomics try to answer the questions “Who is there?” and “What could they do?” Unlike functional approaches, “shotgun” metagenomics try to elucidate the potential activity of the investigated community. In “shotgun” metagenomics, instead of sequencing a fraction of the DNA, all the DNA information in an environmental

sample is sequenced, resulting in a large number of reads harbouring functional information.

“Shotgun” metagenomics typically comprise 5 main steps.(Quince et al. 2017)

1) DNA extraction, purification and sequencing from the environmental source.

2) Pre-processing of the sequencing reads. This step ensure data of a sufficient quality to go further in the characterization.

3) Taxonomic affiliation and functional characterization. This step aims at disclosing the data “within the data”, which means identifying the OTU present in the community, finding their functions etc. An example of analysis is shown in **Figure 16**, where the data analysis is based on regrouping individual reads in order to reconstruct whole genomes.

4) Statistical analysis. This step enables the scientist to interpret its data.

5) Cross validation.

Shotgun metagenomics makes it possible to reconstruct genomes and identify species in an environment, find theirs genes and thus understand the potential role of each microorganisms in an environment. However, it is not the only method available. In particular, when considering very big dataset, the construction of gene catalogues of a complex environment can give valuable information about the structure of the environment and mapping contigs onto known catalogues can be sufficiently informative(Commichaux et al. 2021; F. Xie et al. 2021). It is also possible to work directly on reads without assembling them. This can be useful for obtaining taxonomic and functional information in a shorter time. However, this may give less accurate results and is dependent on the database used to examine the reads(Tran et Phan 2020).

In the same way, it is also possible to do shotgun metatranscriptomics or metaproteomics.

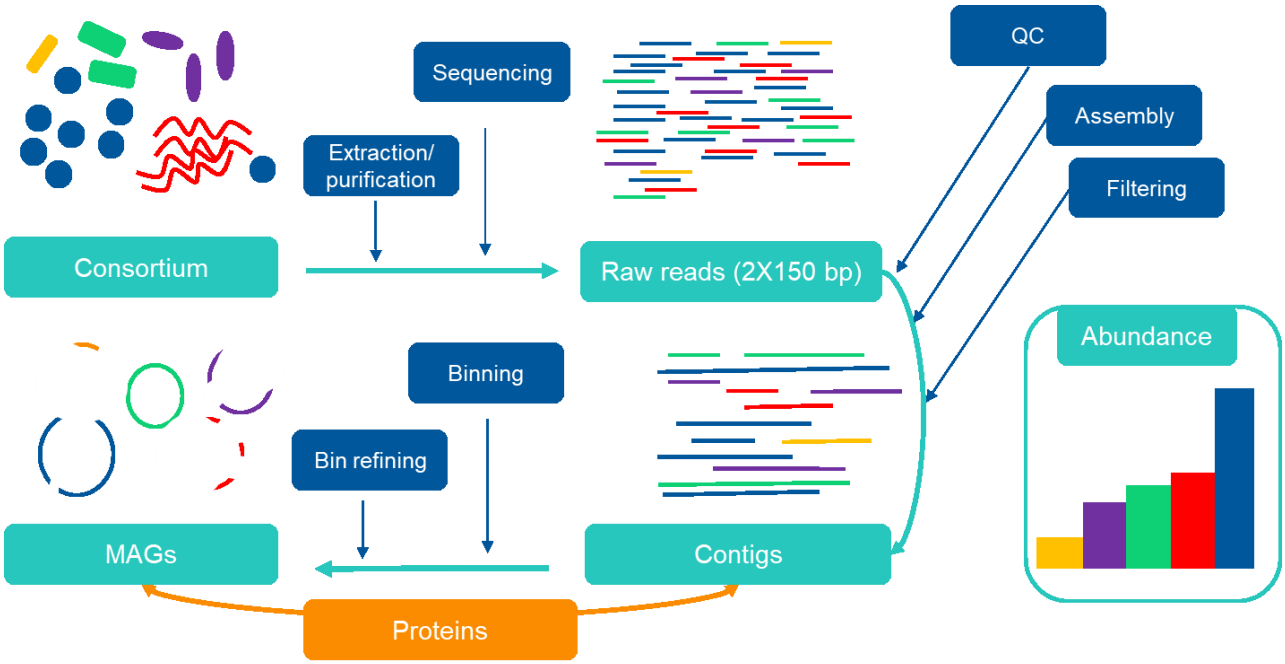


Figure 16: Schematic representation of shotgun metagenomics strategy. The microbial sample is sequenced into short reads. These short reads are then assembled into contigs that are later clustered into metagenome-assembled genomes.

I.4. NEW GENERATION SEQUENCING (NGS) DATA ANALYSIS

No matter what method is used to do metagenomics sequencing, NGS generates a significant amount of data that need to be analysed. The analysis depends of the sequencing technology as well as what is sequenced.

I.4.1. DNA extraction, sequencing technologies and their influence on the data

The first step in any metagenomics workflow is to define what kind of analysis is needed to answer a question. As such, the choice of both the DNA extraction protocol and the sequencing device is important as it defines both the format and the quality of the data and impacts the downstream analysis.

The DNA extraction method itself can introduce a bias that will affect both the quantity of reads obtained, their quality and the resulting taxonomy, as has been

demonstrated in the sequencing of 16S rRNA of oral microbiota(Teng et al. 2018) or soil.(Zielińska et al. 2017) Another key feature in DNA extraction is its ability to produce large fragments (3-40 kb) that are needed to construct fosmid/cosmid libraries(Kakirde et al., 2010). The use of bead-beating extraction methods produce shorter fragments that can be used in amplicon or shotgun sequencing but are not ideal for cloning large-inserts. With the rise of long-read sequencing, fragment size matters and there is needed to use DNA extraction methods that produce longer fragments(Jones et al. 2021).

Each sequencing device has its own key application and output. It is therefore necessary to know which kind of information is needed to choose the appropriate device to use. For instance, the Illumina platform offers the MiSeq series instruments, which have an output of 25 million of 2X300 bp reads. This instrument is best suited for 16S rRNA metabarcoding. HiSeq instruments can provide as much as 6 billion of 2X150bp paired reads, which is a basis for shotgun metagenomics, and NovaSeq 6000 advertises up to 20 billion reads on a S4 flow cell. This requires a suitable computer architecture and data storage to analyse the data, which must be taken into account for sequencing.

Roche 454 pyrosequencing can produce longer reads (400 bp) than HiSeq but this technology introduces insertions/deletions due to its inability to accurately detect homopolymers. This means that downstream analysis will slightly differ.

Long read sequencers such as PacBio produce reads of size higher than 100kbp. This type of sequencing also have a high error rate and necessitate different softwares to analyse the results.

Another factor influencing NGS data analysis is the depth of sequencing. Both taxonomic profiling and relative abundance of identified species are highly dependent on sequencing depth, as rare microorganisms may be lost with low sequencing depth(Pereira-Marques et al. 2019). As a corollary, if identification and characterization of very low abundance species is a key element of an analysis, a high depth is required.

An optimal depth of coverage should be around 50 X (between 20X and 200X) for a sequencing single unknown genome.(Ayling et al., 2019; Desai et al. 2013) However, when the abundance and the number of genomes expected to be found in a sample are unknown, it is difficult to accurately define the depth of sequencing to apply.

On the other hand, if genomes reconstruction is the main objective, it is necessary to consider the limitations and specificities of assembly softwares. For instance, for algorithms based on overlap-layout-consensus, high depth may drastically increase computational time, and for algorithm using de Bruijn-graphs, a higher depth may correlate to a higher error rate.(Lapidus & Korobeynikov 2021) In this case, downsizing may be relevant.

Hence, DNA extraction methods, sequencing technologies and sequencing depth used for sequencing will have a tremendous impact on NGS data analysis and must be recognised.

I.4.2. Metabarcoding

I.4.2.1. Databases

The main usage of metabarcoding sequencing using marker genes such as 16S rRNA is to identify OTUs based on a marker gene. However, this require to query sequences in a database. On this basis, the richness and accuracy of a database is of paramount importance. Indeed, for an OTU to be identified, it is necessary that the correct sequence is registered in the user's database. This also has the side effect of favouring more common markers such as 16S rRNA over niche molecular markers for which an accurate database may be lacking.

Several rRNA database concurrently coexist. SILVA (Quast et al. 2013) is a database that provides ribosomal RNA sequences for bacteria, archeae and eukarya. It includes datasets for small (16S/18S, SSU) and large subunit (23S/28S, LSU). It has the advantage of being quite exhaustive as well as being regularly updated (last version SILVA release 138.1). It currently provides 9,469,124 sequences (510,508 with NR99 i.e. filtered at 99% identity to reduce

redundancy of the database) as well as hosting the LTP database (Living Tree Project) (Yilmaz et al. 2014) a small curated database of known bacteria up to the strain level. The choice of database must be selected in the light of the experiment as LTP is more suited to known environments but poorly informative for unknown ones.

Other database for 16S includes Greengenes, EZbiocloud and the 16S rRNA database from the NCBI. These curated databases may be used when the precision of the data is of critical importance.

Database for other markers are also available. For ITS database such as UNITe or the NCBI database provides dataset for fungi identification.

Finally, it is possible to use a database constructed with a specific environment in mind. For instance, there is a database, now integrated on SILVA, targeted at the xylophagous insects microbial flora: DictDB (Mikaelyan et al. 2015). This type of database can be useful to analyze datasets from similar environment as it allows an easy comparison.

I.4.2.2. Main steps and associated tools for metabarcoding analysis

Data from metabarcoding provides reads on a fasta or fastq format. This initial file must be transformed into relevant biological information through different steps using dedicated software. A brief summary of the main steps and relevant tools is presented here.

I.4.2.2.1. Pre-processing and quality control of the data

Quality control is a crucial step as the quality of the analysis directly depends on the quality of the analyzed dataset. First of all, most of the time, NGS sequencing require all different samples to be mixed during the sequencing process. In order to circumvent this, samples are indexed in a process called multiplexing. It is then necessary to remove such indexes in the pre-processing of the dataset. It is also necessary to remove sequence-adapters that were generated through PCR; the fact that each fragment possess the same sequence may interfere with, for instance, the alignment process.

Some popular tools to remove adapters include Cutadapt (Martin 2011), Timmomatic(Bolger et al., 2014) and AdapterRemoval v2.(Schubert et al., 2016)

Quality control also involves the analysis of sequence quality, sequence length, GC content, the presence of adaptors, ambiguous bases, and overrepresented k-mers in order to detect sequencing errors, PCR artefacts or contaminations. This type of information can be provided by tools such as FASTQC(« FASTQC. A quality control tool for high throughput sequence data | BibSonomy » s. d.) or NGS GC toolkit(Patel et Jain 2012) and allows for a filtering of the data based on the quality of the sequences.

I.4.2.2.2. Chimera removal

Chimera are artefacts generated by PCR when the polymerase is detached from its fragment during the elongation process and is hybridising on another fragment. As such, a new sequence is produced that is a chimera of 2 or more different fragments (**Figure 17**). Chimeras are relatively common in metabarcoding and have a sizeable impact on the diversity and richness observed on samples as they represent new OTUs that have no real existence. They must be detected and removed in order not to have false results.

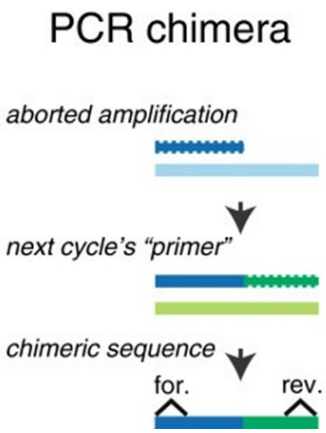


Figure 17: **Principle of chimera formation** (Fichot et Norman 2013)

This can be done based on two different ways. The first method is to query the fragments of a sequence in a database and perform an alignment (e.g. with BLAST). If different fragments from a same sequence are assigned to different taxonomies, the

sequence is considered chimeric. On a similar basis, *de novo* identification is performed by constructing a database from the sequences themselves. As chimeras are side PCR products, it is assumed that they are not chimeras. The sequences with the higher abundance are then used as controls and the sequences are aligned with the control by decreasing abundance. When a non-chimeric sequence is detected, it is added to the control pool. In all cases, removing chimeras on the basis of alignment always means working with compromises. For instance, setting loose parameters allows more information to be retained at the cost of inducing more false positives.

UPARSE(Edgar 2013), UCHIME2(Edgar 2016) or CATch (Mysara et al. 2015) are within the most used tools for chimera detection.

I.4.2.2.3. Clustering

Clustering sequences into OTUs is one of the most important steps. It involves grouping together reads that are thought to come from the same source. Once the reads have been clustered, they are considered to be a single. All subsequent analyses are then performed on the newly formed OTU instead of reads, which both reduces the computational time and greatly simplify the analysis. However, the parameters used for the clustering have a clear impact on diversity, as the larger the clusters, the less diversity is observed. It is therefore crucial to think about the parameters and software to be used(Falentin et al. 2019) .

Most clustering tools use sequence identity as their main metrics but the way to use it can vary greatly as shown in **Figure 18**. Two main strategies have been considered for OTU clustering: using a heuristic approach identifying centroids and a clustering threshold (usually 97%) (a) or identifying seeds and using single linkages to iteratively form a cohesive OTU (b).

The first method is used by programs like USEARCH(Edgar 2010), VSEARCH(Rognes et al. 2016) or CD-HIT(Fu et al. 2012). It has the advantage of being able to process all dataset, at the cost of overestimating the number of OTUs(Wei et al. 2021). On the other hand, SWARM (Mahé et al. 2015) uses the second approach, which identify OTUs more accurately but can be slow on large datasets process all datasets.

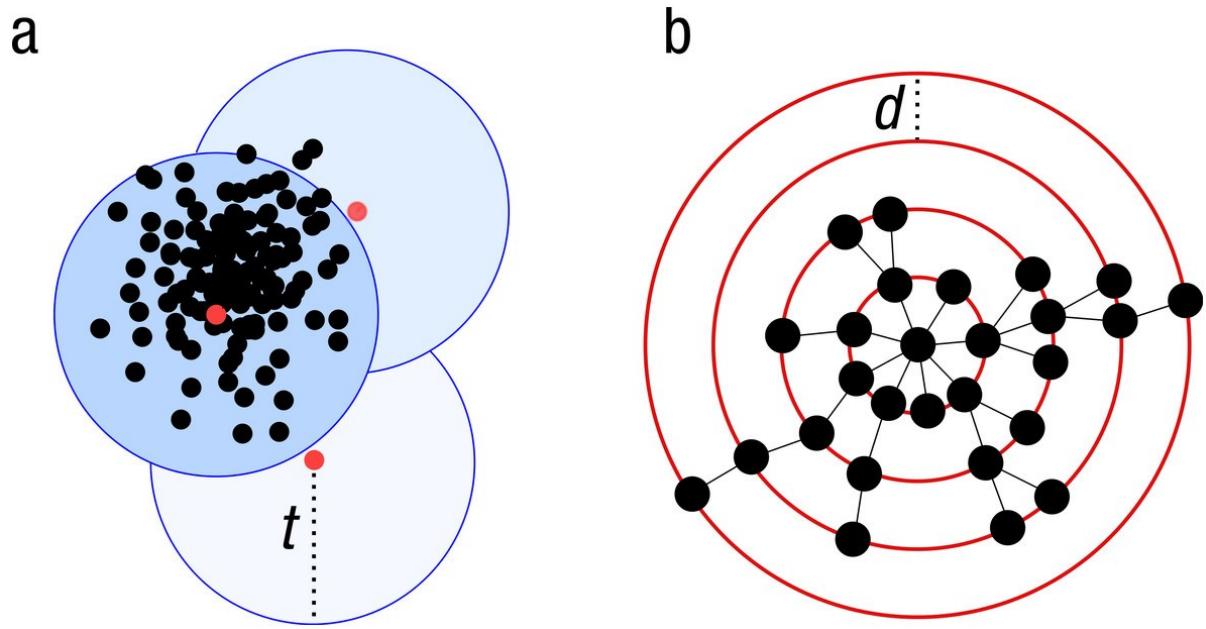


Figure 18: Schematic view of the greedy clustering approach and comparison with swarm.

(a) Visualization of the widely used greedy clustering approach based on centroid selection and a global clustering threshold, t, where closely related amplicons can be placed into different OTUs.

(b) By contrast, Swarm clusters iteratively by using a small user-chosen local clustering threshold, d, allowing OTUs to reach their natural limits.(Mahé et al. 2014)

I.4.2.2.4. Filtering

Filtering data is a crucial step as it reduces the number of false positives and speeds up downstream analyses of the data(Falentin et al. 2019). Although many filters can be applied, it is generally considered good practice to use abundance-based filters as most very low-abundance OTUs are either induced by sequencing errors or by the presence of chimeras. A standard is to remove all OTUs whose relative abundance is inferior to 0.005% (Bokulich et al. 2013).

When biological replicates are available, filtering by keeping only those OTUs found in all replicates allows only data with demonstrated repeatability to be kept. Finally, it is also possible to tailor the filtering by using mock communities. In this case, it can be possible to filter until the mock community is accurately represented, as technical biases need to be removed in this operation.

I.4.2.2.5. Taxonomic affiliation

We can distinguish two main types of algorithms for taxonomy affiliation of OTUs. The first type use Basic Local Alignment Search Tools (BLAST). With this approach, the sequence is aligned against sequences present in a reference database. It allows for the precise identification of differences between two sequences and give an estimation of how close they are through metrics such as identity percentage, coverage percentage, e-value (probability that a match was obtained by chance), or a bit-score that combines information about identity and coverage. These methods also provides information about the reliability of the taxonomy given and allow for consensus of the n best matches to provide strong evidence of an assignation. However, BLAST is slow, which explains the development of other kind of algorithms.

Algorithms based on k-mer (motif composed by k nucleotide) detection such as RDP provide quicker assignation without performing an alignment. Instead, they rely on the fact that certain k-mers are specific to specific taxa. The k-mers are randomly extracted from the sequences and tested against the k-mer database. Although this bypass the long alignment process, it may miss chimeras that would be detected with BLAST and has less robust assignation.

In summary, a trade-off between accuracy and speed is always necessary when making taxonomic assignments using metabarcoding. However, it should be noted that differences in assignment may be less influenced by the algorithm than by the database used.

OTUs are then stored into abundance tables in BIOM format to be used in the downstream analyses.

I.4.2.2.6. Integrated tools

With time, the analysis of metabarcoding has become more mature and several integrated tools have been proposed to analyse datasets. This tools aims at transforming initial raw data into abundance tables through an integrated pipeline.

Mothur (Schloss et al. 2009) is an open source command line software that provides a variety of tools for each of the different steps of metabarcoding analysis and thus offers great versatility. For instance, chimera detection can be done by Vsearch, Uchime or Perseus. It can also provide statistical information on samples, such as rarefaction curves or diversity measures. However, the clustering step is limited by its OUT clustering algorithm.

Qiime (now Qiime2) (Bolyen et al. 2019) is a tool aggregator for metabarcoding analysis. It allows existing tools to match their inputs/outputs. Like Mothur, it offers a variety of options and most of the differences are only " cosmetic" (Schloss 2020); the relevant differences lie in the choice of clustering algorithms where Qiime offers several solutions instead of one very efficient one.

FROGs (Escudié et al. 2017) have advocated for a one-size-fit-all strategy by providing a dedicated tool for all biologists that does not require extensive knowledge in command line and where tools have been selected to offer optimal performance. The main differences between FROGs and the other tools mentioned above is the clustering step that is performed by SWARM (Mahé et al. 2015). FROGs also offers the choice of how taxonomic affiliation should be performed and which filters should be applied. In addition, a number of tools have been included to perform statistical analysis of the dataset. It has recently been expanded to allow taxonomic identification of fungi using the ITS.

DADA2 (Callahan et al. 2016) is an R package that, instead of clustering sequences into OTUs, follows a denoising strategy to create Amplicon Sequence Variants (ASV) that allow strains to be separated when necessary.

Thus, depending on the user, the dataset or the main objective of a project, the choice of an integrated tool can facilitate the analysis of metabarcoding datasets.

I.4.3. Whole metagenome sequencing

Shotgun metagenomics can be divided in 5 major steps. DNA extraction has been discussed previously, so this section will focus on the steps from quality assessment of the reads to statistical analysis.

I.4.3.1. Quality assessment of data

Shotgun metagenomics output is provided as a FASTA or FASTQ file. The quality assessment of the data is a very important step at the outset as it allows the detection of possible contamination and ensures that the bases are not too degenerated. IT tools have been developed, to filter out those poor quality reads. All these tools are necessary for quality control (QC).

One of the quality control tools is FastQC, which provides an overview of the overall quality in the form of an HTML report. Only high quality reads should be included in subsequent analyses. Although these tools can detect problems, no correction tools are included. Other tools include Fastx-Toolkit, NGS QC Toolkit, PRINSEQ or QC-Chain, which are quality control tools that include processing functions. One of the main correction needed is the removal of adapters, which can be done by Cutadapt, Trimmomatic, trimgalore and others.

I.4.3.2. (Co)assembly of data

When analysing metagenomic data, one must first decide whether or not an assembly step is necessary. The process of assembling and binning reads is time-consuming and computationally intensive, which for an application where only a taxonomic/functional profile of a community is required, may be perfectly usable. For instance, methods overlooking assembly are often found in medical studies on the gut microbiome(Zuo et al. 2020; Greenblum et al., 2012). They can be used to quickly profile a microbial community or identify known functions. However, this type of analyses lives or dies depending on the quality of the available databases.

Although it is possible to analyse the reads directly, most metagenomic analysis are performed on contigs. Specific features like K-mer can be used to find a taxonomy for each read, however, the information will be limited to a single read. Unfortunately, longer structures such as genes or operon often necessitate longer fragments and thus assembly(Tamames et al., 2019). Assembly is also necessary on poorly studied environments for which databases may lack crucial information. Metagenomic assemblers use two categories of algorithm: Overlap, Layout and Consensus assembly (OLC) and de Bruijn graphs. A schematic representation of both algorithms is shown in **Figure 19**.

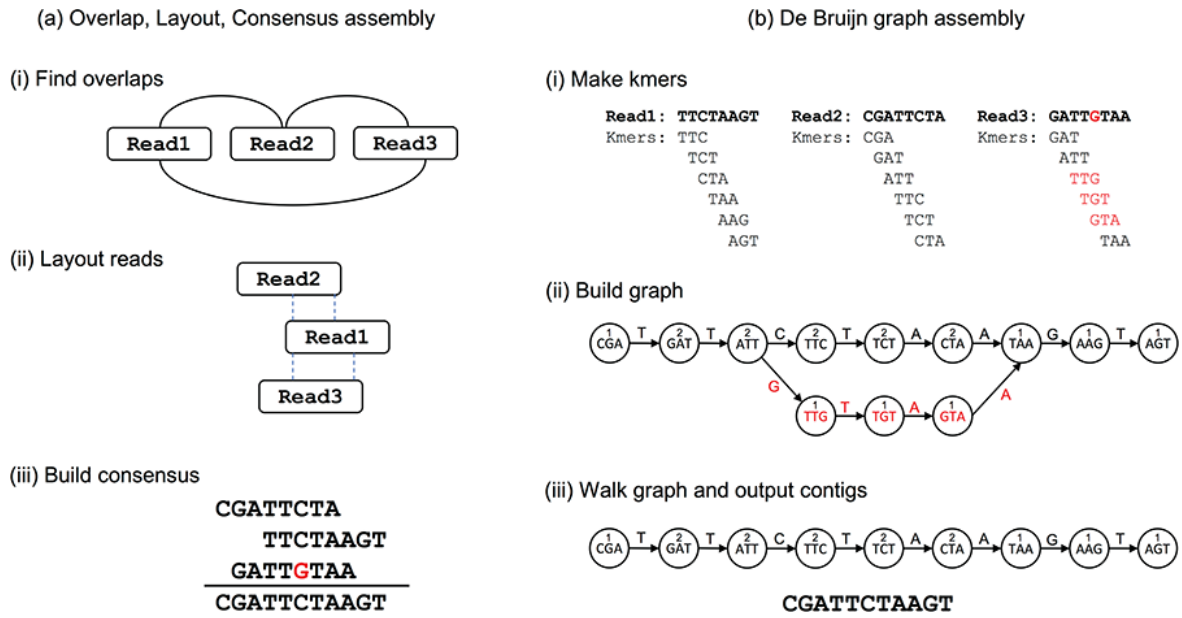


Figure 19: Two different approaches for assembly of short reads Ayling et al.,2019)

In OLC strategy (a), reads are laid out according to their overlapping sequences (ii), a probabilistic consensus is then done to build the more likely contigs (iii). This strategy has been used successfully to make contigs assemblers. There are some differences in the way both overlapping and consensus steps are done. For instance, GENOVO(Laserson et al., 2011) try to map reads into contigs in an iterative way. If no sufficient overlapping is found, a new contig is created. In order to prevent chimeric contigs, every 5 iterations the edges of the contigs are removed. If these reads were indeed part of the contig, they will find their place back later, if not they were chimeric. Other assembler using the OLC strategy have been developed; some examples include

BBlast (Lin et al. 2017), which is a blast-based overlapping assembler, IVA(Hunt et al. 2015) and SAVAGE (Baaijens et al. 2017), which were developed for viruses *de novo* reconstruction.

These strategies are often used by long reads sequencing assembler. Best performing long reads assemblers include software such as Flye, Miniasm/Minipolish, NextDenovo/NextPolish and Raven(Wick & Holt 2021; Latorre-Pérez et al. 2020). Assembly tools targeted at long reads are rapidly evolving and new tools are regularly added to the overall pool of usable software.

In de Bruijn graph assembly (b), the reads are decomposed into kmers which are short sequences of fixed size k. The selection window slides across the reads to make a complete kmer decomposition (i). The kmers are then mapped accordingly to their overlapping edges. Nodes are formed whenever two possible kmers branches occur (ii). The coverage of each kmer is calculated to determine which branch is more likely. The assumption used is that a contig should have the same kmer coverage from one end to the other (iii). The main problem of this assembly strategy is that it is highly dependent of the size of the kmer. In order to correct this drawback, popular assembler such as MEGAHIT(D. Li et al. 2015) have an iterative process with increasing kmer size computed to distinct de Bruijn graphs. In order to reduce computational time, MetaVelvet (Namiki et al. 2012) decomposes the whole de Bruijn graph into subgraphs and separate contigs from chimeric graphs. Most assemblers use this strategy. Examples are RayMeta (Boisvert et al. 2012), MetaSpades (Nurk et al. 2017) etc.

Given the large number of different assemblers, several benchmarks have been carried out to help scientists choose the most appropriate assembler. The most exhaustive one is The Critical Assessment of Metagenome Interpretation (CAMI) challenge which is a global comparative assessement of assembly and binning tools.(Sczyrba et al. 2017) Six assemblers (MEGAHIT, Minia, Meraga (Meraculous + MEGAHIT), A* (using the OperaMS Scaffolder), Ray Meta and Velour) were assessed for their ability to recover contigs from low and high complexity mock datasets. Assemblers that use multiple kmer sizes, such as Minia, Meraga and MEGAHIT, performed better than assemblers that use single size kmers. MEGAHIT

was the best of these overall. However, genomes with very highly coverage were difficult to recover for all assemblers. MEGAHIT failed to recover circular elements (>150X coverage), Minia cannot recover sequence with 80-200X coverage and MERAGA's recovery rate decrease with coverage over 100X.

Depending on the purpose of the study, the results of the benchmarking may differ, as shown by Vollmers et al.(Vollmers et al., 2017) The authors assessed the performance of MetaVelvet, IBDA-UD, MEGAHIT, MetaSpades, Ray Meta, SOAP-denovo2 and Omega on two real datasets from kelp biofilm and the Marburg forest soil. Overall, MetaSpades demonstrated the best assembly size statistics, especially for the kelp biofilm dataset. MetaSpades allowed for the recovery of the most of the diversity at the expense of microdiversity loss. MEGAHIT provides a good balance between genome recovery and cost-effectiveness, allowing equivalent statistics on the soil samples despite a smaller size on the kelp biofilm. However, it is a better tool to access to microdiversity. It is therefore advisable to use MetaSpades for low diversity dataset or for the recovery of very abundant genomes, and MEGAHIT for those that are very complex or where computing resources are limited. However, all benchmark results should be taken with caution, as the development of these software packages is constantly evolving as they are regularly updated, as can be seen from MEGAHIT/ChangeLog.(D. Li [2014] 2021) The best tools for a certain data sets at a given time may differ. While MEGAHIT v1.0.3 had problems with heavily covered genomes (>150 X), this problem has been corrected in the new version 1.1.3.(Meyer et al. 2021)

I.4.3.3. Read mapping and statistics with aligners

After the contigs have been assembled, all information about the samples is lost. In order to have data on the samples, it is necessary to map the reads on the newly obtained contigs. This step necessitate read mappers/aligners to map reads onto the contigs. Popular aligners include BWA-mem, Bowtie2 and Novoalign. Several benchmarks have been published in recent years focusing on speed and accuracy (Shang et al. 2014) or the percentage of mapped/mismapped reads.(Lee et al. 2018) BWA-mem is the most popular tool and is generally considered both fast and reliable,

as is Bowtie2 with slightly more mismatchs. Novoalign has been shown to be effective with short reads (<70bp). It is also the tool that leads to the fewest mismatches at the cost of fewer mapped reads.

These tools generate a BAM file, which includes quantitative information on the number of mapped reads on each contigs and the distribution of the contigs on the samples.

In addition to aligners, software is required to process the data. It is possible to use the utility set Samtools or Bedtools to transform a BAM file to SAM or to sort and index BAM files, calculate the coverage of reads on contigs etc. These statistics can then be used in downstream analyses.

I.4.3.4. Protein alignment tools

The alignment of proteins on a database is a very important step in metagenomic pipelines. Indeed, it is used to access both the functions and the taxonomy of gene predictions. In order to access this information, it is necessarily to use an aligner. The first tool developed was BLAST(Altschul et al. 1990), which uses a so-called seed-and-extend approach. First, BLAST tries to find hits that are small sequences of exact matches; these hits are then extended. If these extensions pass a given threshold, they are used as seeds for the alignment between the target and query. Alignments that are above a threshold are called High Scoring Pairs (HSPs). This process is rather time consuming on large databases, such as NCBI non-redundant database (NCBI-NR), which is around 30 GB in size.

BLAST may be powerful, allowing for a screening against very big databases. However, with the ever-growing size of both databases and data to analyse, new tools have been necessary. Such examples include DIAMOND(Buchfink, Xie, et Huson 2015), which is able to speed up the analysis up to 20 000 times. By being much faster than BLAST, DIAMOND was designed to be used for complete metagenome alignment that would be dauntingly slow with BLAST. Other tools such as SWORD(Vaser et al., 2016) were designed to be quicker than BLAST (up to 64 times) and maintaining a higher sensitivity than DIAMOND.

I.4.3.5. Gene and Protein annotation and affiliation

The objective of build contigs from reads is to access the genome level. It is then an important step in metagenomics pipelines to accurately predict genes from contigs. Popular tools include Prodigal, GenemarkS-2 and Glimmer3. The predicted genes can then be annotated by complementary tools.

Unlike protein prediction, protein annotation is performed against a database. "All inclusive" tools such as Prokka (Seemann 2014) use gene predictions performed by several tools (including Prodigal for coding sequences (CDS) and use it as an input to be search against different databases, in a hierarchical order. Proteins are first searched in an optional database provided by the user using Blast+ BlastP, then against Uniprot, then Refseq proteins using Blast+. If no protein annotation was found, a series of hidden Markov modules databases are checked (Pfam, TIGRFams) using HMMer3. A predicted CDS is labelled as hypothetical protein if no match is found at this point. Recently, DFAST(Tanizawa et al., 2018) has been proposed as another fast tool for genome annotation. This tool is based on the use of GHOSTX instead of BLAST.

It is also possible to search through protein databases directly using BLAST(+) BlastP or DIAMOND BlastP against NCBI database such as RefSeq or Uniprot database. Pfamscan is another tool that can be used to search through the Pfam database. It is also possible to annotate proteins against the KEGG orthology (KO) database using tools such as GhostKOALA or BlastKOALA.(Kanehisa et al. 2016)

Although generalist tools are used for most analyses, dedicated databases and tools exist for specific applications. For instance, the CAZy database(Cantarel et al. 2009) inventories all Carbohydrate-Active Enzymes. It works both as a database or conducting identification of genes coding for CAZymes and is a defining resource for CAZymes. Other tools are designed for retrieving CAZmes information from a dataset such as dbCAN(H. Zhang et al. 2018). This kind of specific tools provides access to curated information that help standardizing results.

While most proteins are annotated from assembled data, there is still a large quantity of non-assembled data on which protein annotation could be performed. With this aim, tools have been developed to annotate functions on short reads such as

MetaGeneHunt(Berlemont et al. 2020). Other strategies have also been proposed which combine a direct alignment of short reads using DIAMOND followed by taxonomic and functional annotation with MEGANIZER.(Bağcı et al., 2021)

I.4.3.6. Binning

To access gene and protein information from metagenomics data, NGS reads have been assembled into contigs using various tools. However, these tools were not designed to reconstruct entire genomes. To do so, a binning step is necessary. This consists of clustering the contigs together to reconstruct complete genomes, according to different metrics. These clusters, also called metagenome-assembled genomes (MAGs), represent the genomes of metagenomics species.

A huge number of binning tools have been developed in recent years. In the framework of the CAMI initiative(Sczyrba et al. 2017), five metagenomics binners were compared: MyCC(H.-H. Lin et Liao 2016), MaxBin 2.0(Y.-W. Wu, Simmons, et Singer 2016) ,MetaBAT(D. D. Kang et al. 2015), MetaWatt 3.5(Strous et al. 2012), CONCOCT.(Alneberg et al. 2014) Overall, MaxBin2.0 performed the best in all situations by recovering highly pure genomes (>90%) with the best completeness. All other software packages were able to reach genome purity above 80%, with variable completeness. All the proposed methods were able to efficiently reconstruct high quality genomes at the strain level. A comparative study using a specifically designed tool for binner benchmarking AMBER (Meyer et al. 2018) provided further insight into the performance of currently available tools. The best results were obtained by the newly released DAS_Tool 1.1(Sieber et al. 2018) and MetaBat2(Dongwan Kang et al. 2019). Although binners benchmark, like assemblers benchmarks, provide useful information, they only provide a snapshot of the environment at a certain time and must therefore be followed with caution. Furthermore, currently, the use of a single binner is being replaced by the generation of a consensus binning using tools such as DAS_Tool or MetWrap.

I.4.3.7. MAGs analysis

The overall quality of the MAGs obtained by binning can be improved by a refining step. Anvio (Eren et al. 2015) is an integrated software that provides a complete metagenomic workflow. Its special feature is that it offers an interactive function to manually refining bins. Binning_refiner (Song & Thomas 2017) is a tool that allows refining bins by using a combination of binning tools.

Recently, other tools have been proposed to further improve the binning process. Example of these are GraphBin2 (Mallawaarachchi et al., 2021) that allows for contigs to be in multiple bins to improve binning or Jorg that provide semi-automatic pipeline to circularize generated MAGs.(Lui et al., 2021)

Another aspect to take into consideration is the quality of the MAGs. For instance on its blog, Anvi'o team recommends to only add MAGs with >90% completion and <10% redundancy to databases. Other guidelines classify MAGs depending on their quality (high = > 90% completion <5% redundancy, medium>50% completion <10% redundancy)(Bowers et al. 2017). Recently it was suggested to identify high-quality MAGs based on tools such as MiGA or CheckM (>70% quality score) and with high coverage and N50 (>2000 bp) and low number of contigs (<500). (Meziti et al., s. d.)

Taxonomic affiliation of MAGs can be done based on the taxonomic affiliation of the contigs by using various threshold of reads of a specific affiliation or phylogenomic trees of marker genes(Dutilh et al. 2007). MAGs-sepcific taxonomic classifiers also exist(von Meijenfeldt et al. 2019).

I.4.3.8. Taxonomic profilers

MAGs corresponding to the quality criteria are then, taxonomically affiliated. However, this step can be performed on reads or genes for which numerous tools have been designed. Direct taxonomic affiliation of reads allows a very fast taxonomic affiliation of a metagenomics dataset. The most popular tools include Kraken, Kaiju(Menzel et al., 2016), MetaPhlan3 (Beghini et al. 2021). Despite their speed, these taxonomic classifiers are only able to classify a suboptimal amount of reads. The results are particularly poor in less studied ecosystem, such as desert soil (**Figure 20**).

Tools such as Kraken and Metaplan2, which were first designed for human microbiome, tends to affiliate much smaller amount of reads in environment such as bioreactors. (Menzel et al., 2016)

	MEM	Greedy	Kraken	Merged
Human vagina	58.6	63.3	45.5	66.4
Human saliva	56.1	61.8	37.4	64.6
Human gut	51.6	56.7	34.7	61.1
Cat gut	51.8	58.9	23.0	63.1
Lake	26.5	33.7	3.1	34.0
River plume	31.8	38.8	5.8	39.3
Baltic seawater	25.3	33.9	5.6	34.8
Desert soil	19.3	24.4	3.8	24.7
Bioreactor sediment	65.0	72.8	16.2	73.1
Bioreactor compost	50.3	58.5	15.1	59.5
Mean	43.6	50.3	19.0	52.1

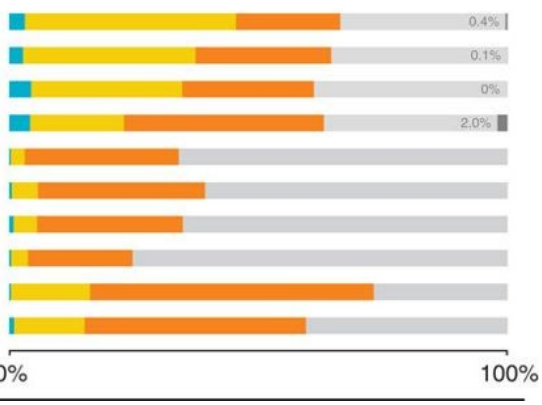


Figure 20: Comparison of the number of affiliated metagenomic reads using different tools. Percentage of classified reads in 10 real metagenomes for Kaiju MEM (*m*=12) and Greedy-5 (*s*=70), as well as Kraken (*k*=31). The Merged column shows the percentage of reads that were classified by at least one of Greedy-5 or Kraken. The Venn-Bar-diagram visualizes the percentage of reads that are classified either only by Kraken (blue), Greedy-5 (orange) or both (yellow).

I.4.4. A case study: metagenomics analyses of the termite gut

I.4.4.1. Microbial diversity of the termite gut

Metagenomics and metabarcoding approaches have been applied to the termite gut microbiome to understand its diversity. Metabarcoding, mainly based on 16S rRNA sequencing, has been used repeatedly to describe the diversity of the termite gut. These studies have allowed us to identify the core of the bacteria found in the termite gut and the differences between termite families or their diet. Recently, ITS have been used to decipher the fungal diversity of the termite gut. However, due to a lack of

representation in databases, identification remains very difficult(Vikram et al. 2021).

In lower termites, most of the diversity consists in flagellated protists that occupy up to 90% of hindgut volume (Hongoh 2010). These protists, mainly *Parabasalia* and *Preatoxysla* represent up to 20 species in the lower termite gut.

In higher termites, bacteria represents the main driving force although fungi and methanogenic archeae can also be found. Methanogentic archaeae are exclusively *Methanobrevibacter* in lower termites while other archeae such as *Methanosarcinales*, *Methanomicrobiales* and *Methanomassiliicoccales* can dwell in higher termites gut(Andreas Brune 2018).

Termites of the *Odontermes* family are known to be fungus-cultivating termites but anaerobic fungi can also be found in the gut of higher termite.(Vikram et al. 2021) Unfortunately, their diversity remains largely unexplored and only Eurotiomycetes and Malasseziomycetes were identified (**Figure 21, b**). The role of fungi in the termite gut is often ignored completely, though it is thought to be marginal.

In most termites, Spirochaetes are the main phylum (Treponema in particular; **Figure 21, a & c**). Bacteroidetes (mainly *Bacteroides*), Firmicutes (mainly *Clostridia*), Fibrobacteres and candidate cluster TG3 are also found in significant abundance. The only exceptions shown here are the two *Termes* species, who mainly feed on humus/soil instead of wood/grass that are the main diet of the other termites. The gut microbiome of *Termes* is comparatively richer in Bacteroidetes and Firmicutes instead of Spirichaetes. Overall, the main differences in the diversity of termite gut comes from their diet. Although the diversity of the gut microbiome of wood-feeding termites is quite similar, it differs greatly from that observed in substrates such as soil or humus with a high moisture content (Rossmassler et al. 2015).

Another important aspect of the measurement of 16S rRNA diversity is to understand its evolution when the gut microbiome is cultivated. It can then be monitored how much of a shift in diversity can be observed between the gut diversity and the diversity after 15 days of cultivation (**Figure 21, c**). Indeed, most Spirochaetes and Fibrobacteres were not be maintained in bioreactors and were replaced by

Baceroidetes, Firmicutes and Proteobacteria which corresponds to the typical diversity of anaerobic bioreactors rather than that of the termite gut(Auer et al. 2017).

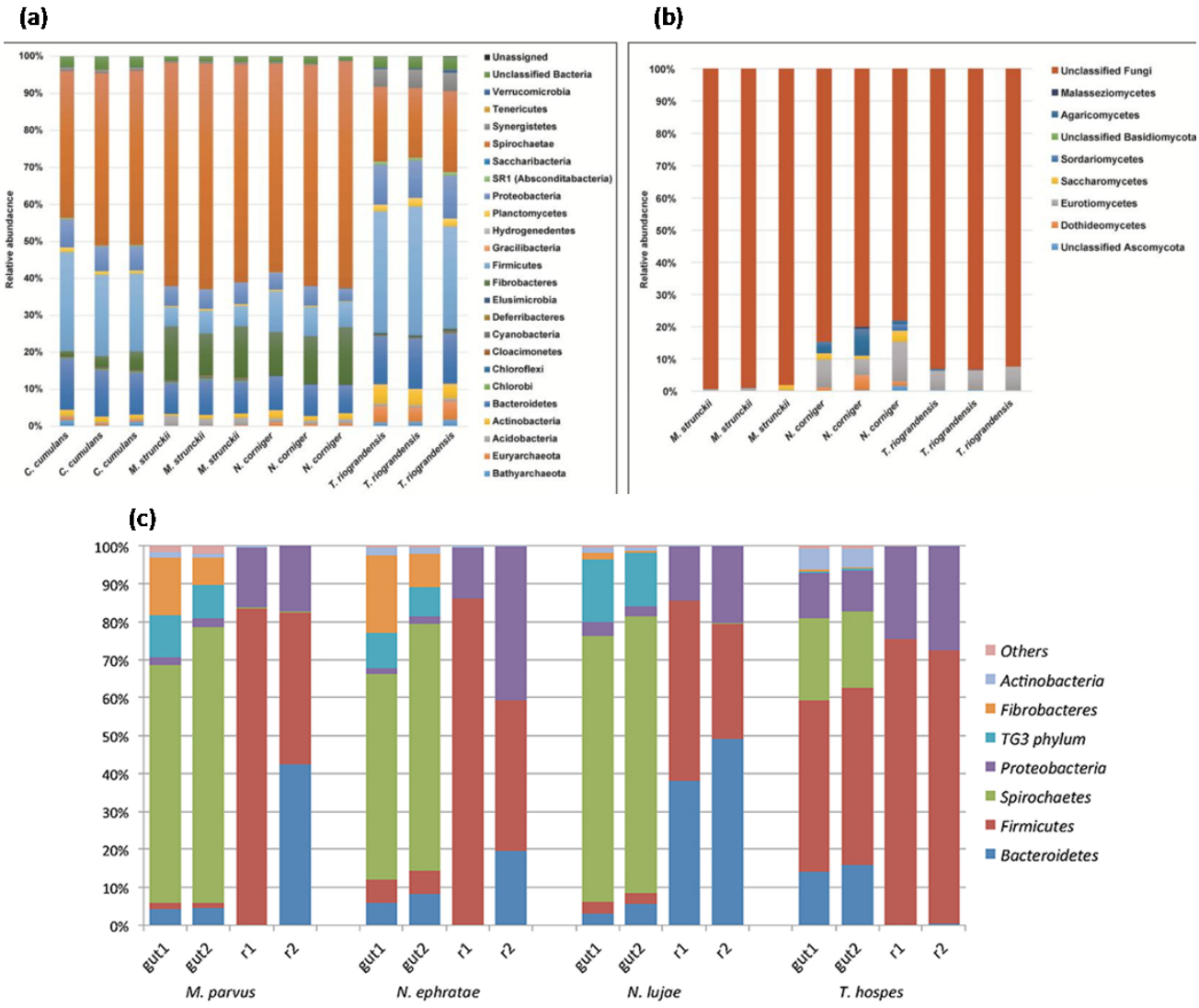


Figure 21: Bacterial (a) and fungal (b) diversity of the gut of four termites: *Coptotermescumulans*, *Microcerotermesstruckii*, *Nasutitermes corniger* and *Termesriograndesis.*(Vikram et al. 2021) (c) Bacterial diversity of the gut of four termite species :*M.parvus*, *N.ephratae, N.lujae*, *T.hospes*(Auer et al. 2017).

I.4.4.2. Metagenomics and metatranscriptomics analyses of the termite gut microbiota.

As termites and their microbiota have displayed a high potential towards lignocellulose degradation, it is natural that both metagenomics and metatranscriptomics surveys of the termite gut microbiota have been done. The first metagenomics analysis of a Nasutitermes termite showed the presence of over 700 GHs including GH5 (cellulases) demonstrating its high potential for lignocellulose degradation.(Warnecke et al. 2007) No lignolytic enzymes were observed. More recent analyses using metatranscriptomics were able to confirm the high efficiency of lignocellulose degrading enzymes from termites as well as attributing them putative origin.(Marynowska et al. 2020)By using these methods the microbiota from numerous termites were deciphered (Calusinska et al. 2020; Marynowska et al. 2020; Grieco et al. 2019).

Metagenomics and metranscriptomics analyses were also used to identify the contribution to the microbiota to its host defence(Peterson et Scharf 2016).

Supplementary Table 1: List of GH families and associated activities and function

FAMILY	MAIN_ACTIVITIES	MAIN_CAZYMES_FUNCTION
GH44	endoglucanase/xyloglucanase	Cellulases
GH48	reducing end-acting cellobiohydrolase	Cellulases
GH5	Cellulases	Cellulases
GH9	endo-glucanases	Cellulases
GH102	lytic transglycosylases	Chitin/Peptidoglycan degrading enzymes
GH103	peptidoglycan lytic transglycosylase	Chitin/Peptidoglycan degrading enzymes
GH104	peptidoglycan lytic transglycosylase	Chitin/Peptidoglycan degrading enzymes
GH108	N-acetylmuramidase	Chitin/Peptidoglycan degrading enzymes
GH109	*N*-acetylgalactosaminidase	Chitin/Peptidoglycan degrading enzymes
GH123	*N*-acetyl-β-galactosaminidases	Chitin/Peptidoglycan degrading enzymes
GH129	α-N-acetylgalactosaminidase	Chitin/Peptidoglycan degrading enzymes
GH136	lacto-N-biosidase	Chitin/Peptidoglycan degrading enzymes
GH154	β-glucuronidase	Chitin/Peptidoglycan degrading enzymes
GH156	exo-α-sialidase	Chitin/Peptidoglycan degrading enzymes
GH163	endo-β-N-acetylglucosaminidase	Chitin/Peptidoglycan degrading enzymes

GH18	chitinases /endo-β-N-acetylglucosaminidases	Chitin/Peptidoglycan degrading enzymes
GH20	*β-N*-acetylglucosaminidases	Chitin/Peptidoglycan degrading enzymes
GH33	sialidase or neuraminidase	Chitin/Peptidoglycan degrading enzymes
GH73	Muramidases	Chitin/Peptidoglycan degrading enzymes
GH84	N-acetyl β-glucosaminidase	Chitin/Peptidoglycan degrading enzymes
GH85	endo-β-N-acetylglucosaminidase	Chitin/Peptidoglycan degrading enzymes
GH89	α-N-acetylglucosaminidases	Chitin/Peptidoglycan degrading enzymes
GH128	β-1,3-glucanase	Glucanases
GH144	endo-β-1,2-glucanase	Glucanases
GH16	β-glucanase / β-glucan synthetase	Glucanases
GH50	β-agarase	Glucanases
GH55	exo-β-1,3-glucanase	Glucanases
GH63	processing α-glucosidase	Glucanases
GH64	β-1,3-glucanase	Glucanases
GH66	Dextranases	Glucanases
GH81	endo-β-1,3-glucanase	Glucanases
GH87	mycodextranase	Glucanases
GH10	Endohemicellulases	Hemicellulases
GH11	endo-β-1,4-xylanase	Hemicellulases
GH141	α-L-fucosidase/xylanase	Hemicellulases
GH26	Endohemicellulases	Hemicellulases
GH53	β-1,4-galactanase	Hemicellulases
GH74	xyloglucanase/endoglucanase	Hemicellulases
GH8	Cellulase/chitosanase/endo-1,4-β-xylanase	Hemicellulases
GH93	exo-α-L-1,5-arabinanase	Hemicellulases
GH95	1,2-α-L-fucosidases	Hemicellulases
GH19	Lysozymes/chitinase	Lysozymes
GH22	lysozymes/α-lactalbumin	Lysozymes
GH23	Peptidoglycan lyases/lysozymes	Lysozymes
GH24	Lysozymes	Lysozymes
GH25	Lysozymes	Lysozymes
GH127	β-L-arabinofuranosidase	Oligosaccharides debranching enzymes
GH137	β-L-arabinofuranosidase	Oligosaccharides debranching enzymes
GH142	β-L-arabinofuranosidase	Oligosaccharides debranching enzymes
GH146	β-arabinofuranosidase	Oligosaccharides debranching enzymes
GH153	poly-β-1,6-D-glucosamine hydrolase	Oligosaccharides debranching enzymes
GH159	β-D-galactofuranosidase	Oligosaccharides debranching enzymes

GH51	α-L-arabinofuranosidase	Oligosaccharides debranching enzymes
GH54	α-L-arabinofuranosidase/β-xylosidase	Oligosaccharides debranching enzymes
GH62	α-L-arabinofuranosidase	Oligosaccharides debranching enzymes
GH67	α-glucuronidase/xylan α-1,2-glucuronidase	Oligosaccharides debranching enzymes
GH78	β-glucuronidase/α-L-rhamnosidases	Oligosaccharides debranching enzymes
GH1	β-glucosidases/β-galactosidases	Oligosaccharides degrading enzymes
GH100	alkaline and neutral invertase	Oligosaccharides degrading enzymes
GH105	unsaturated rhamnogalacturonyl hydrolase	Oligosaccharides degrading enzymes
GH110	α-galactosidase	Oligosaccharides degrading enzymes
GH115	xylan α-1,2-glucuronidase	Oligosaccharides degrading enzymes
GH116	β-glucosidase/β-glucosylceramidase	Oligosaccharides degrading enzymes
GH117	α-1,3-L-neoagarooligosaccharide hydrolase	Oligosaccharides degrading enzymes
GH120	β-xylosidase	Oligosaccharides degrading enzymes
GH125	exo-α-1,6-mannosidase	Oligosaccharides degrading enzymes
GH13	α-glucosidases/α-galactosidases	Oligosaccharides degrading enzymes
GH147	β-galactosidase	Oligosaccharides degrading enzymes
GH151	α-L-fucosidase	Oligosaccharides degrading enzymes
GH164	β-mannosidase	Oligosaccharides degrading enzymes
GH165	β-galactosidase	Oligosaccharides degrading enzymes
GH17	glucan endo-1,3-β-glucosidase	Oligosaccharides degrading enzymes
GH2	β-galactosidase / β-glucuronidase	Oligosaccharides degrading enzymes
GH27	α-galactosidase/α-N-acetylgalactosaminidase	Oligosaccharides degrading enzymes
GH29	α-L-fucosidase	Oligosaccharides degrading enzymes
GH3	β-glucosidase/β-xylosidase	Oligosaccharides degrading enzymes
GH30	β-glucosidase/β-xylosidase	Oligosaccharides degrading enzymes
GH31	α-glucosidases/α-galactosidases	Oligosaccharides degrading enzymes
GH32	Invertases/β-fructosidases	Oligosaccharides degrading enzymes
GH35	β-galactosidases	Oligosaccharides degrading enzymes
GH36	α-galactosidase/α-*N*-acetylgalactosaminidase	Oligosaccharides degrading enzymes
GH38	α-mannosidases	Oligosaccharides degrading enzymes
GH39	β-xylosidase/α-L-iduronidase	Oligosaccharides degrading enzymes
GH42	β-galactosidase	Oligosaccharides degrading enzymes
GH43	α-arabinofuranosidae/α-arabinase/β-xylosidase	Oligosaccharides degrading enzymes

GH52	β-xylosidase	Oligosaccharides degrading enzymes
GH57	α-amylases	Oligosaccharides degrading enzymes
GH59	β-galactosidase	Oligosaccharides degrading enzymes
GH68	levansucrase	Oligosaccharides degrading enzymes
GH76	α-1,6-mannanase / α-glucosidase	Oligosaccharides degrading enzymes
GH79	β-glucuronidase	Oligosaccharides degrading enzymes
GH91	Inulin lyase	Oligosaccharides degrading enzymes
GH92	α-1,2-mannosidase	Oligosaccharides degrading enzymes
GH97	α-glucosidases/α-galactosidases	Oligosaccharides degrading enzymes
GH98	endo-β-galactosidases	Oligosaccharides degrading enzymes
GH99	glycoprotein endo-α-1,2-mannosidase	Oligosaccharides degrading enzymes
GH106	α-L-rhamnosidase	Pectin degrading enzymes
GH138	rhamnogalacturonan α-1,2-galacturonohydrolase	Pectin degrading enzymes
GH139	α-2-*O*-methyl-L-fucosidase	Pectin degrading enzymes
GH140	β-1,2-apiosidase	Pectin degrading enzymes
GH143	2-keto-3-deoxy-D-lyxo-heptulosaric hydrolase	Pectin degrading enzymes
GH145	L-Rhα-α-1,4-GlcA α-L-rhamnohydrolase	Pectin degrading enzymes
GH28	Polygalagturonase	Pectin degrading enzymes
GH88	d-4,5-unsaturated β-glucuronyl hydrolase	Pectin degrading enzymes
GH112	lacto(galacto)-N-biose phosphorylase	Phosphorylases
GH130	Phosphorylases	Phosphorylases
GH149	β-1,3-glucan phosphorylase	Phosphorylases
GH161	β-1,3-glucan phosphorylase	Phosphorylases
GH65	Phosphorylases	Phosphorylases
GH94	cellobiose phosphorylase	Phosphorylases
GH126	α-amylase	Starch degrading enzymes
GH133	amylo-α-1,6-glucosidase	Starch degrading enzymes
GH148	β-1,3-glucanase	Starch degrading enzymes
GH15	glucoamylase/glucodextranase	Starch degrading enzymes
GH158	endo-β-1,3-glucanase	Starch degrading enzymes
GH37	α,α-trehalase	Starch degrading enzymes
GH4	α-galactosidases6-phospho-α-glucosidases	Starch degrading enzymes
GH77	4-α-glucanotransferases	Starch degrading enzymes

II.

EXPERIMENTAL PROCEDURES

II.1. SELECTION OF THE SOURCE OF INOCULA

II.1.1. Choice of the termite species

As discussed in chapter II, lignin degradation in termite is still not a resolved issue. Thus, data on undoubtedly lignin degrading termites is not available to take a decision based on literature data.

In the other hand, to implement a termite-gut microbiome in bioreactors the termite gut had to fill certain criteria:

Firstly the termites had to be available at the Institut de recherche et de développement (IRD), Bondy, France which is one of the very few laboratory accredited to breed termites.

Secondly, some properties of the termite digestive track had to be taken into consideration. Indeed, termites can be divided into higher and lower termites. Lower termites harbour both protists and bacteria in their gut whereas higher termites only have bacteria and archaea. Lignocellulose digestion is mainly achieved by the microbial community hosted in the termite hindgut excepted the *Macroterminae* genus which is a fungus-grower termite for which *Termitomyces* fungi are involved in lignocellulose deconstruction before its ingestion by the termite (Hongoh 2011). To successfully implement a bioreactor, the gut microbiome should preferentially contain only bacteria and no protists (difficult to grow in bioreactors) or fungi (which is not directly part of the termite gut microbiome). It means that only higher termites except *Macrotermitinae* would be suitable for implementing bioreactors. These constraints excluded the vast majority of termite species. At last, the termites had to feed on a lignin rich material which included xylophages and humus feeding termites.

Considering all these constraints, two termite species were chosen: *Nasutitermes ephratae* and *Microcerotermes parvus* which are known to display a good activity toward lignocellulose (Auer et al. 2017) and whose gut have already been successfully cultivated in the lab. Since only a little number of species was available, it was decided that the lignin degradation screening would be done on the

two termite species but using three different types of raw materials and three technical lignins kindly provided by Zelcor partners.

II.1.2. Collection of the termite guts

The termite guts from two species: *Nasutitermes ephratae* and *Microcerotermes parvus* were collected at IRD, Bondy in January 2018. Both species came from colonies bred in a controlled room in tropical conditions of temperature and humidity and fed with birch wood. Termite workers were collected from their nest and cold anaesthetised. The whole gut dissection consisted in a removal of the gut by tweezers and forceps. The dissection was performed in sterile conditions. The guts were immediately suspended in a physiological solution after removal and pooled in a 50 ml centrifugation tube. Once 100 guts were collected, the tube was frozen at -80°C. A total of 1800 guts of *Nasutitermes ephratae* workers and 800 guts of *Microcerotermes parvus* workers were collected in 100 gut lots. Additional 60 guts in three lots were collected for microbial biomass quantification (qPCR). An additional 1000 Nasutitermes guts were sent by the IRD later on.

II.2. THE SUBSTRATES

II.2.1. Selection of relevant natural substrates

To ensure a wider screening, chosen substrate had to differ in their lignocellulosic composition while being promising source for biorefinery and linked to the industrial lignins provided the first workpackage of the project.

First of all, a lignin from Futurol (FU01), extracted from wheat straw, was received within the frame of Zelcor; thus wheat straw was chosen as the first substrate. This choice was also motivated by wheat straw being a widely available feedstock in France and Europe. 450 Mtons of wheat straw could be available globally each year(Talebnia et al., 2010).

An efficient screening requires more than one substrate. The potential substrates had to meet the following criteria:

1- The chosen feedstocks should be widely available in Europe now or in a near future. Feedstocks not grown in Europe but already purchased at a broad scale can also be considered.
2- The chosen feedstocks should ideally have a high lignin content.
3- The chosen feedstocks should be already used (or expected to be used in a near future) in biorefinery/paper industry/energy production.
4- The chosen feedstocks should have a high yield and should be grown in a sustainable way (in accordance to the European research roadmap for biorefinery.

In order to make fair comparisons, it was decided to do a pre-treatment on the wheat straw so its lignin content change.

Selected feedstocks

Wheat straw from the winter wheat variety Koreli was collected at an experimental farm (INRA, Boissy-le-Repos, France) in August 2011. After harvesting, the straw was milled to 2 mm and stored at room temperature (20–25°C) (Auer et al., 2017). The dryness content was determined to be higher than 95%.

II.2.2. Selection of the relevant conditions for the substrate usage

Using these substrates as raw biomass (without pre-treatment) for a screening to select lignin-degrading microorganisms would probably result on a preferential selection of strong cellulose/hemicelluloses degraders since studies have already shown a convergence of microbial community by the substrate (Wong et al. 2016). Conversely, a selection on purified lignin may lead to a decimation of the bacterial diversity. For instance, *Coptotermes formosanus* was not able to use lignin as the sole carbon source (Tarmadi et al. 2017) but it can effectively modify lignin in presence of cellulose/hemicelluloses(Tarmadi et al. 2018). Therefore, it was chosen to use a polysaccharide exhausted biomass obtained by enzymatic pre-treatment.

II.2.3. Enzymatic pre-treatment of the raw biomass and LWS production

Cellic Ctec2® (Novozymes) was purchased from Sigma. Its activity was test against carboxymethylcellulose (CMC) and xylan. The activity was determined to be 1076 U/ml CMCase. A selection of xylanolytic cocktails were tested in addition to Cellic Ctec2®. The best polyssacharide degradation rate was obtained with 40 U Cellic Ctec2®/gWS +96 U **Pentopan Mono BG® (purchased from Merck). The pretreatment was performed as follows.**

In a Sartorius 2l reactor, 100 g of wheat straw were added. 1.2l of pH 4.8 Na-Acetate 0.1M buffer, prepared by mixing 1.2l of a 0.1M solution of sodium acetate and 1l of 0.1M acetic acid and adjusted to pH4.8 was added. The suspension was heated at 45°C and stirred at 300 rpm during 2 hours until the suspension is homogeneous. 200 ml of an enzymatic solution was prepared containing 3.7 ml of Cellic Ctec2® amounting to 4000U (40 U/gWS) and 3.95 g of **Pentopan Mono BG®** amounting to 9600 U (96U/g) in buffer. When the reactor was homogeneous this solution was added. Finally the reactor volume was adjusted to 2l by adding buffer. Once the reactor was ready stirring was slowed down to 250 rpm. After 48h the reactor was stopped and left to decantate overnight. The solid phase was withdrawn and dried for 48h at 55°C. The dried biomass was then pooled together.

Dried pre-treated wheat straw was then milled to 0.5 mm and pre-treated again. The second bash was washed with distilled water, dried for another 48h at 55°C and stocked. The result of this pre-treatment was labelled LWS (lignin-rich wheat straw) and used as substrate for most following applications.

II.2.4. Using technical lignin as a substrate

Three different technical lignins were received in the framework of the project. Their composition is presented in Table 2. Two fractionated lignins were also provided by the University of Wageningen (Majira et al. 2019) and a nitrated lignin was provided by the University of Warwick (Ahmad et al. 2010; Taylor et al. 2012)

Table 1: Composition of the different available technical lignins

Lignin	Original biomass	Process	Klason lignin (%)	%H %G %S (% total monolignols)	% glucose % hemicelluloses
Kraft (Indulin AT™)	Pine wood	Kraft	88.8	10.5 89.5 0	0.21 0.56
GV03	Mix wheat/sarkanda grass	Soda	89.4	1.2 50.1 48.7	0 2
FU01	Wheat straw	Procethol 2G	54.9	2.5 46.8 50.7	33,9 4,8

II.3. THE BIOREACTORS

II.3.1. Technical lignins degradation experiments

II.3.1.1. Experimental design for technical lignins degradation bioreactor experiments

The degradation of technical lignins was assessed using a homemade 400 ml bioreactor presented below (Figure 1)

The screening of the lignin degradation by a microbial community from termites was performed in a house-designed 400 ml bioreactor. Gut samples were thawed at 4°C, centrifuged at 7,197g, 10min, 4°C and the saline solution was eliminated. 100 guts were used to inoculate 300 ml of mineral medium.

The mineral medium consisted in (per litre) 3.21g KH_2PO_4, 1.57 g K_2HPO_4, 0.4g de NH_4Cl, 0.5g NaCl, 0.16g $MgCL_2.6H2O$, 0.09g $CaCl_2.2H2O$, 250µL of V7 vitamin

solution and 1ml of trace micronutrient solution containing (per litre): 300 mg H_3BO_3, 1.1g $FeSO_4.7H_2O$, 190mg $CoCL_2.6H_2O$, 50 mg $MnCL_2.4H_2O$, 42 mg $ZnCL_2$, 24 mg NiCL2.6H2O, 18 mg $NaMoO_4.2H_2O$ and 2 mg $CuCL_2.2H_2O$. The mineral medium was filter sterilized with a 0.2µm Steritop device.

The bioreactors were autoclaved (121°C, 20 min, 1.1 bar) before the medium was added. 600 mg of technical lignin were then added to achieve a 2 g/l concentration.

The temperature was maintained at 30°C by water circulation and the pH was controlled to pH 7 by addition of 2M NaOH and 3M HCl. Stirring was fixed to 300 rpm. Aerobic conditions were ensured by the continuous injection of 150ml filtered air/min. These conditions were chosen to stay close from termite gut real conditions (Brune, 2014). Each condition was conducted in duplicate and compared to a control reactor without inoculation.

10ml samples were taken after 0, 2, 4, 7, 14, 21 and 28 days of incubation. These samples were used to analyze lignin degradation by COD measurement, Folin method and microscopy. Complementary analysis were performed by INRA Versailles.

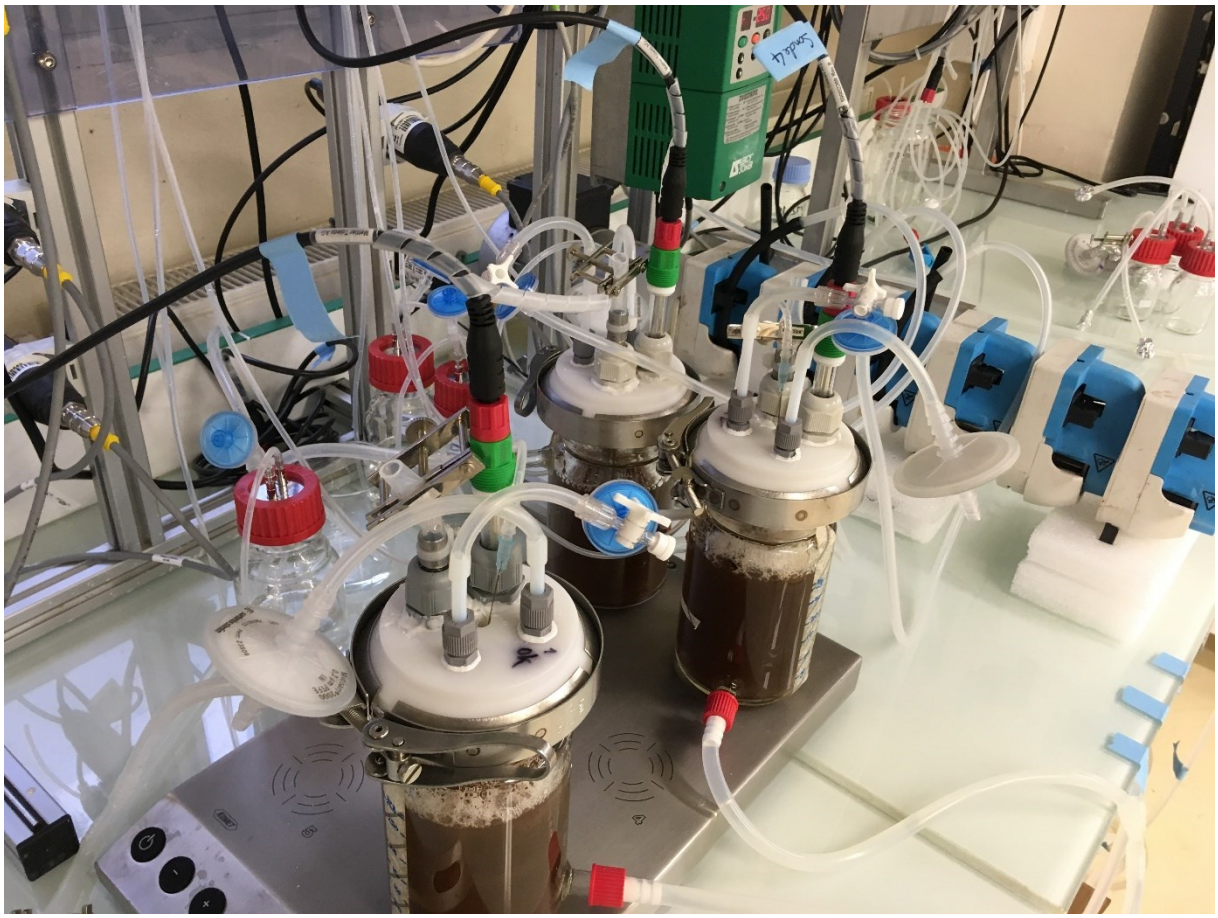


Figure 1: Photograph of the homemade bioreactors for technical lignin degradation

II.3.1.2. Analysis

II.3.1.2.1. Chemical oxygen demand (COD) quantification

1.25 ml samples were diluted with distilled water until reaching 5m. 2 ml of this solution were added to COD test kits HI 93754B-25 (medium range: 0-1500 mg/L) (HANNA instruments). The test tubes were heated for 2h at 150°C. COD was measured at 620 nm on a (Portable datalogging spectrophotometer DR/2010 (Hach); program 435 COD HR).

II.3.1.2.2. Lignin quantification by Folin method

Lignin concentration determination by the Folin method is based on the reduction of Folin-Ciolcateu reactant reduction by the phenol group in the lignin (Standard Methods Committee, 2010 .

1ml samples were diluted to 5 ml in distilled water.

50µL of 2X Folin reactant then 1mL tartrate-carbonate solution were added in rapid succession to 5ml diluted samples. After 30min incubation at 30°C in the dark, the absorbance at 700nm was measured on a spectrophotometer ((Hach LANGE GmbH) DR 1900). Lignin concentration was deduced from lignin standards obtained in the same conditions.

II.3.1.2.3. Microscopy

Drop of samples were put between a slide and a slip cover and used to observe the microbial community with a microscope (Nikon Eclipse 50i). Samples were observed with optical magnification from X100 to X400). Microbial activity was assessed by non-Brownian movement of the microbial species.

II.3.2. Termite gut consortia selection and enrichment in aerobic conditions

II.3.2.1. Experimental arrangements for termite gut consortia selection and enrichment in aerobic conditions

The selection and enrichment process of the termite gut consortia was conducted in 250 ml baffled flasks. The substrate used in this experiment were either wheat straw (WS) or lignin-enriched wheat straw (LWS). The flasks containing 2.2 g of substrate (20g/l) were autoclaved (20 min, 121°C, 1.1 bar) prior to adding the mineral medium (MM). In the case of WS, the flasks were autoclaved twice one day apart to ensure a good sterilization. 100 ml MM was added to the substrate.

Gut samples originating from either *Nasutitermes ephratae* (NE) or Microcerotermes *parvus* (MP) were thawed at 4°C, centrifuged at 7,197g, 10min, 4°C and the saline solution was eliminated and replaced by the same volume of mineral medium. The gut suspension was mixed and 10 ml (50 guts) were used to inoculate 100 ml of mineral medium (composition described at 2.3.1.1.).

Each culture was conducted in triplicate for 14 or 28 days at 30°C, 120 rpm and compared to a control flask containing the sterilized substrate and MM but not inoculated. After 14 (28) days, 10 ml of the flask suspension was used to inoculate a

new flask containing the substrate and the MM. The remaining volume was used to characterize the substrate degradation. This process was done over 5 cycles of successive batch cultivation. The sixth cycle was used to generate bacterial biomass. The three triplicates for each conditions after the sixth cycle were pooled to generate the inoculum used to inoculate 2L bioreactors.

The global experiment plan is presented in Figure 2.

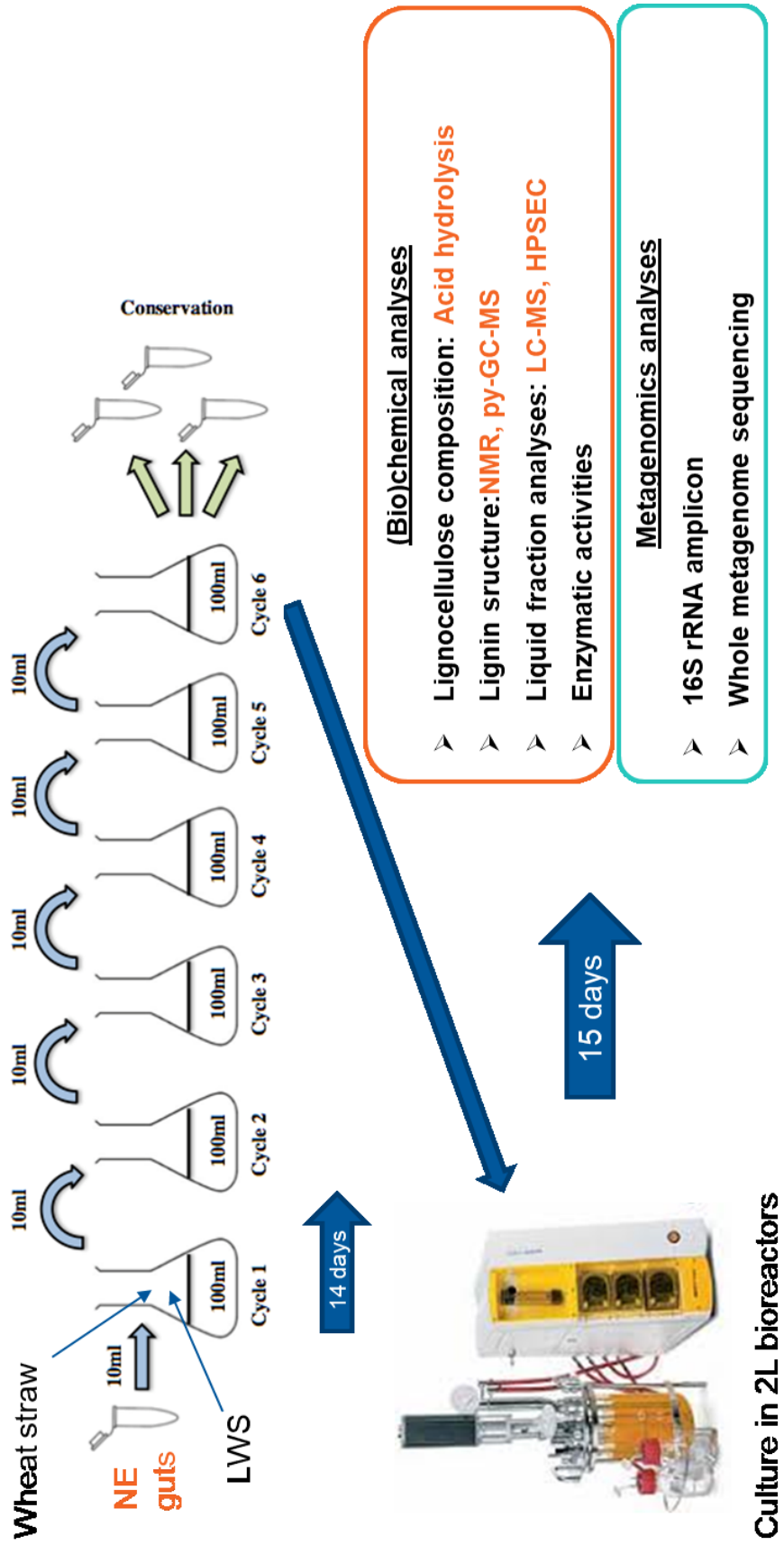


Figure 2: Overall design process

II.3.2.2. Analyses

II.3.2.2.1. Volatile solid (VS) determination

After 7 (14) and 14 (28) days, 15 mL were taken from the flask while maintaining sterile conditions and dried for 48h at 55°C. The dried matter was weighted to determine volatile solid content (VS).

II.3.2.2.2. Acid hydrolysis and Klason lignin determination

Wheat straw composition was determined using the sulfuric acid hydrolysis method (Lazuka et al. 2015; de Souza et al. 2013) on triplicate 80mg samples. The cooled reaction mixture was quantitatively vacuum-filtered through a preweighed 30 mL porosity 3 filtering Pyrex crucible (Fischer Scientific, Illkirch, France) fitted with a type GF/A glass microfiber filter (WhatmanFisher, Illkirch, France). The insoluble residue was washed free of acid with water and dried at 105 °C overnight to determine Klason lignin content. The soluble fraction was filtered and monomeric sugar concentration was determined using the HPLC(Monlau et al. 2012) on an Ultimate 3000 Dionex separation system equipped with a BioRad Aminex HPX 87H affinity column and a refractive index detector (Thermo Scientific). The insoluble fraction was filtered on a sintered filter, washed, dried at 105°C for 24h and weighted to determine Klason lignin content.

II.3.2.2.3. FTIR analysis

Volatile solid was subjected to FTIR analysis. 10 measurement by sample were done on a Nicolet ATR FTIR (Fisher Scientific). Data were analysed with a custom R program.

II.3.3. LWS degradation kinetics of the selected consortium in aerobic conditions

II.3.3.1. Experimental arrangements LWS degradation kinetics of the selected consortium in aerobic conditions

The LWS degradation kinetics by the selected NE15 consortium was conducted in 2L monitored bioreactors BIOSTAT A+ (Sartorius). The substrate used in this experiment was lignin-enriched wheat straw (LWS). The bioreactors containing 1L of distilled water was autoclaved autoclaved (20 min, 121°C, 1.1 bar). After sterilization, water was replaced by 1L of MM (pH7). 40 g (20g/l) of sterilized LWS was then added. The bioreactors were then stirred at 300 rpm for 2h until the LWS suspension was homogeneous. The volume was finally adjusted to 2L with MM (pH7). The bioreactors were maintained at 30°C with a heating coat, stirred at 250 rpm and 0.5 VVM of filtered air was injected.

75ml of the selected consortium were thawed at 4°C overnight before being used as the source of inoculum.

The bioreactors were conducted in triplicate for 15 days. A reactor was also conducted with no inoculation as a control. Samples were taken after 0 (3h after inoculation), 2, 5, 8, 12 and 15 days in order to characterize the macro kinetics of the bioreactors.

The samples taken were used to characterize the substrate degradation by measuring VS content, cellulose, hemicellulose and lignin content by acid hydrolysis and FTIR. Lignin modifications were assessed by thioacidolysis, FTIR, solid state NMR and py-GC/MS. The evolution of various enzymatic activities thought to be involved in lignocellulose breakdown was also monitored (xylanase, CMCases, laccase, peroxidase, nitrated lignin activity). Finally, the dynamics of the bacterial communities were monitored by either 16S rDNA and 16S rRNA sequencing or whole genome sequencing.

II.3.3.2. LWS breakdown by the selected consortium: Chemical analyses

Volatile solid determination, acid hydrolysis and FTIR were performed as described in II.3.2.

II.3.3.2.1. Enzymatic assays

For all enzymatic assays, 0.1 M potassium phosphate buffer pH 7 prepared by adding 9.34g K2PO4 and 6.309g KH2PO4 in 1l of distilled H20 was used to assess enzymatic activity at the pH of cultivation.

Xylanase and **CMCase** activity measurements were performed as previously described (Lazuka et al., 2015). 10 ml samples were taken at different times from the reactor. Samples were centrifuged at 7197g, 4°C for 10 minutes. The supernatant was tested for extracellular enzymatic activity while the pellet underwent 4 cycles of sonication (20s sonication with a probe (20s) followed by 20s in ice) for cell-bound and intracellular enzyme activity. Xylanase and CMCase activity were measured using respectively 1% w/v beechwood xylan (VWR) and 1% w/v carboxymethyl cellulose (Sigma) as a substrate. Activities were estimated by the DNS method. The DNS solution was prepared by adding 10g dinitrosalycilic acid and 16g NaOH to 1l of distilled water. When solubilized, 300g $C_4H_4KNaO_6.4\ H_2O$ was added. 600 µl of DNS solution was added to 600µl of samples. The mix was then heated for exactly 5 min in a dry bath then put in ice. Absorbance at 570 nm was then measured on a 96 well microplate using a Teclan instrument. 1 UI was defined as the volume of enzyme necessary to produce 1µmol of reducing sugars from the substrate per minute.

Laccase, lignin peroxidase and manganese peroxidase activity tests were adapted from (Zhou et al., 2017) were estimated by measuring the oxidation of 2,2'-azino-bis(3-Ethylbenzothiazoline-6-sulphonic) acid (ABTS) at 405nm using $\varepsilon 405_{nm}$ ABTS=36800 M^{-1} cm^{-1}. 1IU was defined as the volume of enzyme necessary to oxidize 1µmol of ABTS in 1 minute.

Nitrated lignin assays were adapted from (Ahmad et al., 2011; Taylor et al., 2012)(Ahmad et al. 2010; Taylor et al. 2012) using GV03 nitrated lignin kindly provided by the university of Warwick. The initial lignin solution (0.76 $g.l^{-1}$ 5-dinitrovanillin, $\varepsilon 300_{nm}$ ABTS=3255 M^{-1} cm^{-1}) was diluted 100X in 0.1M phosphate

buffer (Ph7) to make a stock solution. In a 96 well microplate, 30 µl of bacterial extract was added to 160µl of nitrated lignin stock solution and 10µl H202 40mM (10µl buffer) were added to test peroxidase activity on nitrated lignin. The microplate was incubated for 20 min in a plate reader (Teclan instruments) and A420nm was monitored (1 measure/min)

II.3.3.2.2. Cell wall extraction and grinding

Extract-free cell wall residues (CWR) were obtained by sonication-solvent extraction according to a published protocol(Mansfield et al. 2012) . In a 50 ml centrifugation tube, 40ml of ultrapure water were added to 2g of wheat straw. The mix was sonicated for 20 min in a sonication bath (reference materiel) before centrifugation for 5min at 7197g, 20°. The solvent was removed and replaced by 80% Ethanol:water and the sonication/centrifufation cycle was repeated twice. The final solvent was 100% acetone. CWR was left drying at ambient T°C for 24h and 24h 55°C.

Wheat straw was then milled to powder in a ball mill (Retsch) by doing two successive 15s mill with a metal ball.

II.3.3.2.3. Thioacidolysis

Thiacidolysis of the samples was done according to a published protocol(Méchin et al. 2014). The thioacidolysis solution was prepared by adding 20ml ethanethiol and 5ml BF_3 etherate in 40 ml dioxane and adjusted to 200ml with dioxane. 10 mg of dried extractive-free samples were weighted in a 15 ml glass tube with Teflonlined screwcap. 10 ml of thioacidolysis solution (1ml/mg) and 100 µl of standard solution (1.25 mg/ml tetracosane ($C_{24}H_{50}$) in dichloromethane (DCM), 0.125 mg/ml heinecosane ($C_{21}H_{44}$) in DCM) were added. The mix was gently stirred for 4h at 100°C. In a 30 ml glass tube the mix was added to 7ml 0.4 M NaHCO3 to quench BF3 excess. 100µl 6M HCl was then added and 7ml of DCM were finally added. The tube was then mixed. The lower organic phase was carefully extracted with a Pasteur pipette and dried on anhydrous sodium sulfate. An aliquot of this dried solution (10 µL) was trimethylsilylated (TMS) with 50 µL of N,O-bis(trimethylsilyl)trifluoroacetamide and 5 µL of ACS-grade pyridine for 1 h at room temperature. The TMS sample was injected (1 µL) onto a Trace1300 GC/MS

(Thermoscientific) fitted with an autosampler, a splitless injector (280 °C), and a Triple Quadripole-Ion trap in electronic impact mode with a source at 220 °C, an interface at 280 °C, and a 50 to 650 m/z scanning range. The samples were analyzed on an Agilent DB-5 column (Agilent Technologies) operated in the temperature program mode (from 50 to 110 °C at +30 °C/min, then 110 to 320°C at +6 °C/min), with helium carrier gas at a 1.5 mL/min flow rate. The GC-MS determinations of the H, G, and S lignin-derived monomers were carried out on ion chromatograms obtained with Xcalibur™ software (Thermo scientifics) respectively reconstructed at m/z 239, 269, and 299, as compared to the internal standard hydrocarbon evaluated on the ion chromatogram reconstructed at m/z (57 + 71 + 85)(Ployet 2017)

II.3.3.2.4. Solid state 2D HSQC NMR

CWR samples were swelled in DMSO-d_6/pyridine-d_5 (4:1, v/v) for the gel-state NMR analysis as described previously.(Tarmadi et al. 2018) NMR spectra were acquired using the Avance III 800US system (Bruker Biospin, Billerica, MA, USA) equipped with a cryogenically cooled 5 mm TCI gradient probe (Bruker Biospin). 2D HSQC NMR experiments were carried out using the standard Bruker implementation ('hsqcetgpsp.3') with the parameters described in literature.(H. Kim et Ralph 2010; Mansfield et al. 2012) Data processing and analysis were performed using TopSpin 4.0 software (Bruker Biospin) as described previously(Tobimatsu et al. 2013; P. Y. Lam et al. 2017). The central DMSO-d_6 solvent peaks (δ_C/δ_H: 39.5/2.49 ppm) were used as internal references. For volume integration analysis of the aromatic contour signals, C_2–H_2 correlations from G, G′ and F, C2–H2/C6–H6 correlations from S, S′ and P, and C2′–H2′/C6′–H6′ correlations from T were manually integrated and S, S′, P and T integrals were logically halved. For the polysaccharide anomeric signals, C1–H1 correlations from Gl, X, X′, X″, X‴, A and U were integrated. For the lignin side-chain signals, Cα–Hα correlations from I, II, III, III′ and IV″, Cγ–Hγ correlations from IV‴ and Cγ–Hγ correlations from IV and IV′ were integrated, and III, III′, IV, and IV‴ integrals were logically halved. The obtained integration values were normalized on a $½S_{2/6} + ½S'_{2/6} + G_2 + G'_2 + ½H'_{2/6} = 100$ basis for each spectrum.

II.3.3.2.5. Quantitative py-GC-MS with 13C lignin as internal standard

Analytical pyrolysis coupled to gas chromatography with high-resolution mass spectrometric detection (Exactive Orbitrap, Thermo Scientific, Waltham, MA, USA) was performed as previously described(van Erven et al. 2019). To each CWR sample (~80 µg), 10 µL of a ^{13}C wheat straw lignin internal standard (IS) solution (1 $mg.mL^{-1}$ ethanol/chloroform 50:50 v/v) was added and dried prior to analysis. All samples were prepared and analyzed in triplicate. Lignin-derived pyrolysis products were monitored in full MS mode on the most abundant fragment per compound (both nonlabeled and uniformly ^{13}C labeled). Pyrograms were processed by TraceFinder 4.0 software. Lignin contents and relative abundances of lignin-derived pyrolysis products were calculated as described previously(van Erven et al. 2019).

II.3.3.3. Analyses of the supernatant extracts

II.3.3.3.1. Sample preparation

The supernatant (80mL) of the inoculated and control bioreactors were extracted with 20 mL ethyl acetate (EA) after pH adjustment to 3–4 (2M HCl aqueous solution). The aqueous layer was further extracted by 2x20 mL EA. The combined EA extracts were dried over $MgSO_4$ and dried under reduced pressure bellow 45 °C. The extracts were dissolved in THF and filtered (0.45 µm acrodisc GHP filter).

II.3.3.3.2. HPSEC analysis

HPSEC analysis of the ethyl acetate extracts was performed using two different columns: 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, mixed-C pore type) and 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, 100Å). THF (1 $mL.\ min^{-1}$) was used as eluent and detection was performed at 280 nm. All the chromatograms were normalized in time scale according to the peak of toluene, injected as standard. Only chromatograms obtained with the 100 Å are shown here since they were found more informative.

II.3.3.3.3. LC-MS analysis

For LC-MS analysis of the ethyl acetate extracts, aliquots of the extracts were evaporated to dryness and dissolved in acetonitrile and filtered (0.45µm, GHP

Acrodisc, Pall Gelman, Merck, Molsheim, France). Samples were injected into an UHPLC apparatus (Thermo Fisher Scientific) coupled to an electrospray ionization mass spectrometer (ESI-MS) and photodiode array (PDA) co-detection. UHPLC analysis was performed using a C18 column (2.7 µm, 50 mm x 2 mm I.D.mm; Highpurity, Thermo Fisher Scientific), a 12–95 vol.% aqueous acetonitrile, 1% HCOOH gradient (Millipore, ST Quentin-en-Yvelines, France) for 30 min and a 1 $mL.min^{-1}$ flow rate. Negative-ion ESI-MS spectra (120–2000 m/z) were acquired using a quadrupole–time-of-flight (Q-TOF) spectrometer (Impact II, Bruker, Leipzig, Germany) setting needle voltage at 4 kV and desolvating capillary temperature at 350°C.

II.4. MOLECULAR BIOLOGY TOOLS

II.4.1. DNA/RNA extraction and purification

7.5 ml were taken from the bioreactor, aliquoted in 2 ml tubes and centrifugated at 13000g, 4°C for 5 min. The supernatant was discarded, the sediment was frozen in liquid nitrogen and stocked at -80°C.

Coextraction of DNA/RNA was done with the PowerMicrobiome RNA Isolation (Qiagen) kit using the protocol proposed by constructor. Cell lysis was performed by ball milling with a Fastprep instrument (MP Biomedicals, 2X30s at 6 $m.s^{-1}$ with 30s in ice between each lysis step). DNA and RNA were then separated using the AllPrep DNA/RNA minikit (Qiagen) following the constructor protocol.

The quality of the purified DNA/RNA was checked by migration on a 0.8% Agarose gel revealed by Ethidium Bromide and by a measure of absorbance at 260 nm and 280 nm using a Nanodrop 1000 device (ThermoScientific).

DNA traces in the RNA samples were removed using a TURBO DNA-free kit (Ambion) following the constructor protocol.

II.4.2. Reverse transcription and PCR1

Microbial diversity of samples was studied by 16S rRNA gene sequencing, targeting the V3-V4 region.

After RNA cleaning (TURBO DNase kit, Invitrogen),, reverse transcription of the RNA was achieved using the M-MLV reverse transcriptase kit 5invitrogen) and random hexamers as primers, in accord with manufacturer's instructions.

Illumina sequencing of the V3-V4 region of 16S rRNA genes was performed after PCR amplification using the bacterial primers 343F and 784R, modified to add sequencing adaptors during a second PCR (343F = 5'-CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC GGR AGG CAG CAG-3' , 748R=5'-GGA GTT CAG ACG TGT GCT CTT CCG ATC TTA CCA GGG TAT CTA ATC CT-3')).

A second PCR was performed to add indexes to the sequences using the following primers (FP2 = 5'-AAT-GAT-ACG-GCG-ACC-ACC-GAG-ATC-TAC-ACT-CTT-TCC-CTA-CAC-GAC-3' and RP2 = 5'-CAA-GCA-GAA-GAC-GGC-ATA-CGA-GTA-index-GTG-ACT-GGA-GTT-CAG-ACG-TGT-3') (Auer et al. 2017).

II.5. METAGENOMICS DATA ANALYSIS

II.5.1. Generation of an abundance table of OTUs and their taxonomic affiliation with FROGS.

Library preparation was performed as previously detailed (Auer et al. 2017)and loaded on a MiSeq Illumina cartridge using reagent kit v3 to obtain paired-end 300 base pair (bp) reads. 16S rRNA gene sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France) and generated between 13,000 to 32,000 paired-reads per sample (20,500 on average) that were analyzed using the FROGS pipeline (Find Rapidly OTU with Galaxy Solution) implemented on a Galaxy instance (http://sigenae-workbench.toulouse.inra.fr/galaxy/)(Escudié et al. 2017). In summary, paired reads were merged using FLASH(Magoč & Salzberg 2011); primers and adapters were

removed with cutadapt and a de novo clustering was performed using SWARM(Mahé et al. 2015) with an aggregation distance d = 3 after denoising using d =1. Chimera were removed using VSEARCH (Rognes et al. 2016). Singletons were filtered out from the dataset and affiliation was done using Blast+ against the Silva database (release 132). In agreement with recommendations (Bokulich et al. 2013), in order to avoid false OTUs, an OTU abundance threshold of 0.005% of the total sequences was applied. OTU tables were then formatted in standard BIOM for data analyses. For diversity analysis, OTU tables, taxonomic affiliations and bioreactors metadata were imported into Phyloseq v1.26.1 to analyze diversity(McMurdie & Holmes 2013a) using R v3.5.3 (R core team, 2019).

II.5.2. Pipeline for "shotgun" metagenomics analysis of a microbial community

II.5.2.1. Libraries constructions

For shotgun metagenomics, DNA libraries for individual samples were prepared following Illumina HiSeq3000/NovaSeq instructions using True Seq DNA-PCR free kit v2 (Illumina, San Diego, CA, USA). Libraries were pooled in identical quantities and paired-end sequenced using one line of the Illumina NovaSeq platform to obtain 150 bp reads. Shotgun sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France).

II.5.2.2. Quality control

Shotgun sequencing with Hiseq3000/ NovaSeq resulted in 300-400 M paired reads in the dataset (range: 19-31 million per sample). All samples were submitted to quality control using FastQC. As all samples displayed a minimal Phred quality score of 20, no remaining adapter, no overrepresented sequence were found, the kmer profiles were normal and the desired length of 150bp they were kept for genome reconstruction.

II.5.2.3. Co-assembly

For genome reconstruction reads form all samples were used for *de novo* co-assembly using MEGAHIT v1.1.3 with default parameters (D. Li et al. 2016). Co-assembly produced resulted in 33,071 filtered contigs (size ranging from 200 to 1.24 Mbp, N50=5393) on the first dataset, 708764 contigs (size ranging from 200 to 1.7 Mbp, N50=16175, L50 = 12143) on the second dataset. On the first dataset, due to the inability to reconstruct *Bacteroides graminsolvens* genome, the assembly of a single sample was done using MetaSpades.

Assembly statistics were obtained with Assemblathon.pl. Paired-end reads were mapped onto the contigs using BWA-MEM v0.7.17 with default parameters (Heng Li 2013) and mapping statistics were compiled using the Samtools suite v1.8 (Heng Li et al. 2009). Contigs were filtered using minimum contig length of 1500 bp and minimal abundance of one RPKM (reads per kilobase per million mapped reads) on at least two different samples.

II.5.2.4. Metagenome assembled genome (MAG) reconstruction

For MAGs reconstruction, contigs were filtered using minimum contig length of 1500 bp and minimal abundance of one RPKM (reads per kilobase per million mapped reads) on at least two different samples. Filtered contigs were the used to generate a contigs database using Anvi'o v6.2 (v5.4 in the first datasaet) (Eren et al. 2015).

Contigs binning is the step on which the main difference in analysis between the two datasets can be found. On the first dataset, contigs were binned using CONCOCT followed by a manual curation of the bins on Anvi'o v5.4. On the second datasets, however, as available tools have improved, contigs were binned using 3 different binning algorithm, namely CONCOCT v1.0.0 (Alneberg et al. 2014), MetaBAT 2 v2.12.1 (Kang et al. 2019) and MaxBin 2 v2.2.7 (Wu et al. 2016). The three independent binnings were then aggregated using DAS Tool (Sieber et al. 2018) to obtain a set of metagenome assembled genomes (MAGs). Each MAG was manually curated using Anvi'o v6.2 interactive based on the GC content and the differential

coverage of the contigs in the MAG. SCG were used to assess the completeness and contamination of the resulting MAGs using Anvi'o v6.2.

We classified the resulting MAGs as high quality (>90% completeness, <5% redundancy), medium quality (≥50% completeness, <10% redundancy) and low quality (other MAGs) in agreement to proposed classification(Bowers et al. 2017). Only MAGs of medium or high quality were kept for analysis.

II.5.2.5. Taxonomic and phylogenomic analysis

Taxonomic affiliation of the MAGs was performed with CAT and BAT in Bin Annotation Tool v5.0.3 (BAT) mode (von Meijenfeldt et al. 2019) using –r=10 –f=0.3 as criteria.

A MAGs phylogenomic tree was constructed with Phylophlan3 using the Phylophlan database of proteins and parameters set to accurate/high diversity (Segata et al. 2013). The best consensus tree over 100 bootstraps was obtained from RAxML-ng v0.9.0 (Kozlov et al., 2019). The phylogenomic tree was drawn using the Interactive Tree Of Life (iTOL v5) (Letunic et Bork 2019).

II.5.2.6. Functional annotation of MAGs: CAZymes, COGs and pfam

Open reading frames (ORFs) were searched on the filtered contigs and structural genes were annotated from the assembly using Prodigal v2.6.3 run in a metagenomics mode (-p meta) (Hyatt et al., 2010). Functional annotation of predicted protein was done with DIAMOND v. 0.9.22 in mode BLASTP against NR database (Buchfink et al., 2015).

Clusters of Orthologous Genes (COGs) were identified on the predicted proteins using eggnog-mapper on eggNOG 5.0 database (Huerta-Cepas et al., 2019). COGs related with VFA-production were specifically extracted from the COGs dataset to be analyzed (Supplementary data 3).

SusC and SusD genes were identified by their COGs. **PULs** were predicted in the MAGs by the presence of a tandem SusC-SusD genes on a contig.

CAZyme annotations were performed on the identified ORfs(Delannoy-Bruno et al. 2021) . A custom script which aims at accommodating the modularity of CAZymes was used to compare sequences to the full-length sequences stored in the CAZy database using BlastP (version 2.3.0+). If sequences are aligned with a sequence already in the CAZy database with 100% coverage, >50% amino acid sequence identity and E-value ≤ 10−6, they are assigned to the same family or subfamily (or the same families if the found sequence contains more than one CAZy module). If not, the sequences have to undergo a similarity search that involves two steps. On the one hand, using BlastP, they are compared to a library of modules (instead of full sequences that can contains multiple modules). On the other hand, a HMMER3 search against a curated collection of hidden Markov models based on each of the CAZy module is performed. The script assigns a family when both BlastP E-value<10^{-4} and it aligns with >90% overlap.

Predicted proteins with a potential lignin degrading activity were identified using **pfam modules**. A set of 31 modules were selected as important for lignin degradation and aromatics metabolism (Moraes et al. 2018). Pfam were predicted using Hmmer v3.2.1on Pfam33.1 release. Only modules with E-values < 1 e10^{-5} were considered.

Count of COGs, CAZymes, pfam and PULs were estimated for each MAG.

II.5.2.7. MAGs temporal dynamics along LWS degradation

To estimate the abundance of the reconstructed MAGs at each sampling point, raw reads were queried against to a database constructed with the MAGs qualified as al least medium-quality using bwa-mem with default parameters. Unmapped reads and reads mapped to more than one location were removed with samtools view (-q =10). The relative abundance of each MAG for each sampling point was computed as the fraction of reads in each sample mapping to the respective MAG normalized on the size of that bin (fraction of reads per nucleotide in bin) and divided by the total number of reads in the corresponding sample.

II.5.2.8. Statistical analysis

Abundance of the functional traits of MAGs along incubation time were explored and statistical differences between different traits was assessed using the R package vegan.

Subsequent data analysis was performed on R v4.0.2. Principal component analysis (PCAs) and Partial Least Square Discriminant Analysis (PLSDAs) were done with the MixOmics v6.6.2(Rohart et al. 2017). Heatmaps were generated through pheatmap and bootstrap was performed using pvclust(Suzuki et Shimodaira 2006). Networks analyses were performed using NetCoMi(Peschel et al. 2020). All analysis were performed using the Tidyverse v1.3.0.

III. TERMITE GUT MICROBIOTA CONTRIBUTION TO WHEAT STRAW DELIGNIFICATION IN ANAEROBIC BIOREACTORS

III.1. TERMITE GUT MICROBIOTA CONTRIBUTION TO WHEAT STRAW DELIGNIFICATION IN ANAEROBIC BIOREACTORS

*Louison Dumond[a] , Pui Ying Lam[b], Gijs van Erven[c,d], Mirjam Kabel[c], Fabien Mounet[e], Jacqueline Grima-Pettenati[e], Yuki Tobimatsu[b], Guillermina Hernandez-Raquet[a]**

[a]Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.

[b]Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto, Japan.

[c]Wageningen University & Research, Laboratory of Food Chemistry, Bornse Weilanden 9, 6708 WG, Wageningen, The Netherlands.

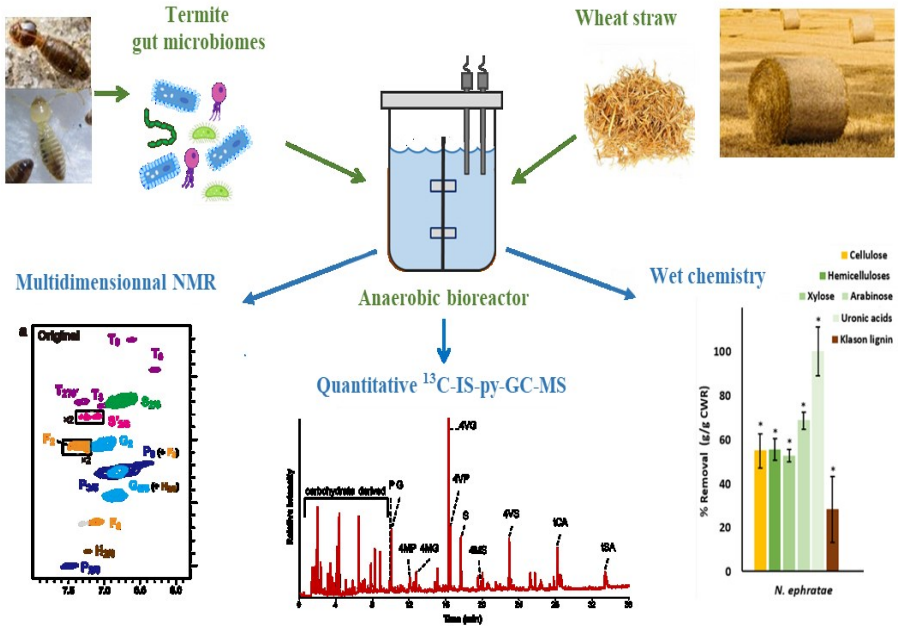
[d]Wageningen Food & Biobased Research, Bornse Weilanden 9, 6708 WG, Wageningen, The Netherlands

[e]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse III, CNRS, UPS, UMR 5546, 24 Chemin de Borde Rouge, 31320 Castanet-Tolosan, France

* Corresponding author

III.2. ABSTRACT

GRAPHICAL ABSTRACT



Lignin is a major lock for lignocellulose valorization in biorefineries, prompting a need to find new ligninolytic systems. Termites are efficient lignocellulose degraders and a large part of this ability comes from its anaerobic gut microbiome. However, the potential of termite gut microbiomes to degrade lignin under anaerobic conditions has yet to be elucidated. By applying wet chemistry, multidimensional NMR spectroscopy and quantitative ^{13}C-IS py-GC-MS, we determined the chemical and structural characteristics of wheat straw digested by the gut microbiomes of higher termites *Nasutitermes ephratae*, *N. lujae*, *Microcerotermes parvus* and *Termes hospes* implemented in anaerobic bioreactors. Interestingly, all gut microbiomes managed to remove lignin (up to 37%), although hemicellulose (mean 51%) and cellulose (mean 41%) were degraded more efficiently. Important structural differences, indicative of ligninolytic action, were discerned in the residual lignin. For the studied termite gut microbiomes, a slight decrease of S/G ratio was observed upon digestion, whereas lignin's interunit linkages and Cα-oxidized moieties, including benzaldehyde and hydroxypropiovanillone/syringone substructures, accumulated. Additionally, tricin

terminal units were clearly depleted, along with decreases in *p*-coumarate and ferulate pendant units. At least 80% of the observed delignification is estimated to be due to the removal of 'true' lignin, with the remainder being explained by the removal of hydroxycinnamic acids and tricin. Collectively, our findings suggest a partial deconstruction of lignin, with 'peripheral' lignin subunits being preferentially targeted. The present work thus provides new insights into the anaerobic deconstruction of lignin by termite gut microbiomes.

KEYWORDS: lignin; lignocellulose; termite gut microbiome; NMR spectroscopy; py-GC-MS; delignification; anaerobic digestion

III.3. INTRODUCTION

The increasing global demand for producing energy, transportation fuels and sustainable commodity chemicals from non-food renewable resources has motivated the development of lignocellulose biorefineries.(*Directive (EU) 2015/1513* 2015) However, lignocellulosic biomass is highly resistant against bioconversion that can largely be ascribed to the recalcitrance of the polymers it contains.

Lignocellulosic biomass consists mainly of the three polymers, cellulose, hemicelluloses and lignin, of which the former two are polysaccharides and the latter is an aromatic polymer. These polymers are elaborately interwoven with each other to form a complex three-dimensional structure, and constitute the majority of the lignocellulosic agricultural residues, accounting for 30-50% (cellulose), 20-30% (hemicelluloses) and 15-30% (lignin) of the dry matter, respectively(Lapierre et al., 1995; del Río et al. 2012). The composition of the biomass is completed by ash, proteins and wax.

Lignin plays a key role in plants as it provides mechanical strength and protection against environmental and microbial attacks(Liu et al., 2018). While lignin, therefore, forms the main recalcitrant factor against polysaccharide conversion in biorefinery approaches,(Kamimura et al. 2019) it is also the most abundant natural renewable source of aromatic compounds for chemical industries. These features have prompted a multitude of research into lignin degradation, with biological conversion strategies increasingly receiving attention(Kamimura et al. 2019).

Lignin is mainly synthesized from three monolignols, namely, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, which respectively give rise to the three major phenylpropane units, *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units, when incorporated into the lignin polymer(Ralph et al.,2019). In monocotyledonous grass species, including wheat, lignin additionally contains γ-*p*-coumaroylated G and S units derived from polymerization of γ-*p*-coumaroylated monolignols, and also tricin units derived from polymerization of a flavone tricin(Ralph et al.,2019). In addition, (di)ferulates (FA), mainly ester-linked to hemicelluloses (glucuronoarabinoxylans,

GAX), are partially bonded to lignin in grass cell walls(Ralph et al., 2019; Karlen et al. 2016) .

Despite lignin's recalcitrance to biodegradation, in nature, diverse microorganisms have shown to be able to degrade lignin, with white-rot like *Phanerochaete chrysosporium* and *Trametes versicolor* which have been applied for biotechnological purposes(Sigoillot et al. 2012). Compared to fungi, bacteria display lower ligninolytic activities but their versatile metabolisms enable them to degrade a large variety of lignin-derived aromatic compounds(Masai et al., 2007) Furthermore, bacterial enzymes are relatively easily produced, especially given the genetic tools for their heterologous overexpression,(Lambertz et al. 2014) therefore, several research teams are aiming to decipher bacterial lignin degradation.

Lignin degradation by bacteria was first suggested in Actinomycetes from the genus *Streptomyces*(Crawford 1980). Since then, the ligninolytic capacity of *Streptomyces* has been deeply characterized (Zeng et al. 2013; Riyadi et al. 2020)and members of *Pseudomonas*, *Rhodococcus*, *Bacillus,* and *Acinetobacter* have been described as potential ligninolytic bacterial species(Ahmad et al. 2010). Nevertheless, the lignin degradation capacity of bacteria associated to the partial or strict anaerobic digestive system of herbivorous animals, such as ruminants and phytophagous insects in which overcoming lignin recalcitrance might play an important role, still largely remains to be elucidated. These organisms, therefore, might harbor interesting bacterial species.

Of these phytophagous insects, termites are among the most efficient lignocellulose decomposers,(Andreas Brune 2014) with high potential for lignocellulose biorefinery applications. Termites are historically classified into higher and lower termites depending on their lignocellulose degradation mechanisms. In lower termites, most of the lignocellulose degradation comes from the interaction of the termite with flagellated protists as well as prokaryotes constituting their gut microbiome. While higher termites lack flagellated protists, they shelter a variety of lignocellulolytic prokaryotes within their gut. Some higher termites also recruit ectosymbiotic fungi for efficient lignocellulose degradation(Andreas Brune 2014).

While the degradation of the cell wall polysaccharides in termites has been extensively documented in the literature,(Andreas Brune 2014) it is still controversial

how termites overcome the lignin barrier. Initially, it was hypothesized that there was no lignin degradation in termites at all; the activity of termites being limited to the degradation of mono- or dimeric-lignin model compounds (Kato et al., 1998; Ke, Singh, et Chen 2011). With the advancement of tools for lignin analysis, especially the use of multi-dimensional nuclear magnetic resonance (NMR) techniques,(H. Kim et Ralph 2010; Mansfield et al. 2012) polymeric lignin degradation by termites has been recently reassessed(Tarmadi et al. 2018; Hongjie Li et al. 2017). For instance, a recent study which employed two-dimensional (2D) heteronuclear single quantum coherence (HSQC) NMR for dissecting lignocellulose decomposition by a lower termite *Coptotermes formosanus* provided NMR signatures supporting that lignin polymers can be at least partially decomposed during their passage through the termite gut digestive system(Tarmadi et al. 2018). Similarly, studies employing pyrolysis coupled to gas chromatography with mass spectrometric detection (py-GC-MS) reported structural modifications of lignin albeit with no clear depolymerization of lignin polymer backbones by lower termites *Cryptotermes brevis* and *C. formosanus*(Katsumata et al. 2007; Ke et al., 2013). Recently, the lignin analysis toolkit has been further advanced by the development of a py-GC-MS method that employs uniformly ^{13}C-labelled lignin as internal standard, allowing quantitative analysis of lignin content and structure(van Erven et al. 2017). Proven invaluable for the analysis of fungal ligninolysis this method might provide additional insights into lignin degradation by termites(van Erven et al. 2018).

In a previous work we reported the lignocellulose degradation by gut microbiomes of four higher termite species *Nasutitermes ephratae*, *Nasutitermes lujae*, *Microcerotermes parvus* and *Termes hospes*. These microbiomes were grown on lignocellulosic biomass in anaerobic bioreactors, demonstrating their high potential to convert lignocellulosic substrates into carboxylates(Auer et al. 2017). In this study, we aimed to decipher lignin modification induced by termite gut microbiomes by expanding upon our previous work, using the resulting digested wheat straw to assess lignocellulose modifications by making use of in-depth structural analysis (whole-cell wall HSQC NMR and ^{13}C-IS py-GC-MS), complemented by wet-chemistry.

III.4. EXPERIMENTAL SECTION

III.4.1. Lignocellulose degradation by termite gut inocula in anaerobic bioreactors

The lignocellulose degradation capacity of gut microbiomes from four different species of higher termites *Microcerotermes parvus, Termes hospes, Nasutitermes ephratae*, and an undescribed species closely related to *N. lujae* (herein after *N. lujae*) was assessed in anaerobic bioreactors (Applikon MiniBio 500). Termites guts (500 guts) were used as inocula for two replicate anaerobic bioreactors for each termite species, using a minimal medium and autoclaved (121°C, 20 min, 1.2 bar) milled wheat straw as sole carbon source as previously described(Auer et al. 2017) and detailed in Supporting Information (Supplementary Materials and Methods section). At the end of the 20 days of incubation, VFA and gas production were determined and the whole culture broth was snap frozen in liquid nitrogen and kept at -80°C until further use. For chemical characterization replicate samples were pooled together.

III.4.2. Volatile fatty acid quantification

Volatile fatty acids (VFA) contained in the liquid phase of samples were analyzed using a Varian 3900 gas chromatograph equipped with a flame ionization detector and CP-Wax 58 (FFAP) CB column as described previously(Cavaillé et al. 2013).

III.4.3. Volatile solid quantification

Wheat straw concentration was determined at the beginning and at the end of the 20-day incubation by measuring the total (TS) and volatile (VS) solids, corresponding to the solid organic fraction (For details see Supplementary Information).

III.4.4. Cell wall extraction and grinding

Extractive-free cell wall residues (CWR) were obtained by sonication-solvent extraction of 2g of wheat straw samples (Mansfield et al. 2012). CWR was obtained after drying at room temperature overnight and for 24 h at 55°C. CWR was then milled in a ball mill MM 400 (Retsch) (Supplementary Information).

III.4.5. Polysaccharide analysis and Klason lignin determination

Wheat straw composition was determined on the original wheat straw and the digested samples collected from the bioreactors at the end of the incubation period. Their chemical composition was determined using the sulfuric acid hydrolysis method described by Lazuka et al 2015 (Supplementary Information). The insoluble residue was washed with distilled water and dried at 105 °C overnight to determine Klason lignin content. Sugar composition was determined on an Ultimate 3000 Dionex HPLC with refractive index detector (Thermo Scientific) equipped with a BioRad Aminex HPX 87H affinity column(Lazuka et al. 2015).

To estimate the absolute extent of removal of the different lignocellulose components by the gut microbiomes in the bioreactor experiments, we computed the total remaining mass of polymers (m_p) (Figure 1) per liter as such [1]:

$$[1] m_p = w_p * m_{vs} * \frac{m_{CWR}}{m_t}$$

considering the mass percentage of said polymers (w_p) (Table2) , total mass of volatile solids (VS) remaining in the bioreactors (m_{VS}) (Table 1) corrected by weight loss induced by extraction ($\frac{m_{CWR}}{m_t}$).

III.4.6. Thioacidolysis

Analytical thiacidolysis was performed according to Méchin et al.(Méchin et al. 2014) (Supplementary Information). Trimethylsilylated lignin-derived monomers were quantified on a Trace1300 GC/MS (Thermoscientific) equipped with a Triple

Quadripole-Ion trap and an Agilent DB-5 column (Agilent Technologies).(Ployet 2017)

III.4.7. 2D HSQC NMR

CWR samples were swelled in DMSO-d_6/pyridine-d_5 (4:1, v/v) for the gel-state NMR analysis as described previously.(Tarmadi et al. 2018) NMR spectra were acquired using the Avance III 800US system (Bruker Biospin, Billerica, MA, USA) equipped with a cryogenically cooled 5 mm TCI gradient probe (Bruker Biospin). 2D HSQC NMR experiments were carried out using the standard Bruker implementation ('hsqcetgpsp.3') with the parameters described in literature.(H. Kim et Ralph 2010; Mansfield et al. 2012) Data processing and analysis were performed using TopSpin 4.0 software (Bruker Biospin) as described previously.(Tobimatsu et al. 2013; P. Y. Lam et al. 2017) The central DMSO-d_6 solvent peaks (δ_C/δ_H: 39.5/2.49 ppm) were used as internal references. For volume integration analysis of the aromatic contour signals, C_2–H_2 correlations from G, G′ and F, C2–H2/C6–H6 correlations from S, S′ and P, and C2′–H2′/C6′–H6′ correlations from T were manually integrated and S, S′, P and T integrals were logically halved. For the polysaccharide anomeric signals, C1–H1 correlations from Gl, X, X′, X″, X‴, A and U were integrated. For the lignin side-chain signals, Cα–Hα correlations from I, II, III, III′ and IV″, Cγ–Hγ correlations from IV‴ and Cγ–Hγ correlations from IV and IV′ were integrated, and III, III′, IV, and IV‴ integrals were logically halved. The obtained integration values were normalized on a $½S_{2/6} + ½S'_{2/6} + G_2 + G'_2 + ½H'_{2/6} = 100$ basis for each spectrum.

III.4.8. Quantitative py-GC-MS with ^{13}C lignin as internal standard

Analytical pyrolysis coupled to gas chromatography with high-resolution mass spectrometric detection (Exactive Orbitrap, Thermo Scientific, Waltham, MA, USA) was performed as previously described(van Erven et al. 2019). To each CWR sample (~80 µg), 10 µL of a ^{13}C wheat straw lignin internal standard (IS) solution (1 mg mL^{-}

[1] ethanol/chloroform 50:50 v/v) was added and dried prior to analysis. All samples were prepared and analyzed in triplicate. Lignin-derived pyrolysis products were monitored in full MS mode on the most abundant fragment per compound (both nonlabeled and uniformly ^{13}C labeled). Pyrograms were processed by TraceFinder 4.0 software. Lignin contents and relative abundances of lignin-derived pyrolysis products were calculated as described previously(van Erven et al. 2019).

III.4.9. Statistical analysis

Statistical differences between digested and undigested samples were calculated with t.test, $p<0.05$, n= number of samples

III.5. RESULTS

III.5.1. Wheat straw degradation by termite gut microbiomes in anaerobic bioreactors

Wheat straw biomass was incubated for 20 days in anaerobic bioreactors inoculated with gut microbiomes from higher termites, *N. ephratae*, *N. lujae*, *M. parvus* and *T. hospes.*(Auer et al. 2017) To assess the degradation of wheat straw biomass, total volatile solid (VS) content and volatile fatty acids (VFA) production were determined (Table 1). Regardless of the gut microbiome origins, wheat straw was clearly degraded as determined by the loss of VS, but the extent was different between the gut microbiome origins. The highest degradation rate (%, w/w) was observed for the gut microbiome of *N. ephratae* (45.2 ± 5.0%), followed by that of *N. lujae* (37.1 ± 4.3%), *M. parvus* (31 ± 3.7 %) and *T. hospes* (30 ± 5.0%). Wheat straw in the bioreactors was mainly transformed into VFA, with *N. ephratae* being the best VFA producer. VFA composition was dominated by acetate (>85%) and completed by propionate and butyrate.[24]

Table 1: Volatile solids (VS) content and volatile fatty acid (VFA) production after 20 days of incubation in anaerobic bioreactors. Adapted from Auer et al.(Auer et al. 2017). Values are expressed as means ± SD (n = 3).

	Undigested original wheat straw	Digested wheat straw			
		Gut microbe origin:			
		N. ephratae	*N. lujae*	*M. parvus*	*T.hospes*
VS content (g /l)	18.84	10.31 ± 0.91	11.30 ± 0.83	12.81 ± 0.91	12.81 ± 0.70
VFA production (g/l)	0.32	5.61	3.99	3.46	3.49

III.5.2. Lignocellulose compositional changes in wheat straw degradation by termite gut microbiomes in anaerobic bioreactors

To gain insights into the mechanisms underlying conversions of lignocellulose components by termite gut microbiomes, the samples of cell wall residue (CWR) prepared from the undigested and digested wheat straws were chemically characterized. The changes in lignocellulose composition were determined by sugar analysis and lignin content analysis based on Klason lignin and ^{13}C-IS py-GC-MS approaches (Table 2). Sugar analysis showed slight decreases in cellulose content only in wheat straw CWR digested by *N. ephratae* gut microbiome, whereas apparent cellulose content remained unchanged in the CWR samples digested by the microbes

from *N. lujae* and *T. hospes*, or slightly increased in the CWR samples digested by *M. parvus* microbiome.

Table 2: Lignocellulose compositional analysis of wheat straw samples digested by gut microbiomes from four higher termites in anaerobic bioreactors. Values are expressed as means ± SD g per 100 g cell wall residues (CWR). Values in parentheses indicate percentage change compared to undigested original wheat straw. Statistical differences are marked with * ($p < 0.05$, $n = 3$, Student's *t*-test).

	Undigested original wheat straw	Digested wheat straw			
		Gut microbe origin:			
		N. ephratae	*N. lujae*	*M. parvus*	*T.hospes*
Cellulose (g /100 g CWR)	43.3 ± 2.2	38.4 ± 2.2 (-11.3%)	43.1 ± 2.3 (-0.5%)	45.4 ± 1.1 (+4.8%)	44.2 ± 0.6 (-2%)
Hemicelluloses (g /100 g CWR)					
Total	34.2 ± 1.7	29.9 ± 0.6* (-12.5%)	28.3 ± 0.4* (-17.3%)	29.6 ± 0.7* (-13.4%)	26.3 ± 0.6* (-23.1%)
Xylose	31.2 ± 1.5	29.6 ± 1.2 (-5.1%)	27.0 ± 1.1* (-13.5%)	28.5 ± 0.1* (-8.7%)	25.0 ± 0.6* (-19.9%)
Arabinose	3.5 ± 0.3	2.0 ± 0.1* (-42.9%)	2.0 ± 0.1* (-42.9%)	2.0 ± 0.1* (-42.9%)	1.9 ± 0.1* (-45.7%)
Uronic acid	0.6 ± 0.0	ND (-100%)	0.1 ± 0.1 (-83.3%)	ND (-100%)	ND (-100%)
Lignin (g /100 g CWR)					
by Klason	22.5 ± 3.2	31.7 ± 1.3* (+40.0%)	28.6 ± 4.1* (+27.0%)	25.0 ± 3.4 (+11.0%)	29.6 ± 4.6 (+31.5%)
by ^{13}C-IS py-GC-MS	24.2 ±1.4	NA	28.6 ± 1.5* (+18.2%)	28.6 ± 1.3* (+18.2%)	28.1 ± 0.8* (+16.0%)

ND: not detected; NA: not analyzed.

In contrast, decrease in hemicellulose fraction was observed for all the digested samples, ranging from 12.5% to 23% compared to the initial hemicellulose content of undigested, original wheat straw. Within the hemicellulose fraction, xylose content was reduced by 5-20%, whereas arabinose content was more drastically reduced by ~50%, in all the digested samples. In addition, uronic acid content in the digested samples decreased by more than 80% in all the digested samples. Thus, GAX fractions were preferentially removed by the microbiomes action.

The lignin content in wheat straw samples was assessed by either or both Klason and ^{13}C-IS py-GC-MS (Table 2). The two methods provided overall similar lignin content data, with the ^{13}C-IS py-GC-MS method being superior in precision (RSD <6%), and consistently showed increases of lignin in all the digested samples compared to the original wheat straw. The increment measured by Klason method ranged from 11% to 40% with statistical significance for the samples digested by the gut microbiomes from *N. ephratae* (+40%) and *N. lujae* (+27%) ($p < 0.05$). A more conservative increment was measured by py-GC-MS, varying between 16% and 18% with statistical significance for all the digested samples. Collectively, these data show that, for all the termite gut inocula studied, there is a preferential degradation of hemicellulose fraction, particularly GAX, over cellulose and lignin fractions.

Furthermore, the absolute extent of removal of each lignocellulose fraction by the gut microbiomes was estimated based on the VS loss (Table 1) and wheat straw chemical composition (Table 2) (see Material and Methods). This analysis showed a higher removal of polysaccharides by *N. ephratae* gut microbiome compared to other species, with a 55% loss of both cellulose and hemicellulose fractions (Figure 1). Other termite microbiomes also displayed efficient polysaccharide removal with cellulose degradation ranging from 28% to 47% and hemicellulose degradation ranging from 46% to 55%. Notably, all the termite gut microbiomes displayed higher or at least similar hemicelluloses consumption compared to cellulose. Whereas both *Nasutitermes* (*N. ephratae* and *N. lujae*) gut microbiomes removed similar amounts of cellulose and hemicelluloses, microbiomes from *M. parvus* and *T. hospes* displayed a preferential degradation toward hemicelluloses than toward cellulose. Within

hemicellulose fraction, arabinosyl and uronosyl moieties seemed to be more easily degraded than xylosyl moieties.

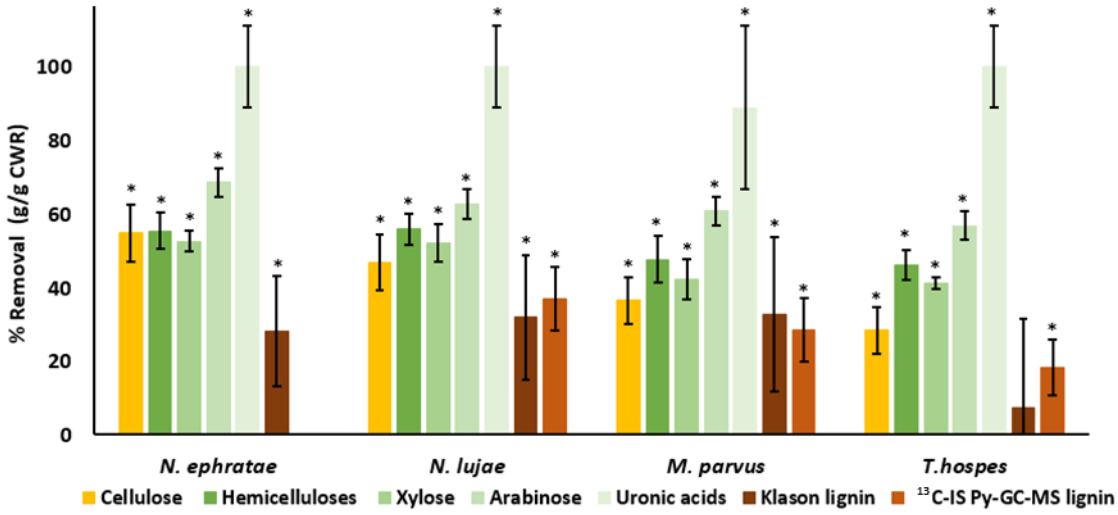


Figure 1. Estimation of absolute lignocellulose compositional changes of wheat straw samples digested by gut microbiomes based on volatile solid loss and chemical composition analysis of the digested residues. Average values and standard deviation are expressed as percentage change compared to undigested original wheat straw (g/100 g CWR), with hemicellulose being the sum of xylose, arabinose and uronic acids. Statistical differences are marked with * ($p < 0.05$, $n = 3$, Student's t-test). ^{13}C-IS Py-GC-MS lignin removal was not measured for *N. ephratae* because of the lack of sample.

Even though the digested wheat straw residues were enriched in lignin content (Table 2), in fact significant amounts of lignin appeared to be removed (Figure 1). The Klason lignin assay suggested that 28-32% of lignin in the original wheat straw was removed by *N. ephratae*, *N. lujae* and *M. parvus* gut inocula, whereas a non-significant lignin removal (7%) was found for *T. hospes*. The corresponding values quantified by py-GC-MS were in a similar order of magnitude for *N. lujae* and *M. parvus*; in contrast, *T. hospes* displayed a lignin degradation of 18% ($p < 0.05$), higher than estimated by the Klason method. Given the accuracy for lignin quantification of the py-GC-MS

method over that of the Klason methodology, we imply that also *T. hospes* is able to remove lignin.

III.5.3. Structural characterization of digested wheat straw residues by thioacidolysis, 2D HSQC NMR and ^{13}C-IS py-GC-MS

To investigate the lignin-unlocking mechanisms of termite gut microbiomes, lignocellulose structures of the original and digested wheat straw CWR samples were further characterized by thioacidolysis, 2D HSQC NMR and ^{13}C-IS py-GC-MS, focusing on the analysis of lignin chemical structures. Due to the limited sample availability of the wheat straw samples digested by *N.ephratae* gut microbes, we report the results obtained for digestion with gut microbiomes from *N. lujae*, *M. parvus* and *T. hospes*.

III.5.3.1. Thioacidolysis

To investigate lignin compositional change before and after the microbial digestion of wheat straw, we performed analytical thioacidolysis, which quantifies lignin-derived monomers released by the chemical cleavage of β–O–4 linkages, the major inter-monomeric linkage type in lignin polymer(Lapierre et al. 1985). Since grass lignins are mainly constituted of G and S units, typically with only a minor fraction of H units(Ralph et al.,2019; Rolando et al., 1992) (as also confirmed by 2D NMR and py-GC-MS below), we only quantified G and S monomers in the present analysis. The total G + S monomer yield based on lignin content was somehow decreased in the CWR samples digested by the *N. lujae* microbe, whereas those determined for samples digested by *M. parvus* and *T. hospes* microbes were overall similar, compared to the undigested wheat straw, though it must be noted that substantial standard deviations were observed (Table 3). An implication for lignin modification by the action of gut microbes was deduced from the changes in S/G monomer ratio (Table 3). For all the three termite gut microbiomes tested, S/G ratio appeared to significantly decrease after

digestion, suggesting a preferential degradation and/or modification of S over G units in lignin β-O-4 units as further demonstrated by NMR and py-GC-MS below.

III.5.3.2. 2D HSQC NMR

2D HSQC NMR was employed to obtain further detailed information on the chemical structures of the wheat straw cell walls digested by the termite gut microbiomes. To gain a global picture of structural alterations in both lignin and polysaccharide fractions, we collected HSQC spectra of the whole CWR samples via the direct dissolution/swelling method using the DMSO-d_6/pyridine-d_5 solvent system.(H. Kim & Ralph 2010; Mansfield et al. 2012) The NMR spectrum of the original, undigested wheat straw cell walls displayed major lignin and polysaccharide signals typical for grass cell walls (Supporting Table 1). The aromatic sub-region (Figure 2a, Supporting Table 1) showed resolved signals from typical S, G and H lignin aromatic units (S, S′, G, G′ and H) and grass-specific cell wall components such as tricin (T) and *p*-coumarates (P)(Karlen et al. 2018) both of which are majorly bound to lignin,(Karlen et al. 2018; W. Lan et al. 2015; P. Y. Lam et al. 2019) and also ferulates (F) mainly bound to GAX.(Karlen et al. 2016) The polysaccharide anomeric sub-region (Figure 2b, Supporting Table 1) displayed anomeric signals associated with cellulose and hemicelluloses, such as those from glucans (Gl), unacetylated (X) and acetylated (X′, X″, and X‴) xylans, arabinan (A) and glucuronan (U)(H. Kim et Ralph 2010; Tarmadi et al. 2018; van Erven et al. 2018).

Furthermore, the oxygenated aliphatic and aldehyde sub-regions displayed signals from various inter-monomeric linkage types, such as β-O-4 (I), β-5 (II) and β-β (III and III′) linkages, as well as polymer end-units, such as cinnamyl alcohol (IV), cinnamaldehyde (IV′), and benzaldehyde (IV″) end-units, in the lignin polymer backbones (Figure 3, Supporting Table 1)(H. Kim et Ralph 2010; A. F. Martin et al. 2019).

Table 3: Thioacidolysis-based lignin compositional analysis of cell wall residues from wheat straw digested by termite gut microbiomes. Values in parentheses indicate percentage change compared to undigested wheat straw. Statistical differences are marked with *($p < 0.05$, $n = 3$, Student's *t*-test).

	Undigested original wheat straw	**Digested wheat straw**		
		Gut microbe origin:		
		N. lujae	***M. parvus***	***T.hospes***
Monomer yield (µmol /g Klason lignin)				
G	**429± 41**	326± 46* (-24%)	544±121 (21%)	446±63 (4%)
S	**543± 90**	380± 46* (-30%)	592±186 (9%)	423±58 (-22%)
G + S	**972± 132**	706 ± 63* (-27%)	1,133 ± 303 (17%)	871 ± 107 (-11%)
Monomer yield (µmol /g py-GC-MS lignin)				
G	**400± 38**	326± 46 (-19%)	476±106 (16%)	504±57 (20%)
S	**504± 84**	380± 46 (-25%)	517±162 (3%)	480±64 (-5%)
G + S	**904± 122**	706 ± 63 (-22%)	993± 268 (+10%)	984±118 (9%)
Monomer ratio (mol/mol)				
S/G	**1.28 ± 0.06**	0.86 ± 0.08*	1.08 ± 0.10*	0.96 ± 0.05*

	(-33.3%)	(-15.6%)	(-25%)

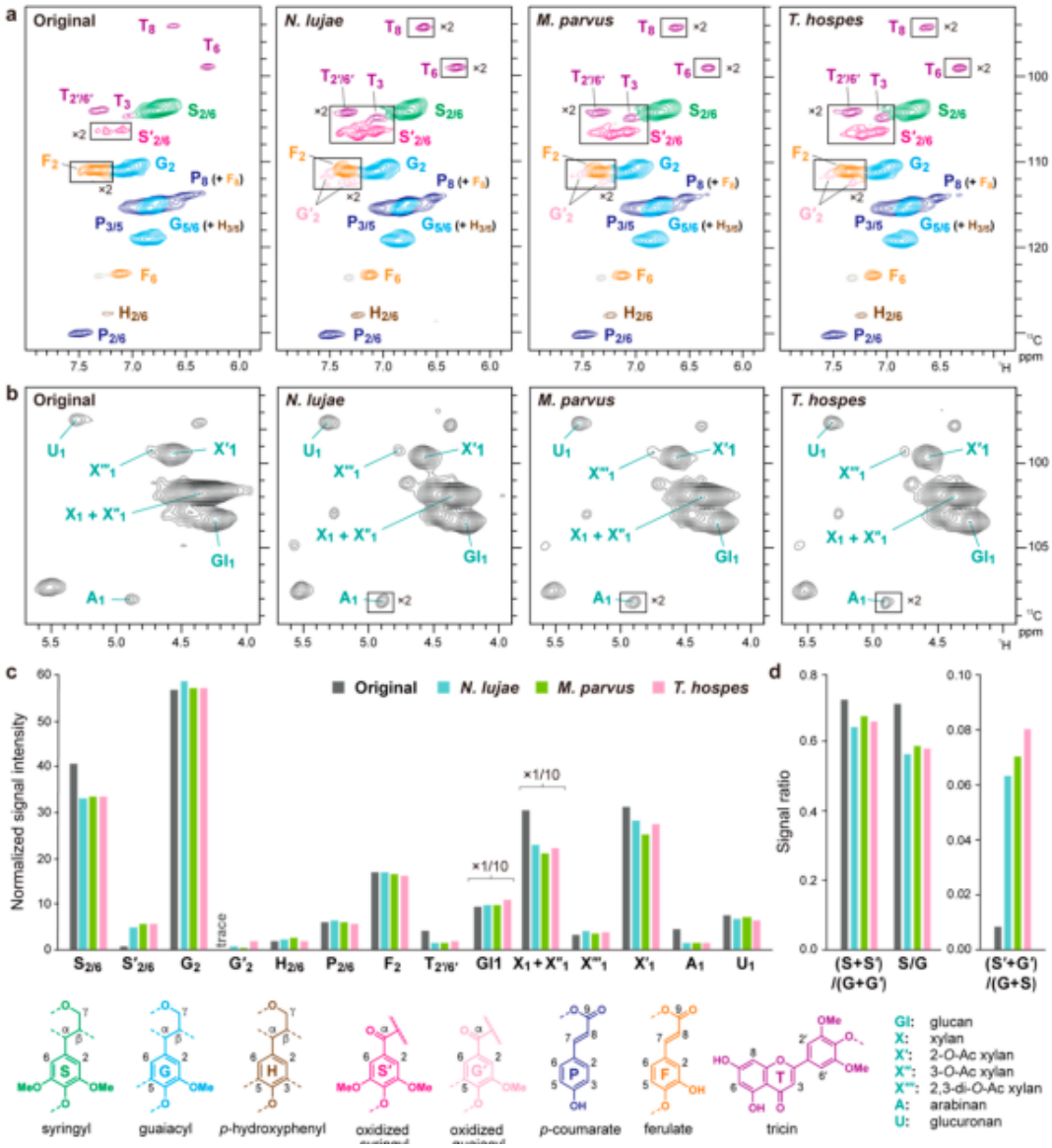


Figure 2. Partial short-range 1H-13C correlation (HSQC) NMR spectra of wheat straw cell walls before (original) and after digestion by gut microbiomes from *N. lujae, M. parvus* and *T. hospes.* Aromatic (a) and polysaccharide anomeric (b) sub-regions are shown. Signal assignments are listed in Supporting Table S1. Boxes labelled ×2 represent regions with the scale vertically enlarged by 2-fold. In (c) and (d) normalized signal intensity values of major aromatic and anomeric signals expressed on a ½S2/6

+ ½S′2/6 + G2 + G′2 + ½H′2/6 = 100 basis (c), and ratios of S (½S2/6 and ½S′2/6) and G (G2 and G′2) aromatic signals (d) are shown. Data labelled ×1/10 indicate that the reported values are divided by a factor 10 for visualization purposes.

The compositional changes in the wheat straw cell walls before and after termite microbiome digestion were deduced by comparing the volume integrations of the HSQC contour signals. For comparison between the original and digested cell wall NMR spectra, major lignin aromatic and polysaccharide anomeric signals were integrated and normalized based on the sum of the total of major lignin aromatic signals ($S_{2/6}$ + $S'_{2/6}$ + G_2 + G'_2 + $H_{2/6}$) (Figure 2c). Glucan (Gl) and xylan (X, X′, X″, and X‴) were the predominant polysaccharide signals detected in all the cell wall spectra (Figure 2a) and reflect the proportional changes of cellulose and hemicellulosic arabinoxylans in cell walls, respectively. Glucan signals from crystalline cellulose, however, are substantially underestimated by the current gel-state NMR method due to its incomplete gelation in the solvent.(H. Kim et Ralph 2010; Mansfield et al. 2012) Regardless of the termite gut microbiome tested, slight increases of glucan (Gl) signals and decreases of signals from non-acetylated and mono-acetylated xylan (X, X′ and X″) and arabinan (A) were prominent after digestion (Figure 2c). This result further corroborates the preferential degradation of GAX over cellulose and lignin by the termite gut microbiomes as determined earlier by the wet-chemical analyses (Table 2 and 3). In contrast to clear depletions in non-acetylated and mono-acetylated xylan (X, X′ and X″) and arabinan (A) signals, 2,3-di-*O*-acetylated xylan (X‴) and ferulate (F) signals appeared to be rather similar between all the digested and undigested control cell wall spectra (Figure 2c), suggesting that heavily acetylated xylan and feruloylated arabinan moieties in GAX may be more tolerant to the microbial digestion compared to the non-acetylated and mono-acetylated xylan moieties constituting the main GAX polymer chain. It is noteworthy that the proportion of glucuronan (U) signals was not significantly affected in all the digested samples. This result contradicts the notable reduction of uronic acid contents in the digested cell walls as determined by the wet-chemical analysis (Table 2 and 3).

In line with the altered S/G lignin unit ratio determined by thioacidolysis, which does not include oxidized forms (S' and G') (Table 3), and also partially by py-GC-MS as described below, the ratio of total S and G lignin aromatic units based on the (S + S′)/(G + G′) NMR signal ratio as well as the ratio of non-oxidized S and G units based on the S/G signal ratio decreased in all the digested wheat straw samples (Figure 2c). Intriguingly, we detected considerable augmentations of the Cα-oxidized S (S′) signals in the digested wheat straw cell wall spectra compared to those in the undigested control cell wall spectrum. The Cα-oxidized G signals (G′) also apparently increased but in lesser extent to oxidized S signals (S′) (Figure 2c). Consequently, the (S′ + G′)/(S + G) signal ratio remarkably increased in all digested samples, rising up to ca. 10-fold in the spectrum of the cell wall sample digested by the *T. hospes* gut microbiome (Figure 2c). In addition, lignin-bound tricin signal (T) appeared to be sharply depleted in all the digested cell wall spectra (Figure 2c), suggesting a preferential degradation of lignin-bound tricin units by the termite gut microbes.

We also performed volume integration analysis for the signals from the inter-monomeric linkages and end-units in lignin polymer appearing in the oxygenated aliphatic and aldehyde sub-regions of the HSQC spectra (Figure 3). Overall, no drastic differences in the relative distributions of the major inter-monomeric linkage signals, such as those from β-O-4 (I), β-5 (II) and β-β (III and III′) linkages, were detected between the spectra of wheat cell walls before and after digestion by either of the three termite gut microbiomes tested, although in absolute terms (per 100 aromatic rings) augmentations in these intact lignin inter-monomeric unit linkages could be observed. The intensities of I, II, III and III′ signals normalized based on S + S′ + G + G′ + H generally increased in all the digested cell wall spectra compared to the undigested control spectrum (Figure 3d). Meanwhile, we detected increased aldehyde signals from cinnamaldehyde (IV′) and benzaldehyde (IV″) end-units in all the digested wheat straw samples (Figure 3c).

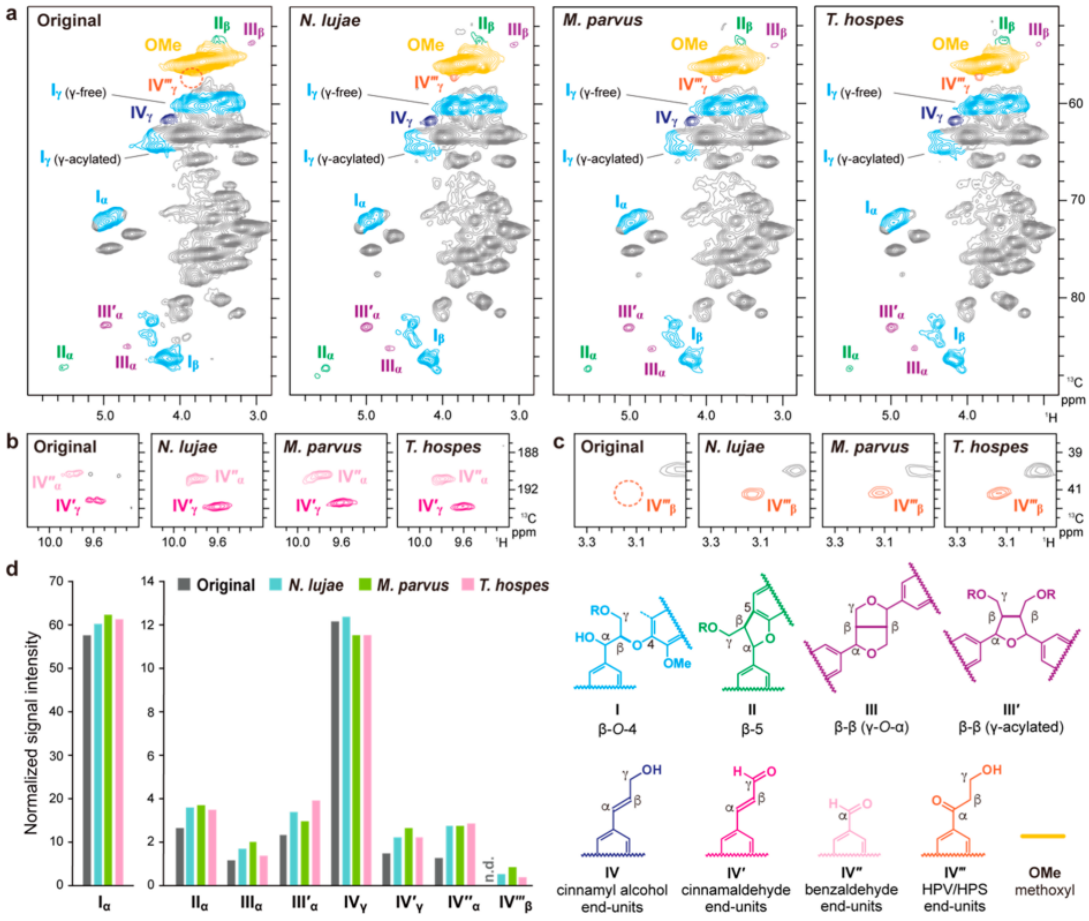


Figure 3. Partial short-range 1H-13C correlation (HSQC) NMR spectra of wheat straw cell walls before (original) and after digestion by gut microbiomes from *N. lujae, M. parvus* and *T. hospes*. Oxygenated aliphatic (a and c) and aldehyde (b) sub-regions are shown. Signal assignments are listed in Supporting Table S1. In (d), normalized signal intensity values expressed on a ½S2/6 + ½S′2/6 + G2 + G′2 + ½H′2/6 = 100 basis (see Figure 2), are shown. n.d., not detected.

More strikingly, a set of new signals attributed hydroxypropiovanilone (HPV) and hydroxypropiosyringone (HPS) end-units (IV‴; δC/δH, Cβ-Hβ at 41.5/3.14 and Cγ-Hγ at 57.0/3.85)[36] appeared clearly in all the digested cell wall spectra, whereas these HPV/HPS (IV‴) signals are undetectable in the undigested cell wall spectra (Figure 3, a, c and d). These data are in line with the augmented Cα-oxidized lignin aromatic unit signals (S′ and G′) in the aromatic sub-regions (Figure 2c) and further consolidate our

notion that lignin polymers are at least partially deconstructed by the action of termite gut microbiomes.

III.5.3.3. ^{13}C-IS py-GC-MS analysis

Besides quantification of lignin content, ^{13}C-IS py-GC-MS analysis also provides valuable information of lignin structures. The different lignin-derived pyrolysis products of the original and digested wheat straw were classified according to their structural moieties and summarized (Table 4 and Supporting table 2).

The original wheat straw displayed a typical structural composition of grass lignin with H:G:S ratio of 10:59:30, in agreement with the composition reported in previous studies(van Erven et al. 2017; 2018). In addition, 4-vinylphenol and 4-vinylguaiacol, majorly derived from *p*-coumaric acid and ferulic acid, respectively, were present in expected levels, as were the other detailed structural features (Table 4).

^{13}C-IS py-GC-MS analysis of the wheat straw digested by termite gut microbiomes showed a similar subunit composition as the undigested wheat straw, with minimal variation between the species (H:G:S *N. lujae* 8:61:31, *M. parvus* 8:62:30, *T. hospes* 9:62:29). Compared to undigested wheat straw, a slight decrease on S/G ratio was observed, particularly for *T. hospes*. These observations are consistent with the aromatic compositional analyses based on thioacidolysis (Table 3) and 2D HSQC NMR (Figure 2).

Even though the overall subunit composition did not differ between undigested and digested wheat straw, a more detailed structural classification revealed significant differences (Table 4), but with little variation between the termite species. A significant decrease of vinyl products (~13%) was observed in digested straws, which was primarily due to decreases in 4-vinylphenol (~20%) and 4-vinylguaiacol (~12%). This suggests the targeted removal of the corresponding hydroxycinnamic acids *p*CA and FA, these products are largely, yet not exclusively, derived of, again in line with the slight decrease of these moieties as observed by NMR analysis (Figure 2c).

Table 4: ^{13}C-IS py-GC-MS relative abundance of residual lignin compounds of wheat straw before and after the digestion by the gut microbiomes of *N. lujae*, *M. parvus* and *T. hospes*. Values are expressed as means ± SD. Statistical differences are marked with * ($p < 0.05$, $n = 3$, Student's *t*-test).

	Undigested original wheat straw	Digested wheat straw		
		Gut microbe origin:		
		N. lujae	*M. parvus*	*T.hospes*
Lignin units (abundance in mol%)				
H	**10.7 ± 1.1**	8.5 ± 0.4*	8.8 ± 0.2*	9.1 ± 0.3
G	**58.9 ± 1.1**	60.9 ± 0.3*	61.9 ± 0.9*	61.9 ± 0.7*
S	**30.4 ± 1.0**	30.6 ± 0.7	29.4 ± 0.8	29.0 ± 0.9
S/G	**0.52 ± 0.02**	0.50 ± 0.01	0.48 ± 0.02	0.47 ± 0.02*
Structural moieties (abundance in mol%)				
Unsubstituted	**5.2 ± 0.14**	6.0 ± 0.53	6.4 ± 0.15*	6.2 ± 1.0
methyl	**3.23 ± 0.17**	3.28 ± 0.19	3.28 ± 0.27	3.69 ± 0.39
Vinyl	**41.48 ± 1.91**	36.19 ± 0.52*	36.14 ± 0.79*	36.17 ± 1.80*
4-VP	**9.07 ± 0.96**	6.96 ± 0.31*	7.26 ± 0.18*	7.46 ± 0.35*
4-VG	**28.31 ± 1.16**	25.29 ± 0.38*	25.04 ± 0.95	24.95 ± 1.23
C_α-ox	**3.89 ± 0.06**	4.06 ± 0.15	3.82 ± 0.09	4.03 ± 0.43
C_β-ox	**2.08 ± 0.03**	2.14 ± 0.06	2.04 ± 0.11	2.14 ± 0.05
C_γ-ox	**40.3 ± 1.96**	44.3 ± 0.35*	44.5 ± 1.36*	43.5 ± 1.84
Miscellaneous	**3.84 ± 0.13**	4.02 ± 0.14	3.86 ± 0.22	4.26 ± 0.19
PhC3	**45.31 ± 2.01**	49.59 ± 0.44*	49.58 ± 1.20*	49.06 ± 1.97

4-VP = 4-Vinylphenol

4-VG = 4-Vinylguaiacol

PhC3 = Phenols with intact α, β, and γ carbon side chain, excluding diketones

Interestingly, Cγ-oxidized pyrolysis products were found to be significantly increased (~10%) upon digestion, concomitantly also increasing the abundance of pyrolysis products with an intact α,β,γ-side chain (PhCγ) to similar extents. We have previously established that these products correlate well with the amount of intact interunit linkages present, as determined by HSQC NMR(van Erven et al. 2018; van Erven, Hilgers, et al. 2019; van Erven, Wang, et al. 2019). Indeed, also the NMR analysis showed a relative increase of intact interunit linkages (per 100 aromatic rings) (Figure 3d).

Despite the fact that the NMR analyses showed a significant augmentation of benzaldehyde units in the digested samples (Figure 3, b and d), nor the grouped Cα-oxidized pyrolysis products (Table 4), nor vanillin and syringaldehyde (Table S2), displayed significant differences in comparison to the undigested wheat straw sample. However, their respective decarbonylated analogues, guaiacol and syringol did increase (Table S2). Aldehyde decarbonylation reactions are likely to occur during pyrolysis, but note that guaiacol and syringol are formed from a wealth of other substructures and, therefore, should not be interpreted as markers for the occurrence of benzaldehyde substructures. Interestingly, coinciding with the increase of Cα-oxidized subunits, the HPV/HPS end-units became apparent in the NMR spectra of the treated samples (Figure 3). We have previously shown that these structures give rise to acetovanillone/guaiacyl vinyl ketone and acetosyringone/syringyl vinyl ketone upon pyrolysis, respectively(van Erven, Hilgers, et al. 2019). Indeed, our py-GC-MS analysis here showed that the former marker products increased approx. 20% in the digested samples (from 0.55 to 0.65% of the relative abundance of lignin-derived pyrolysis products), whereas the latter pair was not significantly different (Supporting Table S2). Overall, our data further corroborate the appearance of the HPV/HPS moieties during the digestion with termite gut microbes.

III.6. DISCUSSION

To gain insights into the fate of wheat straw lignin after digestion by gut microbiomes of higher termites in anaerobic bioreactors, we combined wet-chemistry,

2D HSQC NMR and ^{13}C-IS py-GC-MS measurements. Collectively, our findings suggest that lignin polymer in wheat straw is partially modified and/or depolymerized by the gut microbiomes when implemented in anaerobic bioreactors. Nevertheless, in line with previous observations,(Tarmadi et al. 2018; Katsumata et al. 2007; Ke et al., 2013) all the termite gut microbiomes tested in this study preferentially used polysaccharides, rather than lignin, as preferential carbon sources (Table 2 and 3). In particular, the changes in GAX fractions as assessed by wet-chemistry (Table 2 and 3) and 2D HSQC NMR (Figure 2) suggested that they are more susceptible towards microbial degradation than cellulose and lignin. Consequently, the lignocellulose composition of the digested wheat straw residues presented largely increased lignin contents (by ca. 23% ± 10) compared with the undigested wheat straw (Table 2). Previous studies reported similar results, showing lignin accumulation in termite faeces, suggesting no degradation of lignin by termite(Cookson 1987). Importantly, however, our estimation of absolute lignocellulose compositional changes considering the VS loss (Table 1) suggested that a considerable portion of lignin (26% ± 10) was also removed along with polysaccharides during the microbial degradation (Figure 1), suggesting that the termite gut microbiomes indeed harbor an ability to remove wheat straw lignin in anaerobic bioreactors. We like to point out that, based on our structural analyses by 2D HSQC NMR and py-GC-MS, it is estimated that at least 80% of the observed delignification can be attributed to the true removal of lignin, rather than the removal of *p*-coumaric acid, ferulic acid and tricin moieties.

Our thioacidolysis, 2D HSQC NMR and py-GC-MS analyses provided overall consistent lignin compositional changes induced by the termite gut microbiomes. While the total S and G lignin unit analysis by py-GC-MS and 2D HSQC NMR (based signal ratio of the sums of non-Cα-oxidized and Cα-oxidized **S** + **S′** and **G** + **G′**) suggested overall little changes in the total S/G lignin unit ratio (Figure 2 and Table 4) , the ratio of the non-oxidized S/G lignin units as estimated by thioacidolysis (based on β-O-4-derived S/G monomer ratio) and 2D HSQC NMR (based on signal ratio of only non-Cα-oxidized **S** and **G**) appeared to be decreased in all the digested cell wall samples (Table 3 and Figure 2). Collectively, these data suggested that S units in the original wheat lignin polymer might be more preferentially modified than G units, and

most likely, converted into the corresponding Cα-oxidized S units by the gut microbiomes of higher termites in the anaerobic bioreactors, which will be elaborated on below. 2D HSQC NMR also showed a high level of tricin depletion (ca. 50-65% reductions) in lignin residues left after the digestion. The preferential degradation of tricin or tricin-containing fractions by termites is interesting given the fact that these moieties were shown to (relatively) accumulate in biological oxidative delignification strategies, both by the white-rot fungi *Ceriporiopsis subvermispora* and laccase-mediator treatments. Nevertheless, it has also been shown that tricin can be removed by both chemical (Miyamoto et al. 2018) or biological processes,(van Erven et al. 2018) probably due to its terminal position in the polymeric chain of grass lignin(Lan et al. 2015). For instance, previous studies focused in lignin degradation by *Pleurotus eryngii* showed a strong removal of tricin. Similarly, a preferential removal of oligomeric lignin fractions containing tricin terminal units was observed in the degradation of rice lignocellulose in the *C. formosanus* gut digestive system(Tarmadi et al. 2018).

The evidence for lignin side-chain modifications induced by the termite gut microbes were obtained by both 2D HSQC NMR and py-GC-MS. Most strikingly, they both clearly detected, albeit at low levels, appearance of the HPV/HPS end-units harboring Cα-oxidized side-chains in the digested lignin residues (Figure 3, Table 4 and Supporting Table 2). Side-chain oxidation of Cα is characteristic of lignin modification induced by white-rot-fungi(van Kuijk et al. 2017), and it has also been previously observed in feces of the beetle *Anoplophora glabripennis* and the lower termite *Zootermopsis angusticollis* that digest wood anaerobically(Geib et al. 2008). In the present work, the increase of Cα-oxidized aromatic signals are partially associated with the increases HPV/HPS end-units (Figure 2, Table 4 and Supporting Table 2). The HPV/HPS products have been described for lignin degradation by the action of β-etherases, enzymes that degrade lignin in a reductive cascade.(Higuchi et al. 2018) The presence of β-etherases, particularly of glutathione-dependent-β-etherases cleaving β-aryl-ether lignin compounds was first reported in *Sphingobium* (Gall et al. 2014) Since then, its activity has been shown in *in-vitro* experiments (Masai et al. 1989; Gall et al. 2018) and its presence has also been reported in multiple

bacteria including anaerobic strains such as *Tolumonas lignolytica*(Billings et al. 2015). Most of the bacterial species harbouring homologous glutathione-dependent-β-etherases genes fall into the Proteobacteria phylum, members of which have been also found in termite guts(Kontur et al. 2019; Voß et al. 2020). The gut of termites hosts a large variety of microbes and in the termite species studied here Proteobacteria represented up to 11% of the total gut microbiome. This fraction increased during the bioreactor experiments reaching up to 40% of the total community(Auer et al. 2017). Overall, our results could suggest that bacteria harboring β-etherases genes may have been involved in the cleavage of β-O-4 linkages in the lignin polymer in the current bioreactor experiments. β-etherases are thus far thought to be enzymes associated to the cell membrane (Masai et al. 1989) and their activity on polymeric lignin has only be demonstrated in *in-vitro* experiments (Masai et al. 1989; Gall et al. 2018). Nevertheless, although there is no experimental evidence for this hypothesis, β-etherases could still be transported extracellularly in dedicated outer membrane vesicles as it has been observed for enzymes acting on lignin-derived aromatics, and upon lysis of these vesicles act on polymeric lignin(Salvachúa et al. 2020).To our knowledge, no earlier study on lignocellulose deconstruction by termite gut microbiomes has observed the appearance of HPV/HPS in degraded lignin residues, providing potential evidence of the involvement of lignin β-etherases in the termite gut microbiome.

In addition to the HPV/HPS signatures, 2D HSQC NMR also detected considerably augmented aldehyde end-units, which, unlike HPV/HPS, might be the result of cleavage of the Cα-Cβ bond[51] in the digested lignin residues (Figure 3). Previous py-GC-MS-aided structural studies on lignin degradation by several lower termite species reported pyrograms displaying phenolic compounds annotated as markers for oxidized lignin side-chains (Ke et al., 2011; Ke et al., 2013; Geib et al. 2008). Meanwhile, however, a recent study employing 2D HSQC NMR for dissecting lignin degradation by a lower termite *C. formosanus* detected no clear evidence for such lignin side-chain oxidation(Tarmadi et al. 2018). Although the appearance of the aldehyde end-units in lignin has been described as typical consequences of oxidative lignin degradation undertaken by wood-decaying fungi harboring strong lignin oxidases,(Yelle et al.

2008) it is fairly doubtful that similar lignin oxidation cascades can be operative during lignin degradation by the termite gut microbiomes under strictly anaerobic conditions. Even though bacterial strains with lignin peroxidase activity have been isolated from termite digestive systems,(H. Zhou et al. 2017; Le Roes-Hill et al., 2011) and members of the *Pseudomonas* genus producing an elaborate lignin-active enzymatic arsenal, (Prabhakaran et al. 2015) their involvement during anaerobic lignin degradation remains elusive. Interestingly, putative peroxidase genes, including glutathione and DyP-type peroxidases, and multiple cytochrome oxidases were found in anaerobic lignin-degrading Enterobacteriaceae (γ-Protobacteria),(Woo et al. 2014, DeAngelis et al. 2013) members of which were found in our bioreactors, and hence could have been involved(Auer et al. 2017). Clearly, future studies should further explore this aspect in combination with advanced meta-omics analysis to reveal new lignin degradation pathways, enzymes and microorganisms responsible for the appearance of such oxidized lignin structures, in particular under anaerobic conditions.

Since both 2D HSQC NMR and py-GC-MS suggested overall increases of intact lignin inter-monomeric linkages per total lignin aromatic units in the digested lignin residues (Figure 3 and Table 4), major lignin degradation pathways operated by the termite gut microbiomes may involve cleavages of C–C-bonded (condensed) lignin substructures which can in general not be well assessed by the current NMR and py-GC-MS analyses. Likewise, previous NMR studies on lignin degradations by a lower termite *C. formonous*(Tarmadi et al. 2018) and fungus-cultivating higher termite *Odontotermes formosanus*(Hongjie Li et al. 2017) noted similar views based on their observations of proportionally increased intact β-O-4 linkages in the digested lignin residues. Given the limited degree of structural modifications in the digested lignin residues as assessed by NMR and py-GC-MS, the net removal of lignin measured in the current bioreactor experiments, estimated as to be up to ca. 37% (Figure 2) was rather remarkable. As proposed in previous studies,(Ke et al., 2013) the preferential degradation of hemicelluloses by gut microbiomes might facilitate the removal of lignin fractions cross-linked to hemicelluloses. Indeed, it is well known that a significant portion of grass lignin is covalently cross-linked to GAX via ferulates and the microbial production of feruloyl esterases, have been previously observed in

metatranscriptomic studies of termite microbiome,[60] could provide means to decouple them.(Ralph et al., 2019; Karlen et al. 2016) A more in-depth structural analyses of degraded lignin fractions including those potentially present in the liquid fractions of the bioreactors are required to clarify these aspects. Such lignin structural studies, in combination with metagenomic, metatranscriptomic, metaproteomic, and biochemical characterizations of termite gut microbiomes, would provide further insights into the elusive mechanisms of lignocellulose biodegradation by wood-feeding insects, and contribute to their potential applications for lignocellulose biorefinery.

III.7. ACKNOWLEDGMENT

The authors thank Dr. Hironori Kaji and Ms. Ayaka Maeno (ICR, Kyoto University) for their support in NMR experiments. This study was supported in part by PHC Sakura Program from MAEDI-MENESR and the Japan Society for the Promotion of Science (JSPS), JSPS KAKENHI grants (grant no. JP#16H06198 and JP#20H03044) and a research grant from Exploratory Research on Sustainable Humanosphere Science from RISH, Kyoto University. A part of this study was conducted using the facilities in the DASH/FBAS of RISH, Kyoto University, and the NMR spectrometer at the JURC of ICR, Kyoto University. This work has received funding from the Bio Based Industry Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement no. 720303 (EU-Zelcor, Zero waste ligno-cellulosic biorefineries by integrated lignin valorization). It was also supported by the French National Institute of Research for Agriculture, Food and Environment (INRAE), the Region Languedoc-Roussillon Midi-Pyrénées grant 31000553, the Carnot Institute 3BCAR—Insyme project and the PHC Sakura program from Campus France under the project N° 45094PM. The authors thank the GeT Plage platform of INRAE Toulouse, E. Mengelle, M. Bounouba and S. Dubos for the technical support with bioreactor experiments.

III.7.1.1. Author Contributions

L.D. prepared the samples and carried out and analyzed wet chemistry experiments with thioacidolysis performed under FM supervision. PYL and YT carried out and analyzed 2D HSQC NMR experiments. GvE and MK performed and analyzed py-GC-MS experiments. As project coordinator, GHR designed the study, participated on experimental design and contributed at all stages. The manuscript was written by LD under GHR supervision and with important intellectual contributions ~~of~~ from all authors. All authors have given approval to the final version of the manuscript. GHR and YT participated in funding this research.

III.7.1.2. Conflicts of interest

The authors declare no competing financial interest.

III.8. SUPPORTING INFORMATION

Supporting Table S1. Peak assignments for 2D HSQC NMR spectra of wheat cell walls.

Labels	δ_C/δ_H (ppm)	Assignment
Lignin and cinnamate aromatic signals		
$S_{2/6}$	104.2/6.77	C2–H2 and C6–H6 in syringyl units
$S'_{2/6}$	106.8/7.23, 106.6/7.07	C2–H2 and C6–H6 in Cα-oxidized syringyl units
G_2	111.2/7.06	C2–H2 in guaiacyl units
$G_{5/6}$	119.3/6.88, 114.9/6.78	C5–H5 and C6–H6 in guaiacyl units
G'_2	112.5/7.32, 111.6/7.53	C2–H2 in Cα-oxidized guaiacyl units
$H_{2/6}$	128.0/7.23	C2–H2 and C6–H6 in *p*-hydroxyphenyl units
$H_{3/5}$	114.9/6.74	C3–H3 and C5–H5 in *p*-hydroxyphenyl units
T_3	104.9/7.06	C3–H3 in tricin residues
T_6	98.9/6.31	C6–H6 in tricin residues
T_8	94.2/6.63	C8–H8 in tricin residues
$T_{2'/6'}$	104.3/7.36	C2′–H2′ and C6′–H6′ in tricin residues

$P_{2/6}$	130.1/7.50	C2–H2 and C6–H6 in *p*-coumarate residues
$P_{3/5}$	115.6/6.85	C3–H3 and C5–H5 in *p*-coumarate residues
P_8	113.8/6.36	C8–H8 in *p*-coumarate residues
F_2	111.0/7.36	C2–H2 in ferulate residues
F_6	123.2/7.12	C6–H6 in ferulate residues
Lignin inter-monomeric linkage and end-unit signals		
I_α	71.5/4.87	Cα–Hα in β–*O*–4 units
I_β	86.2/4.22, 83.9/4.38	Cβ–Hβ in β–*O*–4 units
I_γ (γ-free)	60.5/3.86	Cγ–H γ in γ-free β–*O*–4 units
I_γ (γ-acylated)	64.6/4.30	Cγ–H γ in γ-acylated β–*O*–4 units
II_α	87.2/5.50	Cα–Hα in β–5 substructures
II_β	53.6/3.49	Cβ–Hβ in β–5 substructures
III_α	85.0/4.67	Cα–Hα in resinol-type β–β substructures
III'_α	82.9/4.98	Cα–Hα in tetrahydrofuran-type β–β substructures
IV_γ	61.8/4.14	Cγ–Hγ in cinnamyl alcohol end-units
IV'_γ	193.9/9.61	Cγ–Hγ in cinnamylaldehyde end-units
IV''_α	190.8/9.81	Cα–Hα in benzaldehyde end-units
IV'''_β	41.2/3.13	Cβ–Hβ in HPV/HPS end-units
IV'''_γ	57.3/3.83	Cγ–Hγ in HPV/HPS end-units

Supporting Table S1 (continued)

Lignin methoxyl signals		
OMe	55.7/3.74	C–H in aromatic methoxyl groups
Polysaccharide anomeric signals		
Gl_1	103.5/4.44	C1–H1 in (1→4)-β-D-glucopyranosyl units
X_1	102.1/4.28	C1–H1 in (1→4)-β-D-xylopyranosyl units
X'_1	99.9/4.58	C1–H1 in 2-*O*-acetyl-β-D-xylopyranosyl units
X''_1	101.8/4.40	C1–H1 in 3-*O*-acetyl-β-D-xylopyranosyl units
X'''_1	99.1/4.78	C1–H1 in 2,3-di-*O*-acetyl-β-D-xylopyranosyl units

A_1	108.2/4.90	C1–H1 in α-L-arabinofuranosyl units
U_1	97.7/5.27	C1–H1 in 4-*O*-methyl-α-D-glucuronopyranosyl units

Measured in DMSO-d_6/pyridine-d_5 (4:1, v/v). Signal assignment was based on comparison with NMR data in literature.[1-7]

Supporting Table S2. Identity, structural classification and relative abundance of lignin-derived pyrolysis products by quantitative ^{13}C-IS py-GC-MS. Average relative abundance of analytical triplicates.

#	Compound	CAS	Retention time (min)	Structural feature	Sidechain length	M_w ^{12}C (g·mol^{-1})	Quan ion ^{12}C [M-*e*]	Original	*N. lujae*	*M. parvus*	*T. hospes*
1	phenol	108952	9.79	H, unsub.	0	94	94.04132	0.7	0.7	0.7	0.7
2	guaiacol	90051	10.03	G, unsub.	0	124	124.05188	2.8	3.5	3.8	3.6
3	2-methylphenol	95487	11.03	H, methyl	C_α	108	108.05698	0.1	0.1	0.1	0.2
4	4-methylphenol (+3-MP)	106445	12.00	H, methyl	C_α	108	107.04914	0.4	0.4	0.3	0.5
5	4-methylguaiacol	93516	12.71	G, methyl	C_α	138	138.06753	1.6	1.7	1.7	1.9
6	2,4-dimethylphenol	105679	13.18	H, methyl	C_α	122	107.04914	0.1	0.1	0.1	0.1
7	4-ethylphenol	123079	14.25	H, misc.	C_β	122	107.04914	0.1	0.1	0.1	0.1
8	4-ethylguaiacol	2785899	14.83	G, misc.	C_β	152	137.05971	0.2	0.2	0.2	0.2
9	4-vinylguaiacol	7786610	16.29	G, vinyl	C_β	150	150.06753	26.7	23.7	23.4	23.3
10	4-vinylphenol	2628173	16.46	H, vinyl	C_β	120	120.05697	8.1	6.2	6.5	6.6
11	eugenol	97530	16.89	G, misc.	C_γ	164	164.08318	0.2	0.2	0.2	0.3
12	4-propylguaiacol	2785877	16.99	G, misc.	C_γ	166	137.05971	0.1	0.1	0.1	0.2
13	syringol	91101	17.64	S, unsub.	0	154	154.06245	1.7	1.9	2.0	1.9
14	*cis*-isoeugenol	97541	18.25	G, misc.	C_γ	164	164.08318	0.1	0.1	0.1	0.1
15	4-propenylphenol	539128	19.24	H, misc.	C_γ	134	133.06479	0.1	0.1	0.1	0.1
16	*trans*-isoeugenol	97541	19.50	G, misc.	C_γ	164	164.08318	0.8	1.0	0.9	1.0
17	4-methylsyringol	6638057	19.86	S, methyl	C_α	168	168.07810	0.8	0.8	0.8	0.8
18	vanillin	121335	19.99	G, C_α-O	C_α	152	151.03897	1.2	1.2	1.1	1.2
19	4-propyneguaiacol	-	20.23	G, misc.	C_γ	162	162.06753	0.1	0.1	0.1	0.1
20	4-alleneguaiacol	-	20.49	G, misc.	C_γ	162	162.06753	0.1	0.1	0.1	0.1
21	homovanillin	5603242	21.44	G, C_β-O	C_β	166	137.05971	0.4	0.5	0.4	0.5
22	4-ethylsyringol	14059928	21.58	S, misc.	C_β	182	167.07022	0.1	0.1	0.1	0.1
23	vanillic acid methyl ester	3943746	21.82	G, C_α-O	C_α	182	182.05736	0.01	0.01	0.01	0.01
24	acetovanillone	498022	21.89	G, C_α-O	C_β	166	151.03897	0.4	0.4	0.4	0.4
25	4-hydroxybenzaldehyde	123080	22.76	H, C_α-O	C_α	122	121.02848	0.1	0.1	0.1	0.1
26	4-vinylsyringol	28343228	22.90	S, vinyl	C_β	180	180.07810	4.8	4.6	4.5	4.4
27	guaiacylacetone	2503460	23.10	G, C_β-O	C_γ	180	137.05971	0.3	0.4	0.4	0.4
28	4-allylsyringol	6627889	23.31	S, misc.	C_γ	194	194.09373	0.3	0.3	0.2	0.3
29	propiovanillone	1835149	23.79	S, C_α-O	C_γ	180	151.03897	0.03	0.03	0.03	0.03
30	guaiacyl vinyl ketone	-	24.09	G, C_α-O	C_γ	178	151.03897	0.2	0.2	0.2	0.2
31	guaiacyl diketone	2034608	24.32	G, C_α-O, C_β-O	C_γ	194	151.03897	0.3	0.3	0.3	0.3
32	cis-4-propenylsyringol	26624135	24.43	S, misc.	C_γ	194	194.09373	0.2	0.2	0.1	0.2
33	4-propynesyringol	-	25.06	S, misc.	C_γ	192	192.07810	0.2	0.2	0.1	0.2
34	4-allenesyringol	-	25.27	S, misc.	C_γ	192	192.07810	0.1	0.1	0.1	0.1

35	trans-4-propenylsyringol	26624135	25.72	S, misc.	C_γ	194	194.09373	1.2	1.2	1.2	1.3
36	dihydroconiferyl alcohol	2305137	25.81	S, C_γ-O	C_γ	182	137.05971	0.1	0.2	0.2	0.2
37	syringaldehyde	134963	26.34	S, C_α-O	C_α	182	182.05736	0.9	0.9	0.9	0.9
38	cis-coniferyl-alcohol	458355	26.42	G, C_γ-O	C_γ	180	137.05971	1.1	1.2	1.2	1.2
39	homosyringaldehyde	-	27.32	S, C_β-O	C_β	196	167.07027	1.1	1.1	0.9	1.0
40	syringic acid methyl ester	884355	27.66	S, C_α-O	C_α	212	212.06793	0.02	0.02	0.02	0.02
41	acetosyringone	2478388	27.76	S, C_α-O	C_β	196	181.04954	0.7	0.7	0.6	0.7
42	trans-coniferyl alcohol	458355	28.11	G, C_γ-O	C_γ	180	137.05971	21.1	24.8	25.8	25.6
43	trans-coniferaldehyde	458366	28.50	G, C_γ-O	C_γ	178	147.04406	1.4	1.4	1.4	1.4
44	syringylacetone	19037582	28.68	S, C_β-O	C_γ	210	167.07027	0.4	0.4	0.4	0.4
45	propiosyringone	5650431	29.29	S, C_α-O	C_γ	210	181.04954	0.05	0.05	0.05	0.05
46	syringyl diketone	6925651	29.43	S, C_α-O, C_β-O	C_γ	224	181.04954	0.3	0.3	0.2	0.3
47	syringyl vinyl ketone	-	29.57	S, C_α-O	C_γ	208	181.04954	0.03	0.03	0.03	0.03
48	dihydrosinapyl alcohol	20736258	31.13	G, C_γ-O	C_γ	212	168.07841	0.1	0.1	0.1	0.1
49	cis-sinapyl alcohol	537337	31.63	S, C_γ-O	C_γ	210	167.07027	1.0	1.0	1.0	0.9
50	trans-sinapyl alcohol	537337	33.31	S, C_γ-O	C_γ	210	167.07027	15.6	15.9	15.3	14.5
51	trans-sinapaldehyde	4206580	33.54	S, C_γ-O	C_γ	208	208.07301	1.3	1.3	1.1	1.2

IV.

METAGENOME-ASSEMBLED GENOMES FROM A TERMITE-DERIVED LIGNOCELLULOLYTIC MICROBIAL CONSORTIUM PROVIDE NEW INSIGHTS INTO THE ANAEROBIC DEGRADATION OF BIOMASS

IV.1. METAGENOME-ASSEMBLED GENOMES FROM A TERMITE-DERIVED LIGNOCELLULOLYTIC MICROBIAL CONSORTIUM PROVIDE NEW INSIGHTS INTO THE ANAEROBIC DEGRADATION OF BIOMASS

*Louison Dumond[1], Claire Hoede[2], Vincent Lombard[3,4,5], Bernard Henrissat[3,4,5], Guillermina Hernandez-Raquet[1]**

1 Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.

2 INRAE, UR 875 Unité de Mathématique et Informatique Appliquées, Platform GenoToul Bioinfo, Auzeville, CS 52627, 31326 Castanet Tolosan cedex, France

3 CNRS UMR 7257, Aix-Marseille University, F-13288, Marseille, France.

4 USC 1408 AFMB, INRAE, F-13288, Marseille, France.

5 Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

DUMOND Louison dumond@insa-toulouse.fr

HOEDE Claire claire.hoede@inrae.fr

LOMBARD Vincent vincent.lombard@afmb.univ-mrs.fr

HENRISSAT Bernard bernard.henrissat@afmb.univ-mrs.fr

*Corresponding author: Guillermina Hernandez-Raquet, Toulouse Biotechnology Institute - TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, F-31077 Toulouse, France. Tel: + 33 (0) 561 55 99 77. Fax: + 33 (0) 561 55 97 60. Email: hernandg@insa-toulouse.fr

IV.2. INTRODUCTION

Lignocellulose is the main constituent of plant cell walls making it the most abundant biomass on Earth. Lignocellulose is a renewable non-food carbon source that can be obtained from agricultural, urban and forestry residues that represents the most promising biorefinery feedstock for biofuels and biomaterials production. It is composed by three main polymers, cellulose (30-50%) and hemicellulose (20-30%) which are polysaccharides, and lignin, which is an aromatic polymer (15-30%). These three fractions are elaborately interwoven with each other to form a complex three-dimensional structure that confers rigidity to plant cell walls and renders lignocellulose highly recalcitrant to the microbial attack (Lapierre, et al., 1995; del Río et al. 2012).

In Nature, lignocellulose recycling involves of a broad array of carbohydrate active enzymes (CAZymes) (Cantarel et al. 2009) covering a variety of activities, including glycoside hydrolases (GHs), glycoside transferases (GTs), carbohydrate esterases (CEs), polysaccharide lyases (PL) and auxiliary enzymes (AA) that modify, synthesize or cleavage the various linkages found in biomass. There are also no-catalytic substrate-binding proteins (carbohydrate-binding modules, CBMs), as well as laccases and peroxidases that acts on the lignin fraction (Cragg et al. 2015). Fungi, bacteria and protozoa are the main responsible of the production of such wide diversity of enzymes (Cragg et al. 2015).

In herbivorous animals and xylophagous insects, access to the lignocellulosic biomass highly depends on the ability of their gut microbiome to degrade plant polysaccharides (Zilber-Rosenberg et Rosenberg 2008; Cragg et al. 2015). The termite microbiome is of particular interest since termites are among the most efficient lignocellulose degraders playing an important role in the decomposition of plant material and global carbon recycling (König et al.? 2013).

All termite species harbor in their gut microbial symbionts that transform biomass into carboxylates or short chain volatile fatty acids (VFA) such as acetate, that are used as carbon and energy sources by the termite host (Godon et al. 2013; Andreas Brune 2014). In return, the microbiome benefits from the stable gut environment. This system may be assimilated to highly efficient lignocellulolytic microscale bioreactors in which the microbiome is responsible of biomass digestion and product formation (Godon et al. 2013; Andreas Brune 2014). The gut

microbiome of higher termites, such as the wood-feeding subfamily Nasutermitinae, is free from flagellate protozoa and exclusively composed by prokaryotes. Although lignocellulose digestion involves cellulases expressed by the termite tissue (Tokuda et al. 1999), plant biomass is mainly digested by the rich and complex enzymatic cocktail produced by this bacterial gut microbiome (Brune et al., 1995; Nakashima et al. 2002).

In the last ten years, strong efforts have been realized to understand the *in vivo* functioning of the gut microbiome of higher termites, including *Nasutitermes* sp., and its relationship with the termite host (Warnecke et al. 2007; Burnum et al. 2011). Applying metagenomics, metatranscriptomics and metaproteomics, several studies discovered the rich bacterial diversity and the vast repertoire of CAZymes encoded by the symbiotic bacteria associated to these digestive system and their response to changes in diet composition (Calusinska et al. 2020). However, the large complexity found in these gut ecosystems did not enable to reconstruct microbial genomes and identify the specific functional role the different bacterial strains play during plant biomass transformation. A powerful strategy to overcome this problem is to enrich the functionally-relevant fraction of the microbial community by culture under defined conditions (DeAngelis et al. 2010; Delmont et al. 2015). Such approach enable to minimize the ecosystem's complexity and increase the community fraction expressing the lignocellulolytic potential.

In our previous *in vitro* studies, using controlled bioreactors, we have established that the lignocellulolytic potential of termite gut microbiomes can be enriched in anaerobic conditions to exploit them for transforming lignocellulosic biomass into, for instance, VFAs products. To build upon these premises, we hereafter aim to understand the role played by the different community members involved in the anaerobic lignocellulose conversion into VFAs, and decipher their temporal dynamics along biomass degradation. To reach this goal we have sequenced the whole metagenome of a microbial consortium derived from *N. ephratae* gut microbiome enriched for its wheat straw transformation capacity. By reconstructing metagenomes assembled genomes (MAGs) of the most abundant species of this consortium, the annotation of the COGs related to biomass conversion and VFA production, and the actual CAZymes content in the reconstructed MAGs, we established the role and interplay of the key microbial genomes involved in the dynamic process of wheat straw transformation.

IV.3. MATERIAL AND METHODS

IV.3.1. Lignocellulose degradation experiments

A termite-derived microbial consortium (TWS) that has been enriched by sequential batch cultivation on wheat straw (Lazuka et al., 2018) was used as inoculum for lignocellulose degradation experiments in anaerobic bioreactors. These experiments were carried out in biological duplicate anaerobic bioreactors (Biostat B, Sartorius) using wheat straw (20g.L^{-1}) as sole carbon source in a mineral medium (MM), at pH of 6.15 (controlled by NaOH 1N addition) and 35°C, and inoculated with TWS (10% v/v) (Lazuka et al. 2018). The MM contained per liter of distilled water: KH_2PO_4, 0.45 g, K_2HPO_4, 0.45 g; NH_4Cl, 0.4 g; NaCl, 0.9 g; $MgCl_2 \cdot 6H_2O$, 0.15 g; $CaCl_2 \cdot 2H_2O$, 0.09 g. It was supplemented with 250 µL of V7 vitamin solution and 1 mL trace elements solution, containing per liter of distilled water: H_3BO_3, 300 mg; $FeSO_4 \cdot 7H_2O$, 1.1 g; $CoCl_2 \cdot 6H_2O$, 190 mg; $MnCl_2 \cdot 4H_2O$, 50 mg; $ZnCl_2$, 42 mg; $NiCl_2 \cdot 6H_2O$, 24 mg; $NaMoO_4 \cdot 2H_2O$, 18 mg; $CuCl_2 \cdot 2H_2O$, 2 mg; sterilized by filtration (0.2 µm). Methanogenic activity was inhibited by adding brome ethane sulfonate (BES, up to 10 mM). The bioreactors were flushed with nitrogen after inoculation to ensure anaerobic conditions. The replicate bioreactors are therein referred as ST1 (kinetics 1) and ST2 (kinetics 2). Samples were withdrawn at regular intervals throughout the incubation time (Tx indicate the incubation time; x is the number of day) to characterize the lignocellulose degradation and the microbial community dynamics.

IV.3.2. Chemical analysis

All chemical analyses were presented in a previous work (Lazuka et al., 2018). Briefly, wheat straw concentration was determined at each sampling point by measuring the total (TS) and volatile (VS) solids, corresponding to the solid organic fraction. Wheat straw composition was determined using the sulfuric acid hydrolysis method described by Lazuka et al. The insoluble residue was washed with distilled water and dried at 105 °C overnight to determine Klason lignin content (Hatfield et al. 1994). Sugar composition was determined on an Ultimate 3000 Dionex HPLC with refractive index detector (Thermo Scientific) equipped with a BioRad Aminex HPX 87H affinity column.

Volatile fatty acids (VFA) contained in the samples supernatant were analyzed by gas chromatography using a Varian 3900 gas chromatograph equipped with a flame ionization detector and CP-Wax 58 (FFAP) CB column as described previously (Cavaillé et al. 2013).

IV.3.3. Enzymatic activities

Five ml samples were used for enzyme activity measurement. The samples were centrifuged and xylanase and endoglucanase (CMCase) activities were measured using 1% w/v xylan beechwood (Sigma) and 1% w/v carboxymethyl cellulose (CMC) (Sigma) on both the supernatant and the sonicated pellet (Lazuka et al., 2018). One unit of CMCase or xylanase activity (UA, unit of activity) was defined as the amount of enzyme that produces 1 µmol of reducing sugars per minute.

IV.3.4. DNA extraction

DNA was extracted from samples (1.5 ml) taken from the duplicate bioreactors (6 samples each) throughout the incubation time. Samples were centrifuged (13,000 x g, 5min, 4°C), the supernatant was removed and the pellet was snap frozen in liquid nitrogen and stored at -80°C until DNA extraction. Total DNA was extracted using an RNeasy PowerMicrobiome kit (Qiagen) following the manufacturer's instruction but omitting the final DNAse step. Cell lysis was performed with a FastPrep (MP Biomedicals) (2 x 30s at $4 m.s^{-1}$). DNA was purified using an AllPrep DNA/RNA Mini kit (Qiagen) following the manufacturer's instructions. DNA concentration was measured by Nanodrop 1000 spectrophotometer (ThermoScientific) measuring absorbance at 260 nm and 280 nm. DNA purity was assessed by electrophoresis on a 0.8% agarose gel. For shotgun metagenomics, the purified DNA was also quantified by Qubit fluorimetric measurement (ThermoScientific).

IV.3.5. 16S rRNA metabarcoding sequencing and data analysis

Illumina sequencing of the V3-V4 region of 16S rRNA genes was performed after PCR amplification using the bacterial primers 343F and 784R, modified to add sequencing adaptors during a second PCR (343F = 5'-CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC GGR AGG CAG CAG-3' , 748R=5'-GGA GTT CAG ACG TGT GCT CTT CCG ATC TTA CCA

GGG TAT CTA ATC CT-3') (Auer et al. 2017). Library preparation was performed as previously detailed (Auer et al., 2017) and loaded on a MiSeq Illumina cartridge using reagent kit v3 to obtain paired-end 300 base pair (bp) reads. 16S rRNA gene sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France) and generated between 13,000 to 32,000 paired-reads per sample (20,500 on average) that were analyzed using the FROGS pipeline (Find Rapidly OTU with Galaxy Solution) implemented on a Galaxy instance (http://sigenae-workbench.toulouse.inra.fr/galaxy/) (Escudié et al. 2017). In summary, paired reads were merged using FLASH (Magoč et Salzberg 2011); primers and adapters were removed with cutadapt and a de novo clustering was performed using SWARM (Mahé et al. 2015) with an aggregation distance $d = 3$ after denoising using $d = 1$. Chimera were removed using VSEARCH (Rognes et al. 2016). Singletons were filtered out from the dataset and affiliation was done using Blast+ against the Silva database (release 132). In agreement with recommendations (Bokulich et al. 2013), in order to avoid false OTUs, an OTU abundance threshold of 0.005% of the total sequences was applied. OTU tables were then formatted in standard BIOM for data analyses. For diversity analysis, OTU tables, taxonomic affiliations and bioreactors metadata were imported into Phyloseq v1.26.1 to analyze diversity (McMurdie et Holmes 2013b) using R v3.5.3 (R core team, 2019).

IV.3.6. Shotgun metagenomics sequencing, de novo assembly and annotation

For shotgun metagenomics, DNA libraries for twelve individual samples were prepared following Illumina HiSeq instructions using True Seq DNA PCR-free (Illumina, San Diego, CA, USA). Libraries were pooled in identical quantities and paired-end sequenced using one line of the Illumina HiSeq 3000 platform to obtain 150 bp reads. Shotgun sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France).

Shotgun sequencing resulted in 335M paired reads in the dataset (range: 23-31 million per sample) (Supp. Table 1). All samples were submitted to quality control using FastQC (Andrews, 2010). As all samples displayed a minimal Phred quality score of 20, no remaining adapters and the desired length of 150bp, they were all kept for genome reconstruction.

For genome reconstruction, reads form all samples were used for *de novo* co-assembly using MEGAHIT v1.1.3 with default parameters (D. Li et al. 2016). Contigs were filtered using minimum contig length of 1500 bp and minimal abundance of one RPKM (reads per kilobase

per million mapped reads) on at least two different samples. It resulted in 33,071 filtered contigs (size ranging from 200 to 1.24 Mbp, N50=5393). that were used to generate a contigs database using Anvi'o v5.4 (Eren et al. 2015).

Assembly statistics were obtained with Assemblathon.pl. Paired-end reads were mapped onto the contigs using BWA-MEM v0.7.11 with default parameters (Heng Li 2013) and mapping statistics were compiled using the Samtools suite v1.8 (Heng Li et al. 2009).

Open reading frames (ORFs) were searched on the filtered contigs and structural genes were annotated from the assembly using Prodigal v2.6.3 run in a metagenomics mode (-p meta) (Hyatt et al. 2010). Functional annotation of predicted protein was done with DIAMOND v. 0.9.22 in mode BLASTP against NR database (Buchfink et al., 2015).

Taxonomic affiliation of the contigs was performed with CAT (Contig Annotation Tool) v5.0. (von Meijenfeldt et al. 2019) using –r=10 and –f=0.3 as criteria.

Diversity of the metagenome was determined from 16S rRNA sequences extracted from the contigs with Metaxa2 (Bengtsson-Palme et al. 2015) using default parameters, annotated with BLAST 16S ribosomal RNA sequence database (August 2020 update).

IV.3.6.1. Metagenome assembled genome reconstruction and population dynamics

For MAGs reconstruction, the filtered contigs were binned based on their tetranucleotide composition using CONCOCT v1.0.0 (Alneberg et al. 2014). Subsequently, all bins were manually curated using the interactive tool integrated in Anvi'o. Single core genes (SCG) were used to assess the completeness and contamination of the resulting MAGs using Anvi'o v5.4.

In order to improve the assembly of a highly abundant MAG, a single sample (ST1T1) was also separately assembled using MetaSpades v3.11.1 (Nurk et al. 2017) and MAGs were generated with Anvi'o v5.4 using the same procedure. 126 MAGs were generated this way. Duplicated MAGs obtained from the two assembly methods were dereplicated by mapping the contigs from MAGs obtained from the MetaSpades assembly onto the contigs of the MegaHit co-assembly with BWA-mem. For each MAG onto which more than 50% of the contigs were mapped, the one with a higher completion was kept for further analysis.

According to a proposed quality criteria (Bowers et al. 2017) , only high quality (>90% completeness, <5% redundancy) and medium quality (≥50% completeness, <10% redundancy)

bins were retained for further analysis. Bins with size length > 2Mb were also included in the analysis according to the Anvio's recommendations (Delmont et al. 2015). Only bins fulfilling those quality conditions were considered as MAGs.

To assess the temporal dynamic of MAGs in the lignocellulose degradation experiments, the relative abundances of the recovered MAGs at each sampling point was estimated by querying the raw reads of each sampling point against to the database constructed with all MAGs. Unmapped reads and reads mapped to more than one location were removed with samtools view (-q =10). The relative abundance of each MAG was then computed as the fraction of reads in each sample mapping to the respective MAG normalized on the bin size (fraction of reads per nucleotide in bin) and on the total number of reads in the corresponding sample.

The abundances of the functional traits of MAGs along incubation time were explored and statistical differences between various traits were assessed using the R package Vegan v2.5.6 (Oksanen, 2019)

IV.3.6.2. Taxonomic and phylogenomic analysis

Taxonomic affiliation of the MAGs was performed with Bin Annotation Tool v5.0.3 (BAT) (von Meijenfeldt et al. 2019) using –r=10 –f=0.3 as criteria.

A MAGs phylogenomic tree was constructed with Phylophlan3 using the corresponding database of proteins and parameters set to accurate/high diversity (Segata et al. 2013). The best consensus tree over 100 bootstraps was obtained from RAxML-ng v0.9.0 (Kozlov et al., 2019). The phylogenomic tree was drawn using the Interactive Tree Of Life (iTOL v5) (Letunic et Bork 2019).

IV.3.6.3. Functional annotation of MAGs: COGs, CAZymes and PULs

Clusters of Orthologous Genes (COGs) were identified on the predicted proteins using eggnog-mapper on eggNOG 5.0 database (Huerta-Cepas et al. 2019). COGs related with VFA-production were specifically extracted from the COGs dataset to be analyzed (Supplementary data 3).

CAZyme annotations were performed on the identified. A custom script which aims at accommodating the modularity of CAZymes was used to compare sequences to the full-length

sequences stored in the CAZy database using BlastP (version 2.3.0+). If sequences are aligned with a sequence already in the CAZy database with 100% coverage, >50% amino acid sequence identity and E-value ≤ 10−6, they are assigned to the same family or subfamily (or the same families if the found sequence contains more than one CAZy module). If not, the sequences have to undergo a similarity search that involves two steps. On the one hand, using BlastP, they are compared to a library of modules (instead of full sequences that can contains multiple modules). On the other hand, a HMMER3 search against a curated collection of hidden Markov models based on each of the CAZy module is performed. The script assigns a family when both BlastP E-value<10^{-4} and it aligns with >90% overlap.

SusC and SusD genes were identified by their COGs. PULs were predicted in the MAGs by the presence of a tandem SusC-SusD genes on a contig.

Count of COGs, CAZymes and PULs were estimated for each MAG.

IV.3.7. Statistical analysis

Data analysis on the metagenomics dataset was performed on R v3.5.3 (R core team, 2019). Hierarchical clustering and heatmaps were done using Vegan v2.5.6 (Oksanen, 2019). Principal component analysis (PCAs) and Partial Least Square Discriminant Analysis (PLSDAs) were done with the MixOmics v6.6.2 (Rohart et al. 2017). All analysis were performed using the Tidyverse v1.3.0 (Wickham et al., 2019).

IV.4. RESULTS

IV.4.1. Anaerobic wheat straw degradation by TWS

The dynamics of wheat straw degradation by TWS, including cellulose and hemicellulose degradation, the production of VFAs and the activity of major lignocellulolytic enzymes is shown in Figure 1. After 8 days of incubation, the consortium activity reached a plateau for VS degradation, VFA production and enzyme activities. TWS achieved an average of $47.8 \pm 2.9\%$ VS degradation and a production of 188.6 ± 3.1 mCmol VFA L^{-1} (Figure 1.a). After two days, VFA production and VS degradation followed the same trend.

TWS displayed a high cellulose and hemicellulose degradation, reaching 59% degradation of cellulose and 51% degradation of hemicellulose in 11 days of incubation Very shortly after incubation, both cellulose and hemicellulose were found to be degraded in the bioreactor. After respectively 3 and 5 days, the degradation rate of hemicellulose and cellulose decreased until reaching a plateau. Interestingly, it was found that hemicellulose degradation rate reached its peak (29.6% Cmol L^{-1} day^{-1}) after only a single day, predating cellulose degradation (Figure 1.b). Overall, it seems that TWS preferentially degraded hemicellulose in the first 5 days of incubations.

Both CMCase and xylanase activities followed a similar trend (Figure 1.c). After a 2 days lag phase, enzymatic activity sharply increased before reaching a plateau after 5 days of incubation. It is interesting to note that, while the peak activity was reached after 3 days, the peak of hemicellulose and cellulose degradation was already reached at this point in the incubation process.

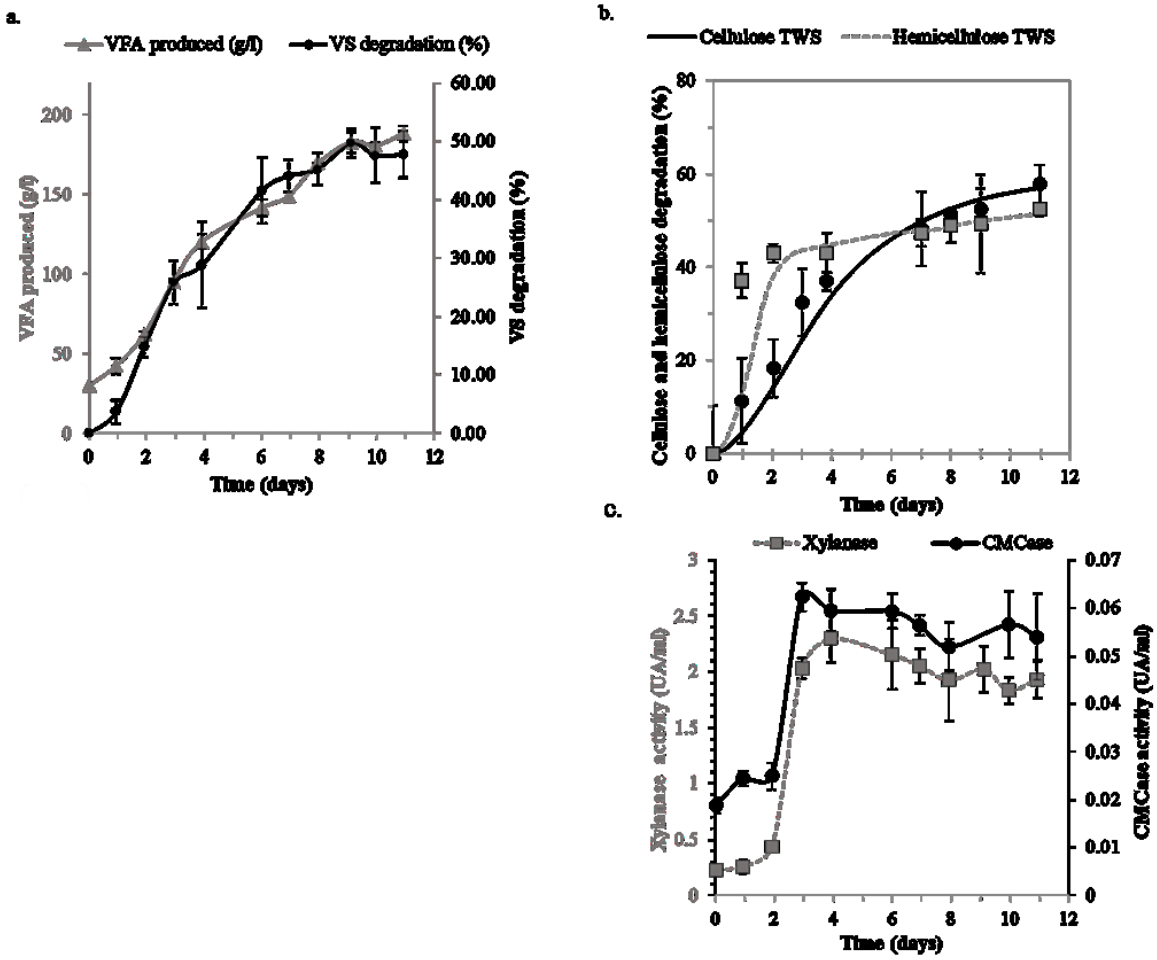


Figure 1: a. Average VS degradation (%) and VFA production (mCmol VFA L^{-1}) and

sampling times for the two kinetics b. Cellulose and hemicellulose degradation of wheat straw by TWS. **c.** CMCase and Xylanase activity of TWS.

IV.4.2. Metagenome of the lignocellulolytic TWS consortium

Based on the wheat straw degradation kinetics, 6 sampling times (indicated Figure 1.a) for each duplicate bioreactor were selected for further metagenomic characterization by shotgun sequencing.

Co-assembly of shotgun sequencing data of all TWS samples followed by contigs filtering, resulted 33,071 filtered contigs (size ranging from 200 to 1.24 Mbp, N50=5393). The length of the largest contig was 1.25 Mbp and 420 contigs were longer than 100 kbp. Reads mapping to the assembled contigs resulted in 90% reads incorporation.

IV.4.3. Taxonomic analysis of TWS metagenome based on assembled contigs

To decipher the microbial community composition of TWS, the taxonomic affiliation of assembled contigs, based both on identified proteins or the taxonomy extracted with MetaTaxa2 was compared to the community composition deduced from 16S rRNA gene amplicon sequencing (**Figure 2**). By all methods, it can be noticed that TWS consortium was constituted quasi exclusively of bacteria accordingly to the prokaryotic composition of *N. ephratae* gut. Since the methanogenesis was inhibited by the addition BES, no archeae were found excepted one low quality MAG affiliated to *Methanobacterium* sp. The prevalence of main phyla was consistent between the different affiliation methods with a domination of members belonging to Bacteroidetes (62-81%) followed by Firmicutes (13-17%) and Proteobacteria (4-11%) though Bacteroidetes were overrepresented in 16S rRNA aamplicon sequencing. Nevertheless, other phyla such as Spirochaetes have got inconsistencies between 16S rRNA and whole genome analysis. While Spirochaetes represented up to 9% of the abundance in the whole metagenome, they were not detected by 16S rRNA amplicon sequencing. This discrepancy could be attributed to a bias of the PCR universal primers used for amplification of the V3-V4 region as they covered 91.4% of Sphaerochaeta genus, sole Spirochaetes representative observed in shotgun metagenomics, of the Silva database (Campanaro et al. 2018) but it could due to the fact that Spirochaetes are difficult to separate based on their 16S rRNA sequence

(Hallmaier-Wacker et al. 2019). Actinobacteria, Synergistetes and Lentispherae were also found within the community at variable but low abundances (<5%) using each affiliation method. At the family level, the prevalent families were Bacteroidaceae (55%-73%) followed by Clostridiaceae (3%-11%) and Enterobacteriaceae (4%-12%); such used. In addition to differences in *Spirochaetaceae* distribution, the repartition of Firmicutes differed between 16S rRNA sequencing and other methods as *Clostriaceae* were majoritarian in the whole metagenome while *Lachnospiraceae* were with amplicon sequencing. Barring this exception, distribution was consistent between the different affiliation methods.

A few bacteria amounted to the majority of the total abundance in the metagenome. More specifically, *Bacteroides graminisolvens* represented up to 70% of the total abundance.

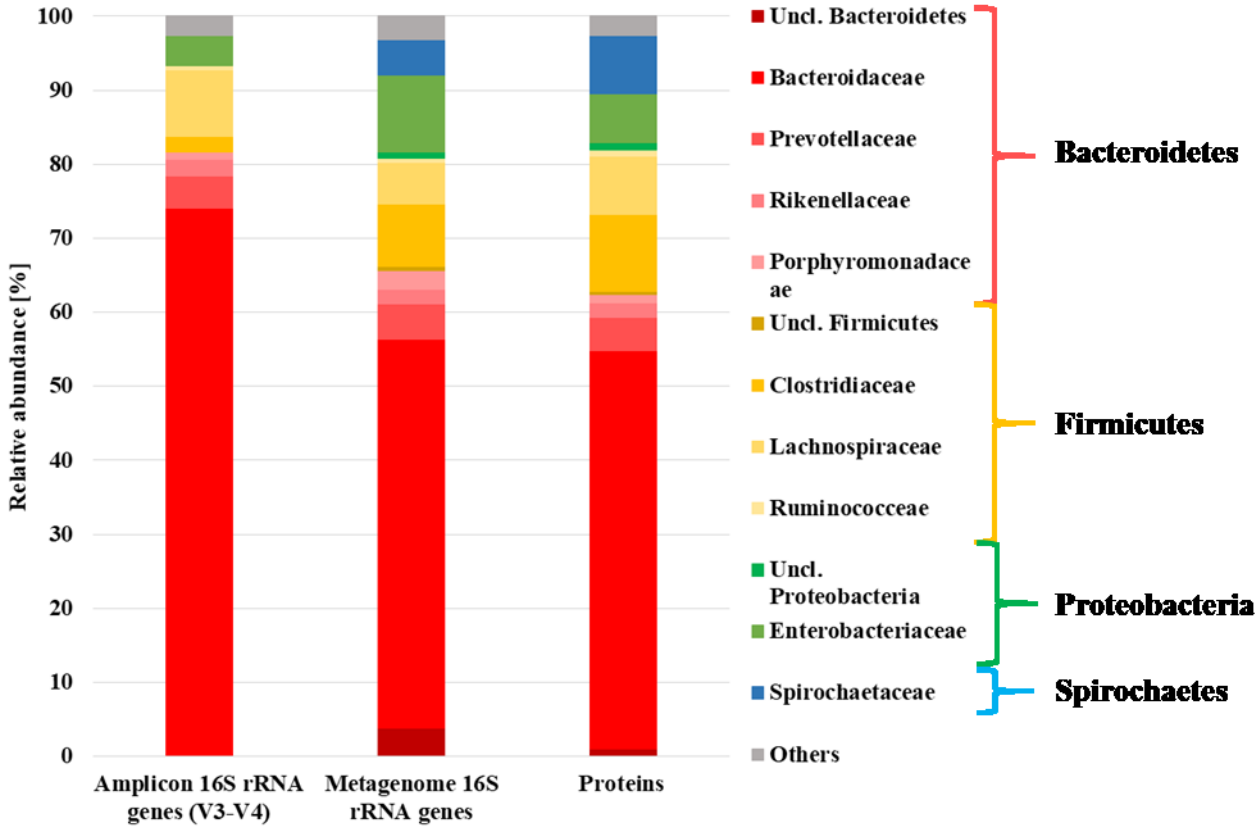


Figure 2: Comparison of the taxonomic composition of TWS metagenome according to amplicon 16S rRNA sequencing, Metaxa2-extracted 16S rRNA genes and the predicted proteins in assembled contigs.

IV.4.4. The consortium diversity evolved over time

In order to understand the dynamic of TWS along lignocellulose degradation, we analyzed the evolution of 16S rRNA gene sequencing data and MAGs taxonomy along the incubation time (**Figure 3**).

Both bioreactors followed the same trend over time with an increase in Bacteroidetes and Firmicutes abundance during the first 4 days. They were replaced by Spirochaetes over time which abundance increased from 3 to 25% after 13 days of degradation for ST2. In both bioreactors, Proteobacteria were quickly depleted.

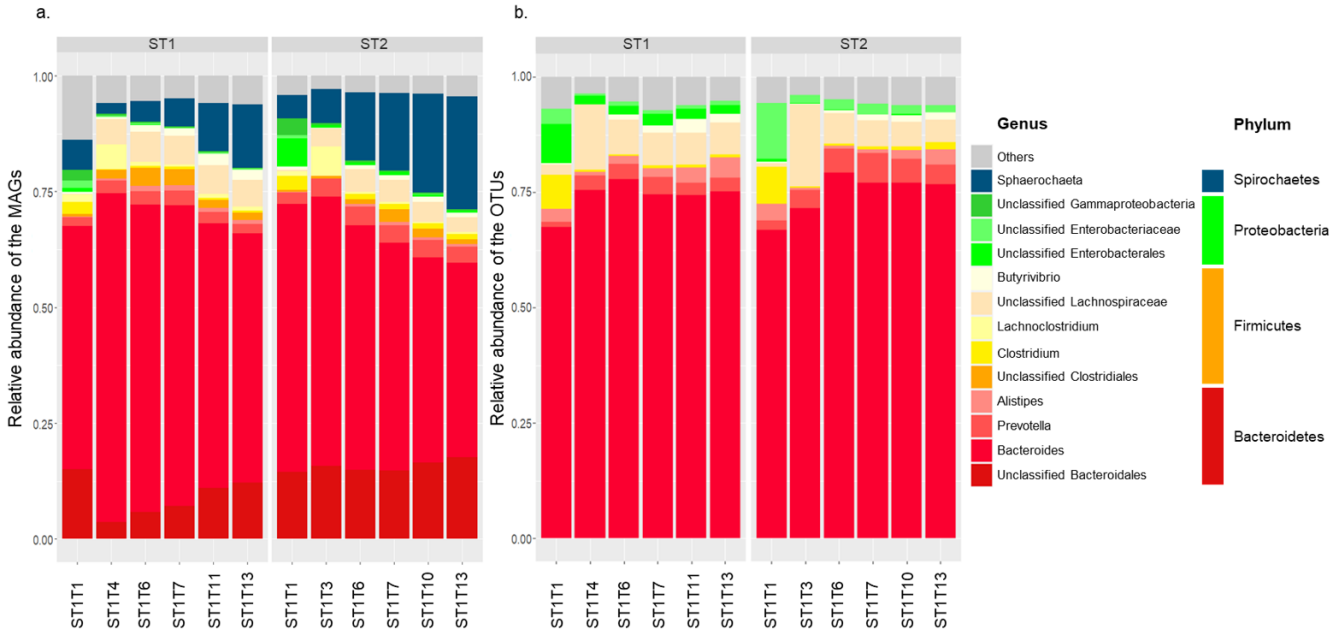


Figure 3: Evolution of TWS community dynamics over time; at genus and phylum level. Description based on **a.** Taxonomic affiliation of MAGs and **b.** Taxonomic affiliation based on 16S rRNA gene data.

37 different genera were observed throughout the degradation however only 7 genera composed at least 1% of the total diversity at one point. Bacteroidetes were largely constituted of Bacteroides (85-95% of Bacteroidetes) with Prevotella and Alistpes being the other main genera. The Firmicutes were mainly composed by Clostridiaceae (mainly *Clostridium*) after one day of degradation (>60% of the Firmicutes MAGs, >80% of the Firmicutes OTUs) but were quickly replaced by the Lachnospiraceae (mainly *Lachnoclostrdium*) and Ruminococceae families (mainly *Butyrivibrio*) afterwards. Proteobacteria were mainly unclassified

Enterobacterales (though affiliated to the *Escherichia-Shigella* genera in 16S). At last Spirochaetes were almost exclusively from the Sphaerochaeta genus.

IV.4.5. Metagenome functional analysis

To gain insight on the metabolic processes of TWS metagenome, the sequences of predicted proteins found in the assembled contigs were functionally classified according to the COG database. From the 378,374 initial predicted proteins, 282,940 (75%) were affiliated to a COG, 213,230 of which (75 %) had known COG functions. The predicted proteins in TWS metagenome were linked to the metabolism (C, E, F, G, H, I, P, Q), information and storage (A, B, J, K, L) and cellular process and signal (D, M, N, O, T, U, V) (Tatusov et al. 2000). About 10.2% of the predicted proteins were associated to transcription process (K), 9.8% to amino acid transport and metabolism (E), 9.4% to carbohydrate, transport and metabolism (G), 7.9% to energy production (C) and 3.5% to lipid transport and metabolism (I) (**Figure 4.a**). A very low abundance of proteins COGs related to RNA processing and modification (A category) and chromatin, structure and biogenesis (B category) were also noticed.

The COG classification of TWS predicted proteins displayed a different profile according to their taxonomic affiliation. On the one hand, Lentispherae, Bacteroidetes, Firmicutes and Spirochaetes metabolism appeared skewed towards carbohydrate transport and metabolism as can be noticed from COG G being one of most abundant COG in their genome **(Figure 4.b.).** Bacteroidetes also displayed a high number of proteins belonging to COGs linked to cell wall/membrane biogenesis (M) while Firmicutes also displayed a high proportion of COGs implicated in transcription (K). These are all important COG categories to consider for lignocellulose degradation. On the other hand, Proteobacteria, Actinobacteria and Synergistetes displayed a lower proportion of predicted proteins involved in carbohydrate transport and metabolism (G) but a comparatively higher proportion of those involved in amino acid transport and metabolism (E COG) or in energy production and metabolism (C COG). This suggest that these phyla may not be directly involved in lignocellulose degradation but using the resulting polysaccharides for their growth. It should be noted that proteins linked to VFAs production are found in the C and E categories. This could imply Proteobacteria, Actinobacteria and Synergistetes may be involved in VFA production.

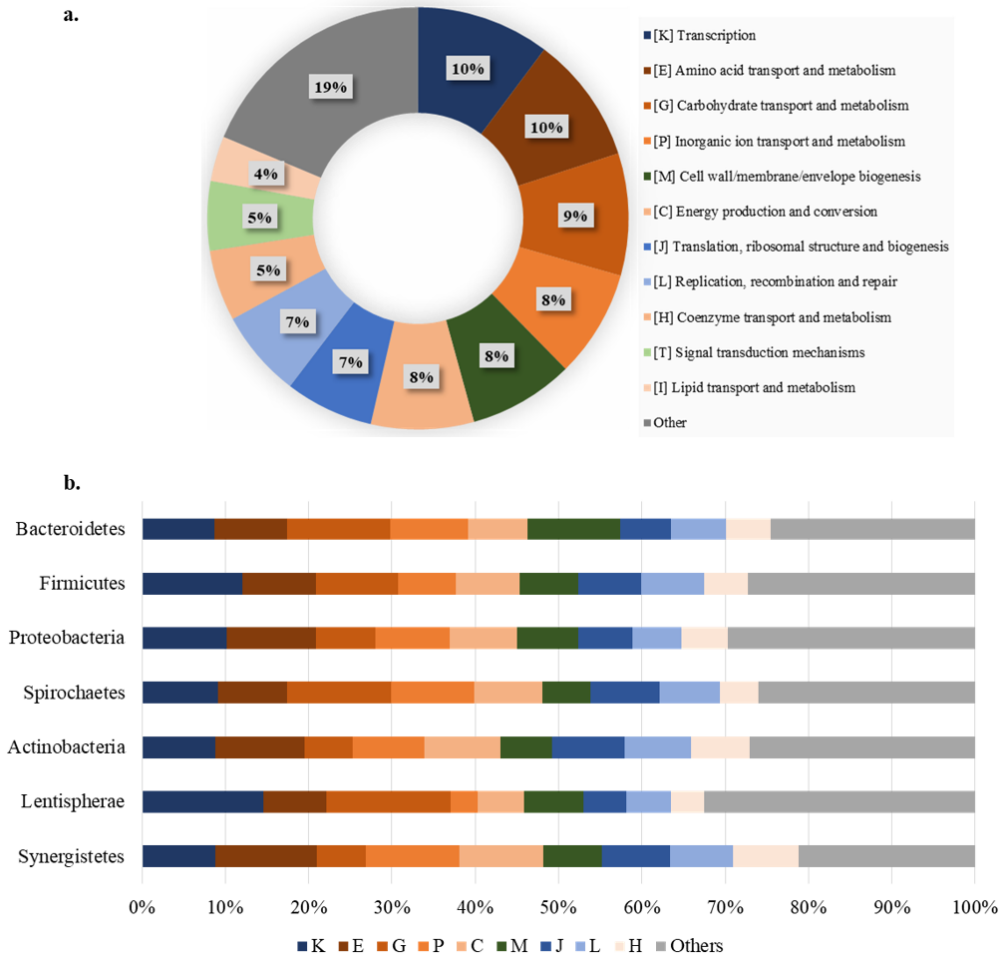


Figure 4: a. Distribution of the COGs families in the metagenome (% COGs), blue color represents the information, storage and processing category, orange the metabolism category and green the cellular process and signal category.

b. Relative abundance of COGs families in each phylum. The nine overall most abundant COGs are represented, all other COGs are added in others.

IV.4.6. CAZymes and VFA related activities

TWS metagenome encoded a large diversity of CAZymes (17,001) and proteins related to VFA production (**Table 1**) which is consistent with the capacity of TWS to transform lignocellulose into VFA. The predicted CAZymes in TWS were mainly classified in the families GH (53 %) followed by GT (27%) and CBM (12%). Proteins AA associated to lignin degradation were detected in a very low proportion (0.3%) which is consistent with the non-lignin degradation

activity observed in TWS (Lauka et al., 2018). A very low number of cohesin and dockerin were also found (1 and 22 respectively)

Table 1: Number of CAZY modules predicted in the whole metagenome, in the 99 selected MAGs and in the CAZy database at the time of submission (March 2020). * % of metagenome CAZymes recovered in MAGs

CAZY Modules	**Whole Metagenome**	**MAGs**
AA	62	45 *(73%)**
CBM	1961	1160 *(59%)**
CE	987	651 *(66%)**
GH	8967	6334 *(71%)**
GT	4637	2843 *(61%)**
PL	387	321 *(83%)**

IV.4.7. Recovering of metagenome assembled genomes (MAGs) from the lignocellulolytic TWS consortium

Co-assembly, CONCOCT binning and manual curation resulted in a dataset comprised 144 MAGs. Additionally, the two most abundant species in the bioreactors identified by 16S rRNA amplicon sequencing (affiliated to *Bacteroides graminisolvens* and *Bacteroides uniformis,* see discussion above), which reconstruction was not achieved in the co-assembly, were successfully reconstructed by MetaSpades assembly of a single sample (ST1T1) and added to the final dataset as well as twelve non redundant bins reconstructed by the same approach. According to Bowers's criteria (Bowers et al., 2017), the dataset comprised 62 MAGs classified as medium quality and 30 high quality MAGs. These 92 MAGs and 7 additional MAGs with size> 2Mb were considered for further analysis. A complete summary of these 99 MAGs is available in the Supplementary file 1.

A phylogenomic tree of the 99 recovered MAGs showed that they were mainly affiliated to Firmicutes, Bacteroidetes and Proteobacteria, according to the overall taxonomic affiliation of TWS metagenome discussed above (**Figure 5**).

47 MAGs belonged to Firmicutes phylum making it the phylum with the highest number of representatives. This phyla was dominated by the Clostridiaceae, Lachnospiraceae and Ruminococceae families. However, with the exception of MAG_66, affiliated to unclassified Lachnospiraceae, most Firmicutes-related MAGs have a low abundance in TWS metagenome. The second most represented phylum was Bacteroidetes with 28 MAGs displaying the highest abundance in the metagenome. This phyla was represented by Bacteroidaceae, Prevotellaceae and Rikenellaceae families. In particular, MAG_91, affiliated to *Bacteroides graminisolvens*, was the most abundant MAG in the dataset representing half the total abundance of MAGs in TWS consortium. The other MAGs found in TWS were affiliated to Proteobacteria (14 MAGs), belonging to the Enterobacteriaceae family, followed by Actinobacteria (5 MAGs) of three different families (Corynebacteriaceae, Eggerthellaceae and Bifidobacteriaceae), Spirochaetes (4 MAGs including MAG_90) and Lentisphearae (1 MAG).

The low bootstrap values at the tree root did not enabled to compare the proximity between these different MAGs. Nevertheless, the tree successfully grouped the MAGs according to their taxonomy and allowed us to resolve the taxonomy of two MAGs (MAG_10 to *Lachnospiraceae Uncl.* and MAG_31 to γ-Proteobacteria Uncl.).

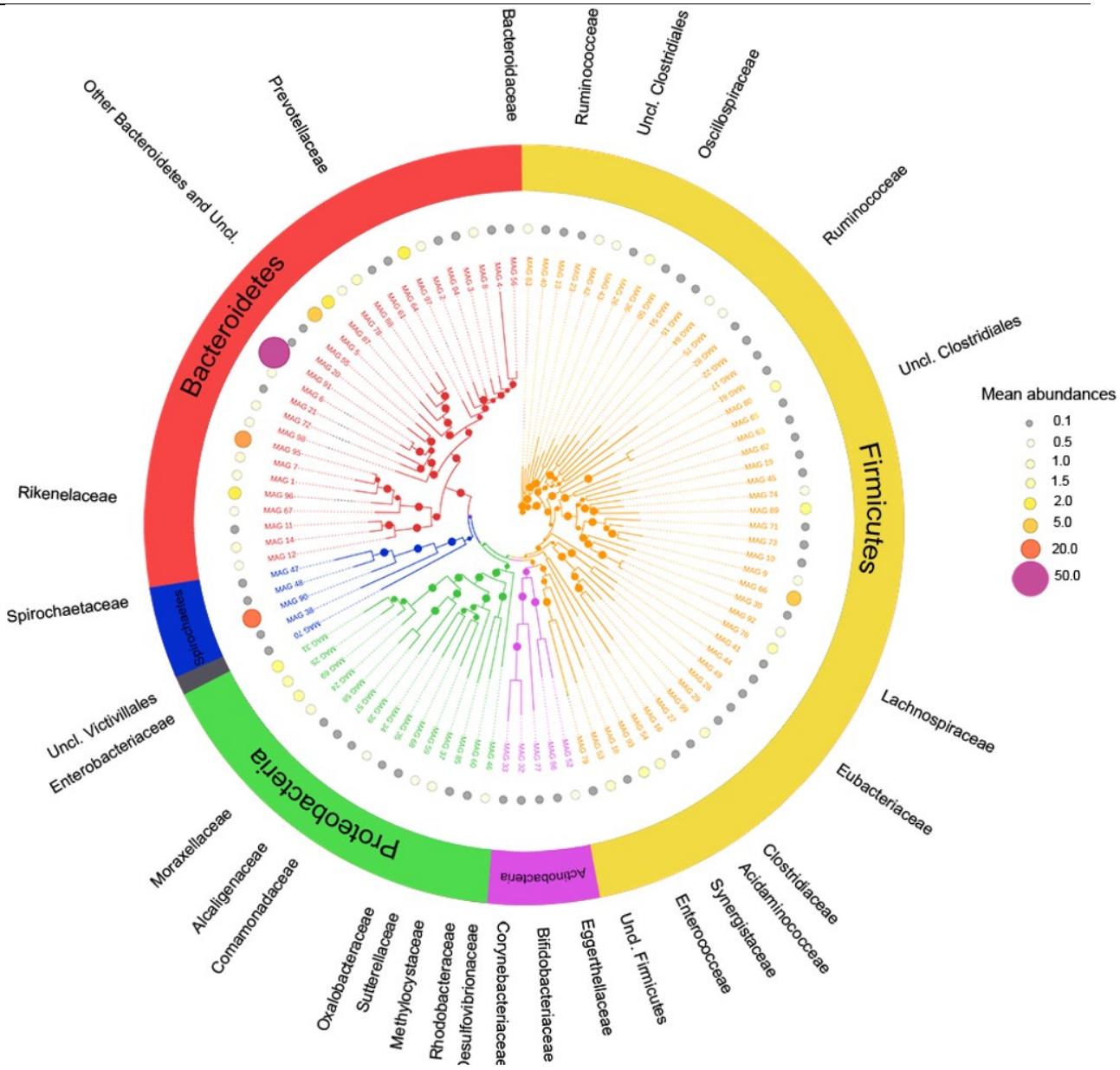


Figure 5: Phylogenomic tree of the selected MAGs with their average abundance.

IV.4.8. Functional potential of TWS-derived MAGs

The functional role of the different MAGs identified in TWS consortium was deduced from the COG classification of their functional gene content. Hierarchical clustering of predicted proteins according to their COG classification roughly clustered the MAGs according to their phylogeny (Figure 4). The first cluster (Cluster 1) depicted in **Figure 6** corresponded to Bacteroidetes-affiliated MAGs and was characterized by a prevalence of the categories G and M which includes CAZymes and are therefore implicated in lignocellulose degradation. Cluster

2 included members belonging to Victivallales (Lentisphearae), some Firmicutes (especially Clostridiaceae and Lachnospiraceae) and all Spirochaetes. This cluster was characterized by a high content of genes involved in transcription process (K) and carbohydrate metabolism (G). Other Firmicutes affiliated MAGs were gathered in Cluster 3 and 4. Cluster 3 is completed by Proteobacteria and Actinobacteria. They overall presents a high number of COGs from the transcription (K) category, as well as either energy production and conversion (C) or amino acid transport and metabolism (E) categories. The relatively lower prevalence of carbohydrate, transport and metabolism (G) COG category indicates a metabolism less centered on polysaccharides degradation. Instead, they harbor a higher content of COGs from the C, E and K categories which denotes a metabolism dedicated to protein and energy production. Cluster 4 contains Firmicutes and one Actinobacteria MAG. They are characterized by a more balanced distribution of COGs in their genome, implying a versatile metabolism less skewed toward a particular action though most MAGs did not present a high number of G COGs. They were clustered because of the over-representation representation of translation, ribosomal structure and biogenesis (J) and replication, recombination and repair (L) COGs in some MAGs.

Overall, Bacteroidetes, Spirochaetes, Actinobacteria and to an extent Proteobacteria were rather convincingly clustered according to their taxonomy while Firmicutes could be found in most clusters. This suggests a high potential of Bacteroidetes and Spirochaetes MAGs to degrade lignocellulose. Conversely, Proteobacteria and Actinobacteria MAGs seem to be active in VFA production. Finally, Firmicutes are presumed to have both activities.

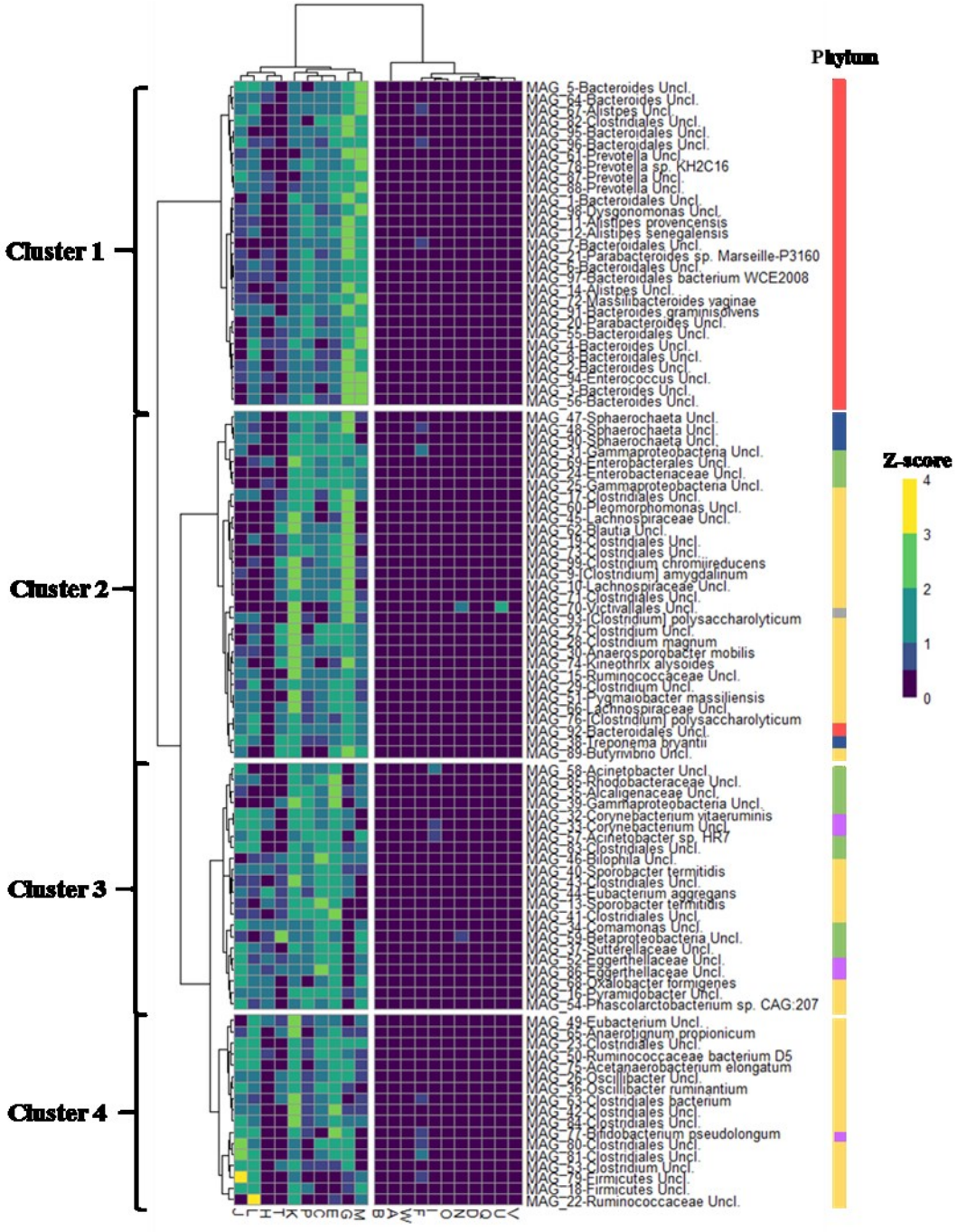


Figure 6: Z-score normalized heatmap representing the prevalence of the different COG categories in the recovered MAGs. The scale represents the Z score of the COG categories. For each MAG, taxonomy at the highest resolved level is shown. The color bar on the side represent the MAGs phyla (red=Bacteroidetes, blue=Spirochaetes, green=Proteobacteria, grey=Lentispherae, purple=Actinobacteria, yellow= Firmicutes).

COG categories: 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

[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*

IV.4.9. TWS MAGs harbors different CAZymes repertoires

To investigate the specific potential role of different MAGs in wheat straw transformation, CAZy modules were predicted in the 99 MAGs. A total of 11,324 CAZymes were found in the MAGs, representing 66% of the number of CAZy modules found in the whole metagenome. Thus, most of the available information in the metagenome was retained in the final MAGs. A total of 6,334 GH were found in the 99 MAGs. The distribution of the modules in the MAGs was not significantly different from the distribution in the whole metagenome or from the distribution of the modules in the CAZy database (Fisher exact test, $p>0.2$) (**Table 1**). With 6,334 glycosides hydrolases, GHs represented the majority of the CAZy modules found in MAGs (56%), followed by GTs (25 %) and CBMs (10 %) (**Figure 7.a**). In contrast, very few CAZymes with auxiliary activity (45) were found in the MAGs which could be expected from an anaerobic system. An exhaustive description of the CAZy modules content in MAGS and their associated activities is provided in the Supplementary data 2. CAZy modules repartition by phyla showed that half of the CAZymes were associated with Bacteroidetes making this phyla the primary lignocellulose degrader of TWS consortium (**Figure 7.b**). Although a majority of MAGs were affiliated to Firmicutes, only 30% of the CAZymes were found in this

phyla. Other phyla represented less than 10% of the number of CAZymes. It is also interesting to note that CAZymes genes represented a substantial portion of the genome of the Lentispherae MAG (4.5% of the total genome length) and Bacteroidetes MAGs (2.7% of the total genome length) (Supplementary data 5) while all other phyla presented a genome where CAZymes took up significantly less of the genome size (around 1%).

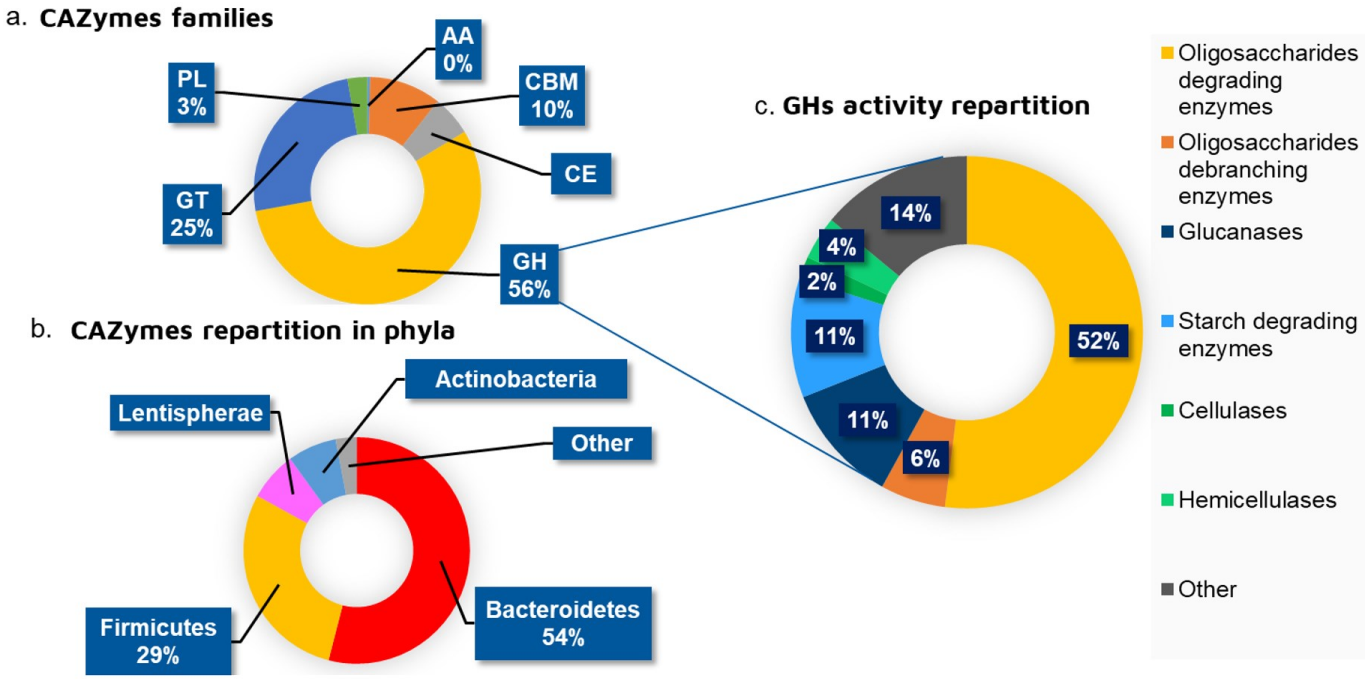


Figure 7: CAZy modules counts and repartition by phylum, module and function

a. Proportion of CAZymes in the MAGs
b. Proportion of CAZymes associated with each phylum
c. Proportion and families of GH associated with polysaccharides degrading functions in the MAGs

A large diversity of GHs which catalyze the hydrolysis of glycosides bonds and are involved in lignocellulose degradation (Cragg et al. 2015) were detected in the MAGs (**Figure 7.c**). In total, 118 GH families were identified, several of which have many subfamilies (Supplementary data 2). They comprised enzymes active on cellulose, hemicellulose, debranching or oligosaccharide degradation (Supplementary data 2).

IV.4.10. Lignocellulose degradation and VFA production potential of the most abundant MAGs

In order to determine how the selected MAGs could operate in the consortium, we focused on the 56 MAGs whose relative abundance was higher than 0.2% at some point during the kinetics. This threshold was chosen in order to keep the representation of all TWS phyla and represent 95.9% of the total abundance of MAGs. Based on the phylogeny of these MAGs, their GHs and VFA production related enzymes content (Polansky et al. 2016; Deusch et al. 2017; Tilocca et al. 2017), we performed multivariate PCA (Supplementary Figure 6) and a PLS-DA (**Figure 8**) to discriminate their function.

As the PCA (Supplementary Figure 6) seemed to separate the MAGs according to their phylogenomy and were separated by GHs on one side, GT and VFA-related COGs on the other, we decided to do a PLS-DA using phyla as the discriminant.

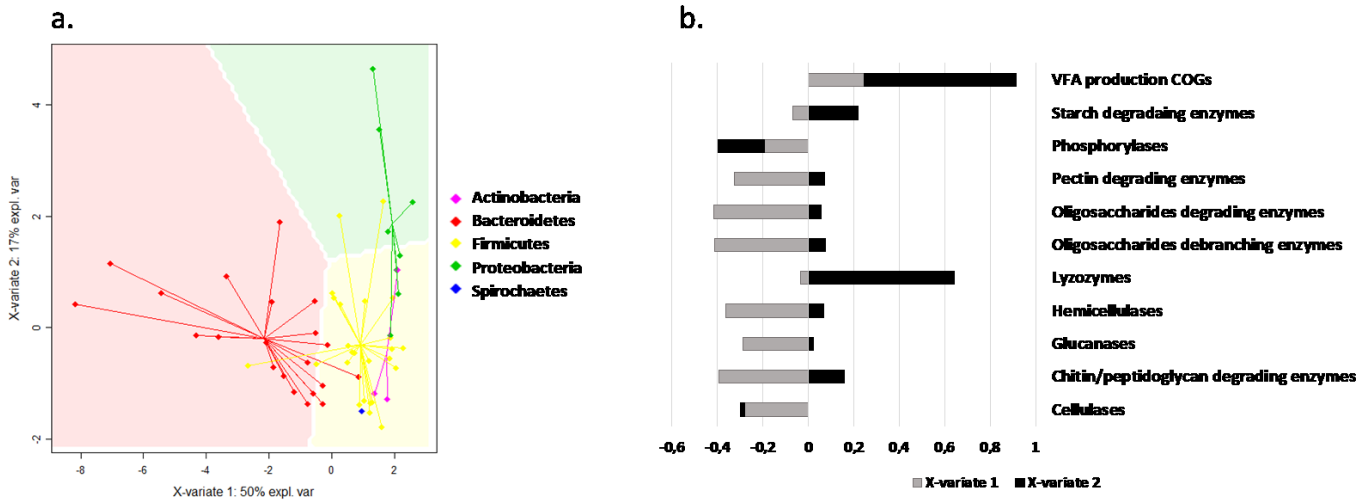


Figure 8: **(a.)** Partial Least Squares Discriminant Analysis (PLS-DA) of *the most abundant MAGs discrimined by their phyla.* Each point represents a MAG and the color represents the phylum of the MAG. (b.) Loadings of the different variables used for the PLS-DA associated with their component.

The PLS-DA successfully separated theses MAGs according to their phylogeny and specific activity in the consortium (**Figure 8.a.**). By analyzing the loadings of the variables used in the PLS-DA (**Figure 8.b**.) it was evidenced that MAGs belonging to Bacteroidetes (on the left side of the PLS-DA) were characterized by their high prevalence of GH and thus their important contribution to lignocellulose degradation. In contrast, Proteobacteria-related MAGs (grouped in the PLS-DA top right corner) contributed importantly to VFA production. Most Firmicutes

and other minor phyla contributed mainly to VFA production but in a lower extend compared to Proteobacteria. These specializations were also backed by the percentage of the genome length dedicated to either CAZymes or VFA production COGs (Supplementary data 5, 8) and is in agreement with what was inferred from their COGs families breakdown. In order to more precisely identify the role of each MAG in the community, we applied hierarchical clustering using Ward method to infer clusters of cohesive activity profile (**Figure 9**). CAZymes modules, separated into their activities (cellulases, hemicellulases, polysaccharides degrading or debranching enzymes, glucan, starch, pectin or chitin degrading enzymes, lyzozymes, CBM, GT, CE, AA, CBM), COGs linked to VFA production, divided between acetate, butyrate and propionate metabolism and lignin-degrading activity were selected for this clustering. Five clusters were identified.

This analysis revealed three MAGs (MAG_2, 56 and 8), all affiliated to Bacteroidetes, with the highest GH content (**Figure 9**) were clustered together (Cluster A). These MAGs harbored both cellulases (GH5, GH9) and hemicellulases (GH10, GH26, GH95) activities but also an activity toward polysaccharides, both degrading and debranching as well as a high number of GH involved in pectin and chitin degradation. Overall they can be considered lignocellulose degradation powerhouse with a very high GH content in the genome. It must be noticed that they also harbor PULs systems (see Supplementary data 7) to maximize their efficiency for lignocellulose degradation and provide information on the targeted polysaccharides (Lapébie et al. 2019). In contrast, the MAGs of this cluster were shown to only have little proteins related to VFAs production.

Cluster B, composed by 8 MAGs (MAG_24, 25, 31, 69, 34, 58, 32 and 99), affiliated to β and γ-Proteobacteria, Actinobacteria and Firmicutes displayed a low content of GHs with the notable exception of some oligosaccharides degrading enzymes (GH1, GH2, GH3, GH13) and a high content of enzymes involved in VFA production . All these MAGs were equipped with genes from the acetate and butyrate synthesis pathway and only MAG_31 lacked the propionate synthesis COGs. Interestingly, while this clusters was found to be the more skewed toward VFA production, it also presented the most lignolytic putative proteins. Indeed, with the exception of MAG_34 lacking laccase and DyP-peroxidase genes and MAG_99 DyP-peroxidase and β-etherase genes, all MAGs in Cluster B presented a high number of potentially lignolytic predicted proteins.

Cluster C had a varied composition with most phyla represented. Overall, MAGs from this cluster had a low number of predicted proteins linked to either lignocellulose degradation or VFA production and may be implicated in totally different aspects of the consortium role.

Cluster D was composed by 6 Firmicutes MAGs (MAG_10, 27, 94, 29, 9, 73). This cluster displayed a high AA and glucanase activity. VFA-related COGs were also found in this cluster although propionate synthesis was more predominant than in Cluster B. Compared to Cluster B it also displayed a higher number of lignocellulose-targeted activities with MAG_10 having access to both cellulases and hemicellulases for instance. Another major difference with Cluster B is the lack of laccases, Dyp peroxidases and β-etherases.

Cluster E, finally included a mix of Bacteroidetes and Firmicutes. Notably, MAG_91-*Bacteroides graminisolvens*, the most abundant MAG in TWS consortium was found in there and displayed both cellulase and hemicellulase CAZymes but also oligosaccharide degrading enzymes, emphasizing is potential role in lignocellulose degradation. It is also quite representative of the cluster as most MAGs in it were found to have a higher number of lignocellulose degrading activity than average and, similarly lower than average VFA producing and lignolytic activities.

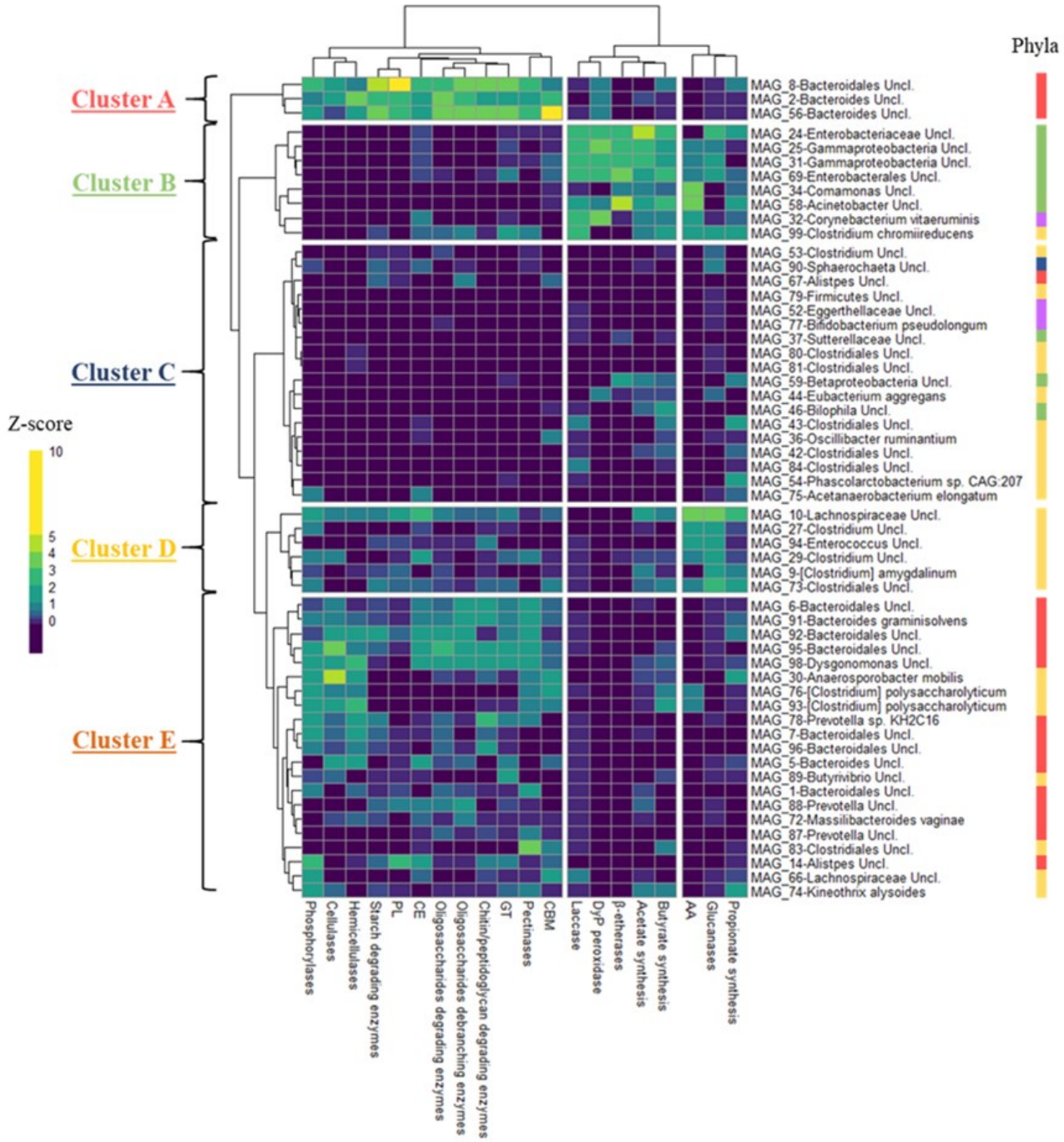


Figure 9: Hierarchical clustering of the most abundant MAGs (Max abundance>0.2%) based on activities of their CAZy modules, VFA related COGs and lignolytic predicted proteins (Count>30) content. Hierarchical clustering is done with Ward method on Z-score normalized data. The five clusters are assigned a color (Cluster A, red, Cluster B, green, Cluster C, Blue, Cluster D, yellow, Cluster E, orange). The color bar on the side represent the MAGs phyla (red=Bacteroidetes, blue=Spirochaetes,

green=Proteobacteria, grey=Lentispherae, purple=Actinobacteria, yellow= Firmicutes).

IV.4.11. Lignocellulose degradation and VFA production over time

To better understand the temporal distribution of enzymatic activities in the bioreactors we assessed the evolution of GHs and VFA related COGs abundance with time and observed the impact of the clusters (**Figure 10**). Five different trends could be observed in the temporal evolution of enzymatic activities of interest (**Figure 10**). First of all, any activity in link with lignin degradation quickly decreased in predicted protein abundance. CE, PL, starch degrading enzymes and phosphorylases, on the contrary increased with time. AA, acetate and butyrate synthesis first decreased by 20-30% between T1 and T2 (2-3 days) before increasing again to initial levels while Propionate synthesis, CBM, cellulases, hemicellulases and oligosaccharides degrading enzymes increased by 20-45% in the same time. Finally some activities remained stable throughout incubation.

Overall, lignin degrading activity was mainly present at T1. This has to be put in relation with Cluster B harbouring the vast majority of proteins with such activity (**Figure 9**) which transpire in Cluster B making up to 95% of Dyp-type peroxidase and β-etherases abundance at T1 and around 40% of AA and laccases. The decrease in abundance of these functions can be linked with the decrease of abundance of Cluster B MAGs during incubation.

Interestingly, while cluster B and Cluster D contained the majority of COGs related to VFA production, they actually do not exceed 45% of the abundance of the predicted proteins (at T1) as Cluster C and most notably Cluster E make up for the majority of the abundance after 3 days of incubation (T2). The difference in behavior between propionate and acetate/butyrate synthesis could be attributed to the weight of Cluster B in the latter activities.

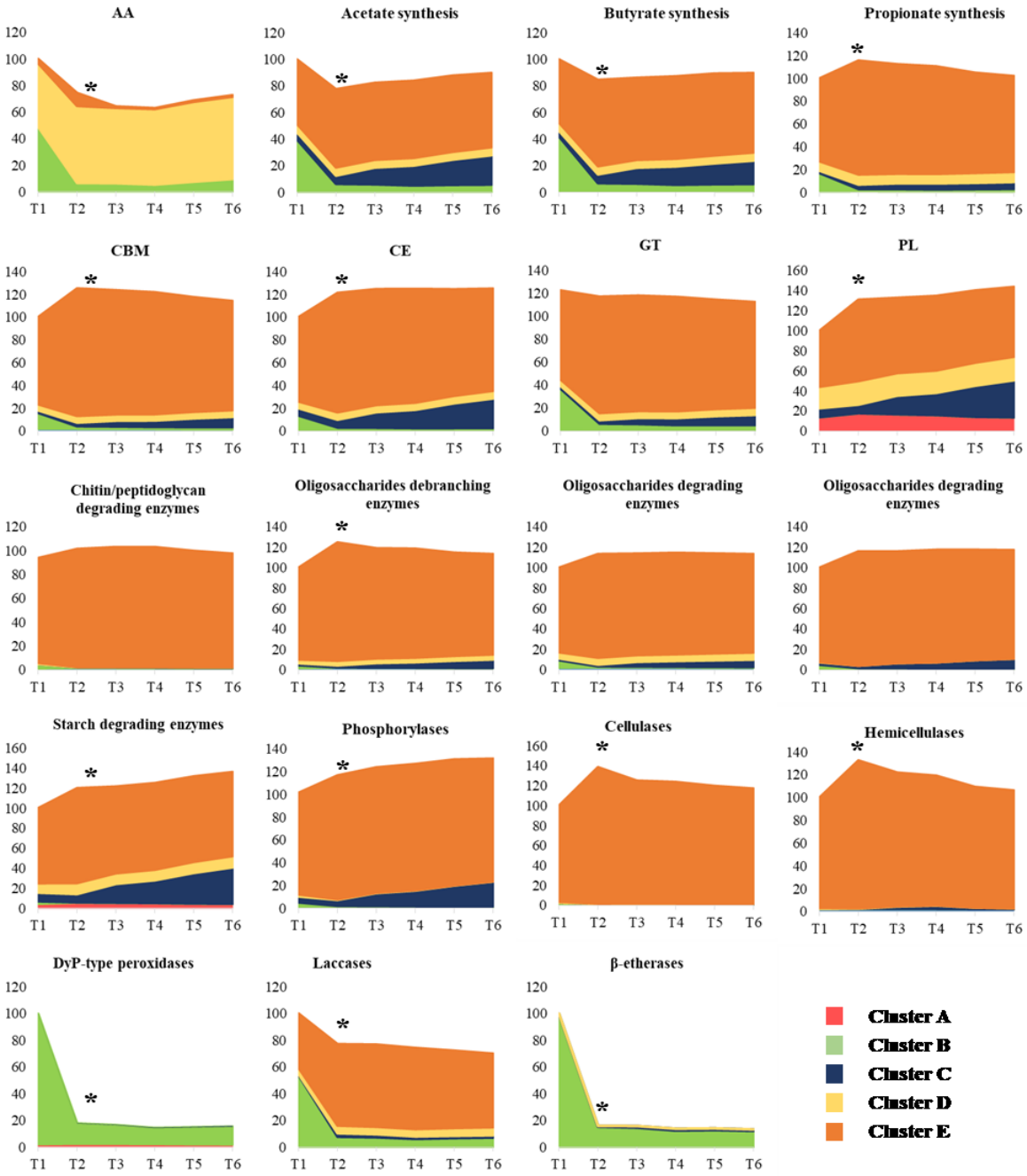


Figure 10: Temporal evolution of abundance of the CAZymes , lignin-degrading predicted proteins and VFA production COGs genes grouped according to their activity in % of their abundance after one day of incubation. Colors represent the proportion of the abundance that can be attributed to the cluster previously defined. * represent statistical difference with T1 (T.test, p-value <0.05)

For all activities related to lignocellulose degradation, while Cluster A was most promising in term of number of enzymes, its actual impact on gene abundance is negligible. With the exception of polysaccharides lyases and starch degrading enzymes where up to 17% of the abundance is due to Cluster A, it never represented more than 1% of the total abundance of proteins for any activity. In fact, Cluster E is the single most important contributor to lignocellulose degrading activities abundance. For some activities, namely cellulases, hemicellulases and chitin/pepttidoglycan degradation it made up more than 99% of the abundance. This clearly show the importance of this cluster in the consortium. However, it also showcased how MAG_91, affiliated to *Bacteroides graminisolvens* shaped the TWS consortium. Though the temporal evolution of the functions is not altered when MAG_91 is removed from the dataset, the contribution of Cluster E is greatly reduced overall (Supplementary data 6).

Carbohydrate esterases, phosphorylases and lyases abundance increased with time. These CAZymes are often described for their role in helping GHs in their role by removing decorations from polysaccharides and depolymerizing some polysachharides (Cragg et al. 2015). Most CE were CE1 and CE4, whose activity is often acetyl xylane esterase necessitate a first step of unlocking by cellulases and hemicellulases that may explain why these genes see their abundance increasing later.

IV.5. DISCUSSION

Over the past decade different studies have try to understand the functioning of lignocellulose degrading ecosystems, particularly those displaying a high degradation efficiency such as the termite gut microbiome (Warnecke et al. 2007; Rossmassler et al. 2015; Marynowska et al. 2017; 2020; Grieco et al. 2019; Waidele et al. 2019; Calusinska et al. 2020; Hervé et al. 2020). However, the high complexity of these ecosystems hinders a deep understanding of the specific role played by different members in the community. In an attempt to reconstruct the genomes involved in lignocellulose bioconversion, we studied here TWS consortium derived from *N.*

ephratae gut microbiome, obtained by an enrichment approach using anaerobic bioreactors and wheat straw as sole carbon source, displaying a lower level of diversity compared to the original termite gut microbiome but high capacity for biomass degradation. Combining 16S rRNA and whole genome sequencing of TWS, we successfully described the microbial diversity dynamics and recovered 99 MAGs. Additionally, we characterized the MAG content in CAZymes and VFA-related proteins and determined their temporal dynamics along wheat straw degradation.

Overall, a similar COGs distribution has been found in other lignocellulose degrading consortia derived from compost (Wang et al. 2016; Alessi et al. 2018) or fungus growing termites (Bastien et al. 2013). And, with the exception of the N category (cell motility), which was comparatively lower, the distribution was close to that observed in *N. takasagoensis* symbionts (He et al. 2013).

Whole genome sequencing enabled to reconstruct 99 genomes with at least a medium quality with 38 MAGs being of high quality. It allowed us to reconstruct the genome of the most abundant species present in the lignocellulose biodegradation experiments, including *Bacteroides graminisolvens*. Although culturing the *N. ephratae* gut microbiome in bioreactors completely altered the original microbial diversity of the termite gut (Auer et al., 2017, Lazuka et al., 2018), the enriched TWS consortium maintains a high capacity to degrade wheat straw into VFAs. The enriched TWS consortium was found to be predominantly composed of Bacteroidetes, Firmicutes and Proteobacteria although initial guts were dominated by Spirochaetes and Fibrobacteres. Interestingly, whole genome sequencing allowed us to observe the presence of Spirochaetes that we were unable to observe with 16S rRNA gene sequencing. Overall, the composition of TWS consortium was closer to that of other lignocellulose-degrading anaerobic systems, such as the moose rumen (Svartström et al. 2017), cow rumen (Lazuka et al. 2015) or biogas bioreactors (Campanaro et al. 2020) than to that of the actual termite gut. The predominance of Bacteroidetes was especially striking given the low amount of Bacteroidetes found in the initial gut. It appears that the growth of Bacteroidetes members was certainly favored by the culture conditions imposed in the bioreactors. Similarly, using wheat straw as sole source of

carbon during the enrichment process, instead of the wood biomass usually used for *N. ephratae* termite as food source, probably had an impact on the selection of species more adapted to grass degradation (Calusinska et al. 2020).

Whole genome sequencing has provided a better understanding of the metabolism of TWS implemented in anaerobic bioreactors. We were especially interested in understanding how TWS degrade lignocellulose and use it as substrate to produce VFAs. Thus, we were interested in MAGs containing CAZymes involved in lignocellulose degradation and proteins related to VFA production. We obtained 6334 GHs on the MAGs, which is in the same order of magnitude as in MAGs obtained from moose rumen (6283 GHs for the same number of MAGs) (Svartström et al. 2017).

Most of the identified cellulases and hemicellulases were found in a cluster of MAGs belonging to Bacteroidales order (cluster A in Figure E), particularly in genomes affiliated to the genus Bacteroides. Interestingly, some cellulases and hemicellulases were also found in MAGs from cluster D and E, affiliated with Firmicutes, particularly with *Clostridium polysaccharolyticum* and *Anaerosporobacter mobilis* and *MAG_10-Lachnospiraceae Uncl.* (Lachnospiraceae) MAGs which is in agreement with the literature as both families have been singled out for their high GHs content in various digestive systems (Thomas et al. 2011; Svartström et al. 2017; Bertucci et al. 2019).

Proteobacteria and Firmicutes (particularly Clostridiaceae) were found to have a high number of COGs related to VFA production. This is not surprising since both have been previously linked to an increase in VFA production (Atasoy et al. 2019). Our data suggests a specialization of the Bacteroidetes towards lignocellulose degradation and a specialization of Proteobacteria towards VFA production. As most proteobacteria were found lacking the genetic tools to degrade cellulose and hemicellulose, they most probably benefit from the breakdown of the polymers into smaller oligosaccharides for which they have oligosaccharides degrading enzymes. However, a few Proteobacteria were found to be asaccharolytic (MAGs affiliated to *Desulfovibrio*, *Acinetobacter* or Sutterellaceae most notably). While they don't directly benefit from lignocellulose breakdown, their high number of genes coding for lysozymes could hint

at dead cell recycling. Though not directly involved in lignocellulose degradation, they may help by reducing inhibition of the organic acids by anaerobic respiration (Su et al. 2017) which may be confirmed by metatranscriptomics.

Besides a global characterization of the selected microbial consortium, we also analyzed the temporal distribution of both the MAGs and the functions of interest. We could thus highlight dynamics in the bioreactors. The beginning of the digestion highlighted the role of bacteria who lack the ability to degrade lignocellulose by themselves and especially Proteobacteria and *Clostridium*. These bacteria were thus starting to produce VFAs from free sugars which is in agreement with previous results where VFAs production was found to precede substrate degradation (Lazuka et al. 2018). It should also be noted that overall, the temporal profile of propionate synthesis activity differed from acetate and butyrate synthesis as differents MAGs were found to be implicated. A second phase (until 3 days of digestion) was observed where activities associated with lignocellulose degradation such as cellulases, hemicellulases, oligosaccharides degrading and debranching enzymes increased. This is mainly correlated to the increase in Bacteroidetes and Lachnospiraceae which possess a genome which is rich in the according GHs. This also coincide with the increase in CMCase and xylanase activities found in the bioreactors and the beginning of lignocellulose degradation. Interestingly, it was found that hemicellulose degradation preceded cellulose degradation and while hemicellulose degradation plateaued after 4 days, cellulose was still being degraded during the next few days (Lazuka et al. 2018). Here we show that bacteria with hemicellulases and cellulases activities are largely the same. Two possible hypothesis could explain this discrepancy. On the one hand, these bacteria could produce both enzymes simultaneously but given the 3D structure of lignocellulose, hemicellulose may need to be degraded first in order to release cellulose. For instance, numerous Bacteroides possess PULs with combinations of cellulases and hemicellulases (GH5+GH10(+GH26)) while Flavobacteriaceae can have PULs with GH9+GH10.(Terrapon et al. 2018) On the other hand, the transcription of the enzymes may be time delayed. Further transcriptomic or proteomic analysis would be needed to verify this hypothesis. A third phase was then observed where the previously mentioned activities were

dwindling. Instead *Sphaerochaeta*, whose exact role could not be identified in this study, were found to increase and replace other bacteria. *Clostridia*, *Butyrivibrio*, *Allistipes* and *Prevotella* were also found to increase in abundance after 4 days of incubation. The fact that *Sphaerochaeta* and Clostridia follow the same trend could be explained as most genes related to the metabolic activity of *Sphaerochaeta* were imported from Clostridia by horizontal transfer (Caro-Quintero et al. 2012). As the contribution of identified clusters to the functions was quantified, we were able to identify MAG_91, affiliated to *Bacteroides graminisolvens* to not only being the most abundant bacteria in TWS but also being the single most influential MAG in the lignocellulose degradation process as it majorly contributed to the abundance of most lignocellulose degradation related activities.

In a previous work, we identified β-etherases and DyP-type peroxidases as potentially involved in lignin degradation by an anaerobic consortium derived from termite gut (Dumond et al. 2021) and though no lignin degradation was observed in the samples analyzed in this study, we were able to identify such enzymes in Proteobacteria MAGs, bringing new evidence to this hypothesis.

IV.6. CONCLUSIONS

Termites have proven to be a valuable reservoir for enzymes displaying lignocellulosic degradation activities. The implementation of a bioreactor taking advantage of this activity to produce VFAs is promising as well. However, while studies have tried to understand the role of the termite microbiome in degrading lignocellulose, there is a lack of information on how this microbiome can degrade lignocellulose outside of its host, in a bioreactor environment. In this study, using whole genome sequencing metagenomics as well as 16S rRNA gene amplicon sequencing, we analyzed the bacterial profiles of a bioreactor using a selected termite gut consortium from *N. ephratae* and its evolution during wheat straw digestion. By analyzing both temporal evolution of microbial diversity and temporal evolution of CAZymes and enzymes involved in VFAs production, we expanded the knowledge on the underlying

mechanisms behind VFA production in a bioreactor inoculated with a selected termite gut microbiome. In this study we have successfully reconstructed the genomes of the most abundant bacteria in the system and have identified the CAZymes associated with these bacteria. Our results clearly showed that the consortium is constituted of a few MAGs with a very high abundance and many with very low abundance. We have also showed there is a specialization of the different phyla with Bacteroidetes heavily involved in lignocellulose degradation while Proteobacteria were more involved in the production of VFAs. Moreover, by investigating the temporal evolutions of microbial diversity and genes associated with them we were able to show that most VFAs producing strains were more abundant in the first few days and were probably producing VFAs from free polysaccharides. We also demonstrated than lignocellulose degrading strains were especially prominent between two and five days of degradation in which probably corresponds to times where all free sugars were already consumed. We further hypothesized that hemicellulose breakdown was the key lock these strains helped to overcome. Overall we managed to reconstruct the genomes of 99 bacteria and their importance related to lignocellulose degradation and VFAs production. However, though the shotgun metagenomics approach allowed us to have an in-depth look into the functions of these bacteria, other “omics” approaches could be used to dig deeper into the mechanisms of lignocellulose degradation by a termite gut microbial consortium. Metatranscriptomics would be an invaluable tool to pinpoint the gene expression and therefore have a better view of the temporal evolution of lignocellulose degrading activities. Moreover both proteomics and metabolomics would give us a better grip onto the underlying mechanisms.

IV.7. AVAILABILITY OF DATA AND MATERIALS

The raw sequence data were deposited at the European Nucleotide Archive (www.ebi.ac.uk) under accession number PRJEBXXX (16S rRNA amplicon sequencing) and PRJXXXXX (Metagenomics). The genome assemblies and all other annotation data are available from the corresponding author under reasonable request.

IV.8. ACKNOWLEDGEMENTS

Authors would like to thanks the GeT-PlaGe for performing sequencing and GenoToul platform for providing computational resource and support for software installation and updates. Authors would also like to thank David Sillam Dussès (French Institute for Research and Development, Bondy, France) for providing the initial termites used in this study.

IV.9. FUNDING

This publication is a part of the Zelcor project and has received funding from the Bio Based Industry Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement no. 720303. It was also supported by the French National Institute for Agricultural Research (INRA), the Region Languedoc-Roussillon Midi-Pyrénées grant 31000553 and by the Carnot Institute 3BCAR—Insyme project.

IV.10. INDEX

AA: Auxiliary activity

CAZymes: Carbohydrate-active enzymes

CBM: Carbohydrate-binding module

CE: Carbohydrate esterase

COG: Clusters of Orthologous Groups

GH: Glycoside hydrolase

MAGs: Metagenome assembled genomes

OTUs: Operational taxonomic units

PL: Polysaccharide lyase

PLS-DA: Partial least square – determinant analysis

PULs: Polysaccharide utilization loci

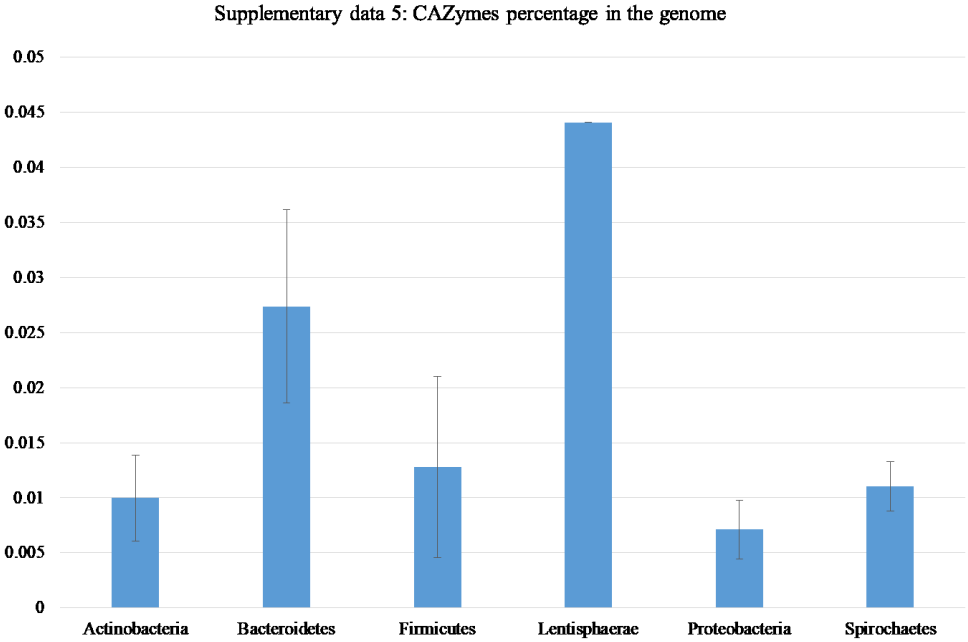
SRA: Sequence Read Archive

VFA: Volatile fatty acid

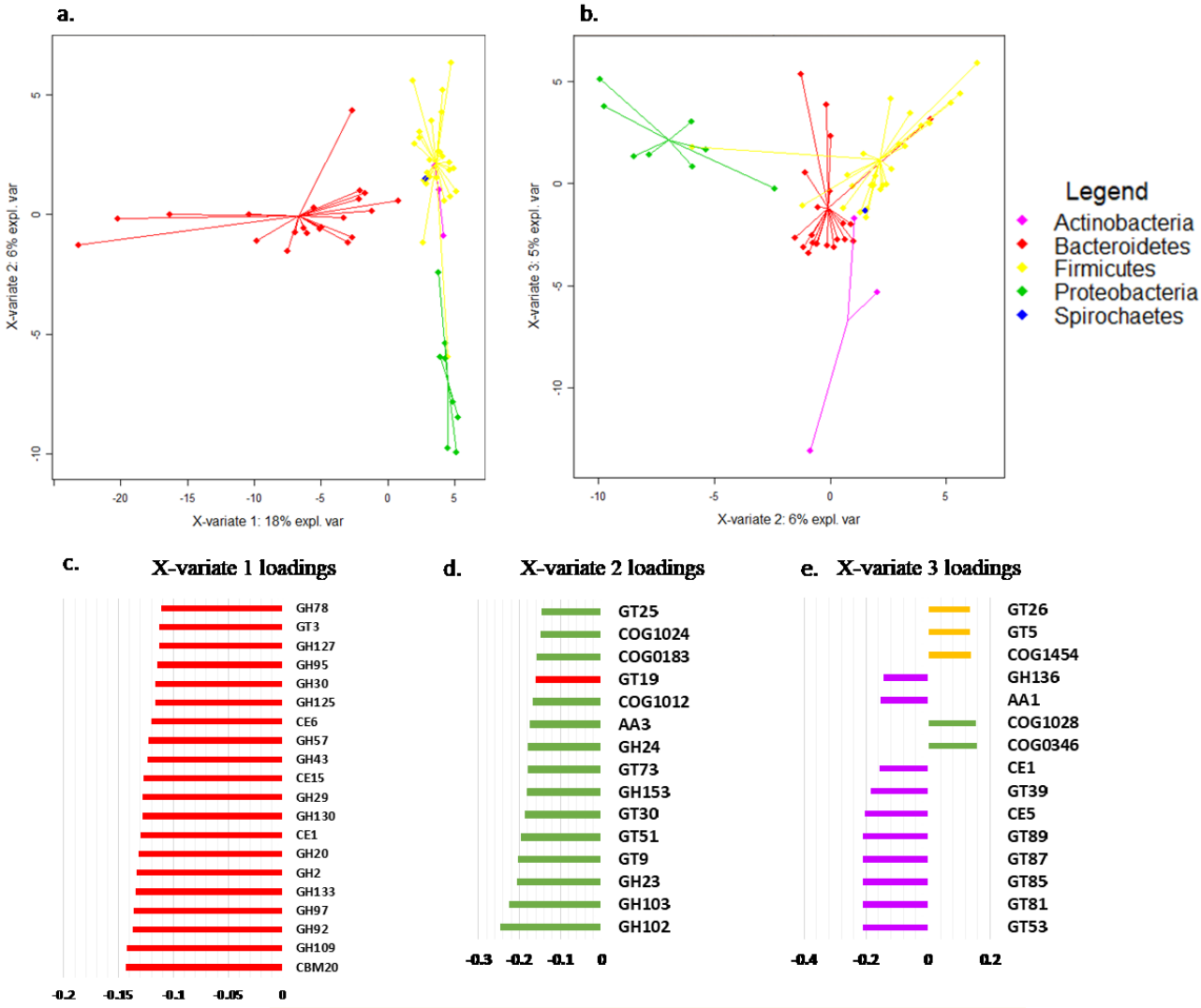
IV.11. SUPPLEMENTARY DATA

Supplementary Table 1: List of samples and number of reads obtained from sequencing

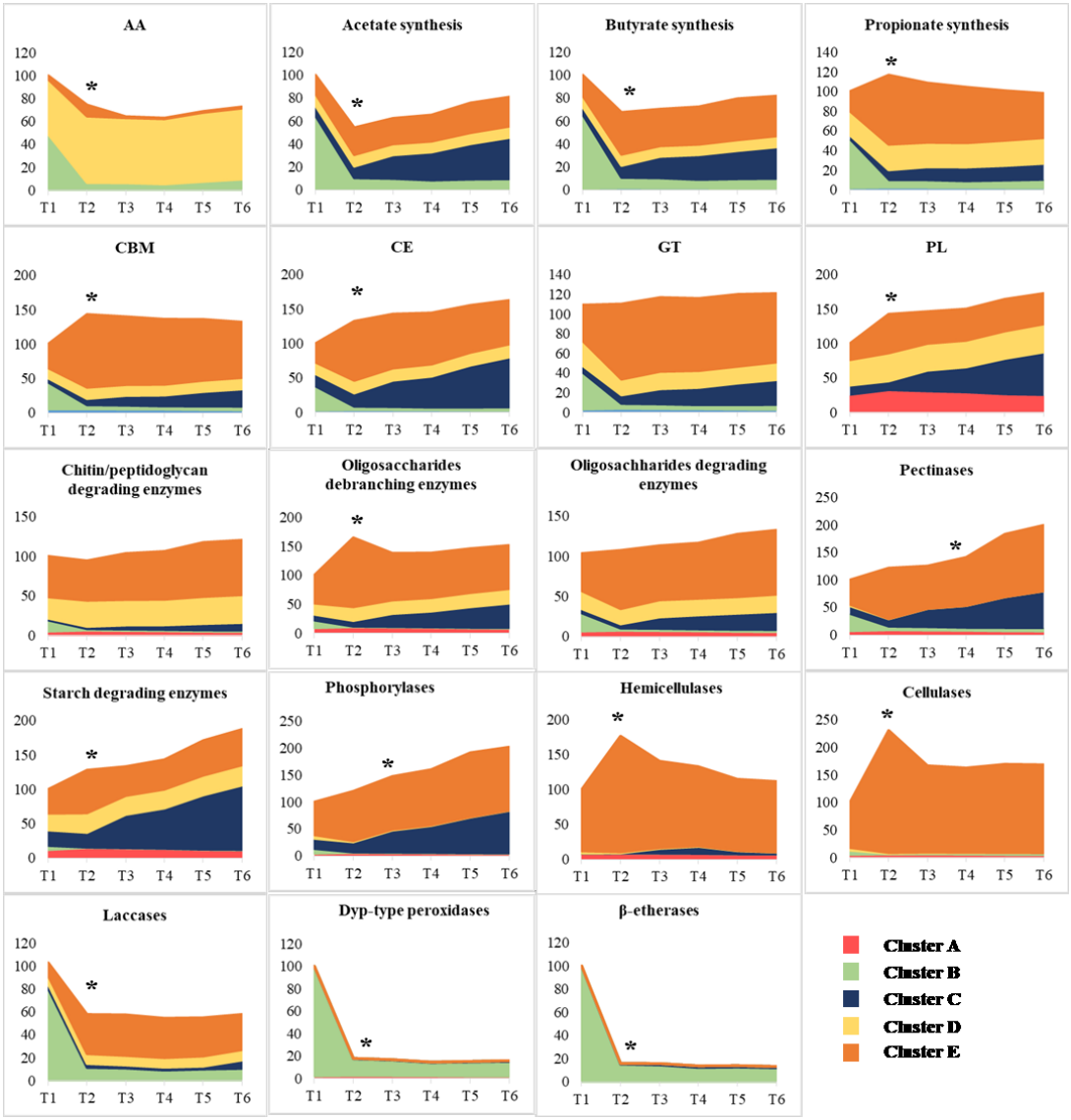
SAMPLE	NUMBER OF READS (MILLIONS)
ST1T1	24.7
ST1T3	26.3
ST1T6	28.1
ST1T7	25.8
ST1T11	23.4
ST1T12	27.1
ST2T1	26.9
ST2T4	30.1
ST2T6	30
ST2T7	29.3
ST2T10	28.6
ST2T12	28



Supplementary data 5 : CAZymes percentage in the genome of phyla present in TWS



Supplementary data 6: Principal component analysis (PCA) of *the most abundant MAGs calculated on predicted proteins numbers*. Each point represents a MAG and the color represents the phylum of the MAG. (c.d.e.) Loadings of the different variates



Supplementary data 8: Temporal evolution of abundance of the CAZymes , lignin-degrading predicted proteins and VFA production COGs genes grouped according to their activity in % of their abundance after one day of incubation without MAG_91. Colors represent the proportion of the abundance that can be attributed to the cluster previously defined. * represent statistical difference with T1 (T.test, p-value <0.05)

V. CHARACTERIZING THE LIGNIN DEGRADATION ABILITY OF *NASUTITERMES EPHRATAE* GUT MICROBIOME AND ITS DERIVED MICROBIAL CONSORTIA ENRICHED ON WHEAT STRAW

V.1. CHARACTERIZING THE LIGNIN DEGRADATION ABILITY OF *NASUTITERMES EPHRATAE* GUT MICROBIOME AND ITS DERIVED MICROBIAL CONSORTIA ENRICHED ON WHEAT STRAW

Louison Dumond[a], Clémentine Duarte[a], Clément Salvagnac[a], L. Cézard[b], A. Majira[b], Betty Cottyn[b], David Sillam-Dussès[c], Edouard Miambi[d], Stéphanie Baumberger[b], Guillermina Hernandez-Raquet[a]*

[a]Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.

[b] Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000, Versailles, France

[c] Faculté des Sciences et Technologie, Université Paris Est Créteil, Département ECOEVO, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES, Paris), 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France

[d] Université Paris 13 - Sorbonne Paris Cité, Laboratoire d'Ethologie Expérimentale et Comparée, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

V.2. ABSTRACT

Termites are crucial to lignocellulose degradation and their microbiome has shown to be potent lignocellulosic consortia. However, while the knowledge of how termite-derived consortia can degrade cellulose and hemicellulose, there is a lack of information on how it could overcome the lignin barrier. Furthermore, as termite gut consortia thrive in anaerobic conditions, little is known about their potential in aerobic conditions. A powerful strategy to overcome the complexity of termite gut microbiome is to enrich the functionally-relevant fraction of the microbial community by culture under defined conditions. Such approach enable to minimize the ecosystem's complexity and increase the community fraction expressing a particular functional potential.

In this study, we cultivated a Nasutitermes ephratae gut microbiome on technical lignins and wheat straw (pretreated and unpretreated) with the goal of enriching a consortium with high lignolytic properties. We coupled chemical and 16S amplicon sequencing to characterize the consortia. We most notably showed using enzymatically lignin-enriched wheat straw as a substrate allows for the enrichment of a lignolyitic consortium and that Pseudomonadaceae may be crucial to its performance.

V.3. INTRODUCTION

The global demand for reducing the greenhouse emissions has motivated the development of lignocellulose biorefineries for a sustainable production of energy, fuels and chemicals from renewable resources. Lignocellulose is the most promising non-food resource since it is abundantly available from agricultural and forestry residues (Bentsen et al., 2014). However, lignocellulosic biomass is highly resistant against bioconversion largely due to its complex polymeric composition. Lignocellulose is composed by 30-50% of cellulose, 20-30% of hemicelluloses and 15-30% of lignin (Lapierre, Pollet, et Rolando 1995; del Río et al. 2012). These polymers are elaborately interwoven with each other to form a complex three-dimensional structure which limits its degradability. In particular, lignin provides the plant mechanical strength and protection against environmental and microbial attack (Liu et al., 2018). As a result, it represents also the main limiting factor to polysaccharide conversion in biorefinery approaches while lignin is the most abundant natural renewable source of aromatic compounds for chemical industries. Therefore, to enhance the entire lignocellulose valorization, finding new catalyst for lignin bioconversion is a topical objective (Kamimura et al. 2019).

Microbial degradation of lignin has mainly been investigated in fungi especially in white-rot fungi such as *Phanerochaete chrysosporium* and *Trametes versicolor* (Sigoillot et al. 2012). There are also many reports of bacteria able to degrade lignin. Proteobacteria and Actinobacteria have been found to be the phyla most often associated with lignin degradation (Brink et al. 2019). While soil is the ecosystem that contributes the most to the identification of new bacteria able to break down lignin

(Brink et al. 2019), several bacteria have also been isolated from the digestive system of xylophagous insects such as termites (H. Zhou et al. 2017; Couger et al., 2020).

Termites play a crucial role in the decomposition of lignocellulose in natural environments (König et al., 2013) and the degradation of the cell wall polysaccharides in termites has been extensively documented in the literature (Andreas Brune 2014). Strong efforts have been realized to understand the *in vivo* functioning of the gut microbiome of higher termites, including *Nasutitermes* sp., and its relationship with the termite host (Warnecke et al. 2007; Burnum et al. 2011). Applying metagenomics, metatranscriptomics and metaproteomics, several studies discovered the rich bacterial diversity and the vast repertoire of carbohydrate active enzymes (CAZymes) involved in lignocellulose degradation encoded by the symbiotic bacteria associated to these digestive systems and their response to changes in diet composition (Calusinska et al. 2020; Marynowska et al. 2020). In contrast, the way termites and their microbiome overcome the lignin barrier is still poorly understood. It was initially hypothesized that no lignin degradation occurred in termites and that the extent of termites activity was limited to the degradation of mono- or dimeric-lignin model compounds (Kato et al., 1998; Ke et al., 2011). More recent studies, making use of the advancement of tools for lignin analysis, especially the use of multi-dimensional nuclear magnetic resonance (NMR) techniques (H. Kim & Ralph 2010; Mansfield et al. 2012) have shed a new light onto lignin degradation by termites and their gut microbiome (Hongjie Li et al. 2017; Tarmadi et al. 2018; Dumond et al. 2021).

Nevertheless, deciphering the functioning of the lignin degrading communities is challenging do to the tremendous complexity of these ecosystems. A powerful strategy to overcome the complexity of termite gut microbiome is to enrich the functionally-relevant fraction of the microbial community by culture under defined conditions. Such approach enable to minimize the ecosystem's complexity and increase the community fraction expressing a particular functional potential (DeAngelis et al. 2010). Such approach has been successfully applied to environments such as sugarcane plantation soil (Moraes et al. 2018), cow rumen supplemented with wheat straw (Lazuka et al. 2015) or soil supplemented with plant-based materials (de Lima Brossi et al. 2016). Therefore, the lignin degrading potential of termite gut microbiome

could be better assessed by selecting and enriching a microbial consortium with high lignin degrading properties that could be relevant for the biorefinery.

In previous *in vitro* studies, using controlled bioreactors, we have established that the lignocellulolytic potential of termite gut microbiomes can be enriched in anaerobic conditions and exploited to transform the polysaccharide fraction of wheat straw into, for instance, volatile fatty acids (VFAs) (Lazuka et al. 2018). We also demonstrated the ability of termite gut microbiome to remove and modify lignin in anaerobic conditions (Dumond et al. 2021). We hereafter investigated the capacity of the gut microbiome of higher termite *Nasutitermes ephratae* to degrade lignin under aerobic conditions. Aerobic conditions were studied because lignin degradation mostly involve oxidation reaction (Bugg et al. 2011) and the termite gut presents micro-oxic zones that may enable the growth of aerobic and facultative anaerobic microorganisms with lignolytic potential (Brune et al., 1995). To achieve this goal, we assessed the potential of *N. ephratae* termite gut microbiome to degrade technical and wheat straw lignin. Then, a sequencing batch enrichment process was performed using raw wheat straw and lignin-enriched wheat straw (LWS) as substrates. Lignin degradation was assessed by deep chemical analysis of the substrate before and after the biological treatment. Additionally, 16S rRNA gene sequencing was used to assess the enrichment process' impact on the microbial community diversity and have new insights on the taxa responsible of lignin degradation.

V.4. MATERIAL AND METHODS

V.4.1. Termite gut inocula

The termite guts from higher termites *Nasutitermes ephratae* were collected at the IRD (Institute for Research and Development, Bondy, France) in January and April 2018. Termites colonies were bred at the IRD in a controlled room in tropical conditions of temperature (28°C) and humidity (80%) and fed with birch wood. Termite workers were collected from their nest and cold anaesthetised. The whole gut dissection consisted in a removal of the gut by tweezers and forceps. The dissection was performed in sterile conditions. The guts were immediately suspended in a

physiological solution after removal and pooled in 50 ml centrifugation tubes (Auer et al., 2018). Once 100 guts were collected, the tube was snap frozen at -80°C. A total of 1800 guts of *N. ephratae* were collected in 100 gut lots. Additional 60 guts in three lots were collected for microbial biomass quantification (qPCR).

V.4.2. Substrates

V.4.2.1. Technical lignins

Two different technical lignins were received in the framework of the project. Their composition is presented in Table 1.

Table 1: Composition of the different available technical lignins

Lignin	**Original biomass**	**Process**	**Klason lignin (%)**	**%H** **%G** **%S** **(% total monolignols)**	**% glucose** **% hemicelluloses**
GV03	Mix wheat/sarkanda grass	Soda	89.4	1.2 50.1 48.7	0 2
FU01	Wheat straw	Procethol 2G	54.9	2.5 46.8 50.7	33,9 4,8

V.4.2.2. Wheat straw (WS) and lignin enriched wheat straw (LWS)

Wheat straw from the winter wheat variety Koreli was collected at an experimental farm (INRA, Boissy-le-Repos, France) in August 2011. After harvesting, the straw was milled to 2 mm and stored at room temperature (20–25°C) (Auer et al., 2017). The dryness content was determined to be higher than 95%.

V.4.2.3. LWS production by enzymatic pretreatment of wheat straw

Cellic Ctec2® (Novozymes) and Pentopan Mono BG® (Novozymes) were used to remove the polysaccharide fraction of wheat straw (WS)

The enzymatic pretreatment was carried out in a 2L Sartorius A (Sartorius, Germany) bioreactor containing 100 g of wheat straw (WS) in acetate buffer (0.1M, pH 4.8; 45°C). Stirring was fixed at 300 rpm during 2 hours until a homogeneous suspension was obtained. Next, 200 ml of the enzymatic cocktail (40 U Cellic Ctec2® + 96 U Pentopan Mono BG® per gWS) were added and the stirring set at 250 rpm. After 48h incubation, the solid fraction was settled overnight, removed and dried (55°C). This pretreated WS was then milled to 0.5 mm (Rosh) and pre-treated again. The final solid, hereafter named lignin-rich wheat straw (LWS) was rinsed with distilled water, dried (55°C) and stored at ambient temperature until use.

To enrich the lignin fraction on WS, an enzymatic pretreatment was performed. To do so, several enzymatic cocktails were considered in order to remove cellulose and hemicellulose from WS (Table x) and select the most efficient enzymatic treatment. Saccharification of a 50g/l suspension of wheat straw was carried out at 48°C, pH 4.8 and agitation 250 rpr, using 4 enzymatic cocktails suspended in acetate buffer. These cocktails consisted of 25 U/g of CTec2 + 100 U/g of either Pentopan MonoBG (Novozyme), Viscozyme (Sigma) or a xylanase from *T. lagnuginosus (Sigma)*. As shown in supplementary Figure 2, the cocktail containing a xylanase from *T. lagnuginosus* plateaued quicker than any other cocktail. However, only 7 g/l of glucose equivalent was released in the supernatant (14% of the total glucose equivalent) which was not better than Ctec2 only. The cocktail containing Viscozyme did similar results than Ctec2 after 72h of saccharification. The cocktail containing 25U/g Ctec2 + 96 U/g Pentopan MonoBG displayed the best conversion, releasing 11.2 g/l glucose equivalent (22.4% of the total glucose equivalent). This cocktail was chosen for wheat straw pretreatment. After this pretreatment, the wheat straw was recovered and dried (48h, 55°C). This dried wheat straw was milled to 0.5 mm. This size was chosen in order to improve the enzymatic conversion of polysaccharides into monomeric sugars (G. G. D. Silva et al. 2012) and prevent eventual adverse effects on bacterial activity by an extensive size reduction (Lazuka et al. 2017). A second enzymatic pretreatment

was performed under the same conditions described above. This second pretreatment resulted on the release of up to 4 g/l glucose equivalent (Supplementary data 3). Overall the hydrolysis yields (ca. 20%) presented here are comparable to other studies on untreated wheat straw (X. Liu et al. 2015; Lazuka et al. 2017; Zheng et al. 2018).

After the first and second steps of enzymatic pretreatment, the composition in cellulose, hemicellulose and lignin in the pretreated substrate was measured by acid hydrolysis and Klason method (Table 2). A clear compositional difference was observed between initial wheat straw and wheat straw after one enzymatic pretreatment. Cellulose, hemicellulose content were reduced by 15% and 3%, respectively, while lignin content increased by 34% going from 22.7 ± 2.0 to 30.4 ± 2.9 g/100g CWR. The second step of pretreatment did not improve much upon this first pretreatment step, but it still managed to port the composition of wheat straw to 36.1 ± 1.5 g/100g CWR cellulose, 30.2 ± 1.8 g/100g CWR hemicellulose and 33.6 ± 2.0 g/100g CWR lignin which consists in a 48% of total increase in lignin content on the final substrate referred as to lignin enriched wheat straw (LWS). Consistently, the saccharification tests of LWS compared to the initial WS showed a significantly lower saccharability (Table 2). LWS is thus a lignin-enriched wheat straw with a reduced amount of easy-available polysaccharides and an increased fraction of lignin.

Table 2: Lignocellulose compositional analysis of wheat straw before and after the two steps of pretreatment. Values are expressed as means ± SD g per 100 g cell wall residues (g/100 g CWR). Values in parentheses indicate percentage change compared to undigested original wheat straw. Statistical differences with WS are marked with * ($p < 0.05$, $n = 3$, Student's t-test).

	Cellulose	**Hemicellulose**	**Lignin**
WS (g/g CWR)	42.6 ± 1.8	35.1 ± 1.5	22.7 ± 2.0
First pretreatment (g/g CWR)	35.5 ± 3.6 (-17%)	33.9 ± 0.3 (-3%)	30.4 ± 2.9* (+34%)
LWS (g/g CWR)	36.1 ± 1.5*	30.2 ± 1.8*	33.6 ± 2.0*

	(-15%)	(-14%)	(+48%)

V.4.3. Bioreactor experiments for technical lignin degradation

The degradation of technical lignins by termite gut microbiomes was assessed using an in house-designed 400 ml bioreactor (Supplementary data 1) using a mineral medium (MM) consisting in (per litre) : 3.21g KH_2PO_4, 1.57 g K_2HPO_4, 0.4g de NH_4Cl, 0.5g NaCl, 0.16g $MgCL_2.6H2O$, 0.09g $CaCl_2.2H2O$, 250µL of V7 vitamin solution and 1ml of trace micronutrient solution containing (per litre): 300 mg H_3BO_3, 1.1g $FeSO_4.7H_2O$, 190mg $CoCL_2.6H_2O$, 50 mg $MnCL_2.4H_2O$, 42 mg $ZnCL_2$, 24 mg NiCL2.6H2O, 18 mg $NaMoO_4.2H_2O$ and 2 mg $CuCL_2.2H_2O$. The mineral medium was filter sterilized with a 0.2µm Steritop device.

The bioreactors were autoclaved (121°C, 20 min, 1.1 bar) before the medium was added. 600 mg of technical lignin were then added as the sole carbon source to achieve a 2 g/l final concentration.

For bioreactors inoculation, gut samples were thawed at 4°C, centrifuged (7,197 x g, 10min, 4°C) and the saline solution was eliminated. 100 guts were used to inoculate 300 ml of mineral medium containing each technical lignin.

The temperature was maintained at 30°C by water circulation and the pH was controlled at 7 by addition of 2M NaOH and 3M HCl. These conditions were chosen to stay close from termite gut real conditions (Brune, 2014). Stirring was fixed to 300 rpm. Aerobic conditions were ensured by the continuous injection of 150ml filtered air/min. For each technical lignin, bioreactor experiments were conducted in duplicate and compared to a control reactor without inoculation.

10ml samples were taken after 0, 7, 14 and 28 days of incubation. These samples were used to analyze lignin degradation by COD measurement, Folin method and microscopy. Complementary analysis were performed by INRA Versailles on samples taken at the initial and final incubation times.

V.4.4. Termite gut microbiome enrichment

The enrichment process of the termite gut consortia was conducted in 250 ml baffled flasks. The substrate used in this experiment were either wheat straw (WS) or lignin-enriched wheat straw (LWS). The flasks containing 2.2 g of substrate (20g/l) were autoclaved (20 min, 121°C, 1.1 bar) prior to adding the mineral medium (MM). In the case of WS, the flasks were autoclaved twice one day apart to ensure a good sterilization. 100 ml MM was added to the substrate.

Gut samples originating from either *Nasutitermes ephratae* (NE) or Microcerotermes *parvus* (MP) were thawed at 4°C, centrifuged at 7,197g, 10min, 4°C and the saline solution was eliminated and replaced by the same volume of mineral medium. The gut suspension was mixed and 10 ml (50 guts) were used to inoculate 100 ml of mineral medium (composition described at 2.3.1.1.).

Each culture was conducted in triplicate for 14 and 28 days at 30°C, 120 rpm and compared to a non-inoculated control flask. After 14 (28) days, 10 ml of the flask suspension was used to inoculate a new flask containing the substrate and the MM. The remaining volume was used to characterize the substrate degradation. This process was done over 5 successive cycles of cultivation. The sixth cycle was used to generate bacterial biomass. The three triplicates for each conditions after the sixth cycle were pooled to generate the inoculum used to inoculate 2L bioreactors.

The global experiment plan is presented in Supplementary data 2.

V.4.5. Chemical analysis

V.4.5.1. Volatile solid (VS) determination

After 7 (14) and 14 (28) days, 15 mL were taken from the flask while maintaining sterile conditions and dried for 48h at 55°C. The dried matter was weighted to determine volatile solid content (VS).

V.4.5.2. Technical lignin analyses

Thioacidolysis was performed on the methanol/water washed samples according to the protocol of Lapierre et al (1993).

Pyrolysis-GC-MS analyses were performed as described in Voxeur et al. (2017).

For Chemical oxygen demand (COD), 1.25 ml samples were diluted to 5ml with distilled water. 2 ml of the diluted solution were added to HI 93754B-25 MR COD kits (HANNA instrument). Tubes were then heated for 2 hours at 150°C. COD was then measured by absorbance at 620nm with a Portable datalogging spectrophotometer DR/2010 (Hach); programme 435 COD HR. Lignin concentration was calculated against standards.

HPSEC analysis was performed on the ethyl acetate extracts after drying and dissolution of the dried extract in THF, then filtration (0.45 µm acrodisc GHP filter) using the following two different columns: 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, mixed-C pore type) and 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, 100Å). THF (1 mL min-1) was used as eluent and detection performed at 280 nm. The mixed C column allows to get the overall molar mass distribution of the sample whereas the 100 Å one has higher resolution in the lower molar mass ranges. All the chromatograms were normalized in time scale according to the peak of toluene, injected as standard. Only chromatograms obtained with the 100 Å are shown here since they were found more informative.

LC-MS was performed in the following conditions: aliquots of the extracts were evaporated to dryness and dissolved in methanol then ultrafiltered (0.45µm, GHP Acrodisc, Pall Gelman, Merck, Molsheim, France) before injection onto an UHPLC apparatus (Thermo Fisher Scientific) combined with an electrospray ionization mass spectrometer (ESI-MS) and photodiode array (PDA) co-detection. UHPLC analysis was performed using a C18 column (2.7 µm, 50 mm x 2 mm I.D.mm; Highpurity, Thermo Fisher Scientific), a 12–95 vol.% aqueous acetonitrile, 1% HCOOH gradient (Millipore, ST Quentin-en-Yvelines, France) for 30 min and a 1 mL.min-1 flow rate. Negative-ion ESI-MS spectra (120–2000 m/z) were acquired using a quadrupole–time-of-flight (Q-TOF) spectrometer (Impact II, Bruker, Leipzig, Germany) setting needle voltage at 4 kV and desolvating capillary temperature at 350°C.

V.4.5.3. Acid hydrolysis and Klason lignin determination

Wheat straw composition was determined using the sulfuric acid hydrolysis method described in (Dumond et al. 2021) on triplicate 80 mg samples. The cooled reaction mixture was quantitatively vacuum-filtered through a preweighed 30 mL porosity 3 filtering Pyrex crucible (Fischer Scientific, Illkirch, France) fitted with a type GF/A glass microfiber filter (WhatmanFisher, Illkirch, France). The insoluble residue was washed with water and dried at 105 °C overnight to determine Klason lignin content. The soluble fraction was filtered and monomeric sugar concentration was determined using the HPLC analytical protocol described by Monlau et al. (Monlau et al. 2012) on an Ultimate 3000 Dionex separation system equipped with a BioRad Aminex HPX 87H affinity column and a refractive index detector (Thermo Scientific). The insoluble fraction was filtered on a sintered filter, washed, dried at 105°C for 24h and weighted to determine Klason lignin content.

V.4.5.4. FTIR analysis

Volatile solid was subjected to FTIR analysis. 10 measurement by sample were done on a Nicolet ATR FTIR (Fisher Scientific). Data were analysed with a custom R program.

V.4.6. Diversity Analysis

V.4.6.1. DNA/RNA extraction and purification

For DNA extraction, 1.5ml samples were taken and centrifuged (13,000 g, 5min, 4°C), the supernatant was removed and the pellet was snap frozen in liquid nitrogen. The samples were stored at -80°C before DNA extraction. Total DNA were extracted using an RNeasy PowerMicrobiome kit (Qiagen) following the manufacturer's instruction but omitting the final DNAse step. Cell lysis was performed with a FastPrep (MP Biomedicals) (2X30s at $4m.s^{-1}$). DNA was purified using an AllPrep DNA/RNA Mini kit (Qiagen) following the manufacturer's instructions. DNA concentrations were measured by Nanodrop 1000 spectrophotometer (Thermoscientific) measuring

absorbance at 260nm and 280 nm. DNA purity was assessed by electrophoresis on a 0.8% agarose gel.

V.4.6.2. Amplicon sequencing

Ilumina sequencing of the V3-V4 region of 16S rRNA genes was performed after PCR amplification using the bacterial primers 343F and 784R, modified to add sequencing adaptors during a second PCR (343F = 5'-CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC GGR AGG CAG CAG-3' , 748R=5'-GGA GTT CAG ACG TGT GCT CTT CCG ATC TTA CCA GGG TAT CTA ATC CT-3'). Library preparation was performed as previously detailed (Auer et al., 2017) and loaded on a MiSeq Ilumina cartridge using reagent kit v3 (paired 300 bp reads).

Sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France using a MiSeq Illumina for 16S rRNA gene sequencing and a HiSeq3000 Illumina for whole genome sequencing.

V.4.6.3. Sequence analysis of metabarcoding datasets

16S rRNA gene sequences were analyzed using the FROGS pipeline (Find Rapidly OTU with Galaxy Solution) implemented on a galaxy instance (http://sigenae-workbench.toulouse.inra.fr/galaxy/) (Escudié et al. 2017). In summary, paired reads were merged using FLASH (Magoč & Salzberg 2011), then, primers/adapters were removed with cutadapt and a de novo clustering was performed using SWARM (Mahé et al. 2015) with an aggregation distance $d = 3$ after denoising. Chimera were removed using VSEARCH (Rognes et al. 2016). Singletons were filtered from the dataset and affiliation was done using Blast+ against the Silva database (release 132) for 16S rRNA gene amplicons. OTU tables were then formatted in standard BIOM for data analyses. OTU tables, taxonomic affiliations and bioreactors metadata were imported into Phyloseq v1.26.1 (McMurdie et Holmes 2013b) using R v3.5.3.

V.4.7. Data analysis

Subsequent data analysis were performed on R v3.5.3 (R core team, (2019)). Hierarchical clustering and heatmaps were done using Vegan (Oksanen, 2019, v2.5.6). PCAs and PLSDAs were done with the MixOmics package(Rohart et al. 2017)v6.6.2). All analysis were performed using the Tidyverse package v1.3.0.

V.5. RESULTS AND DISCUSSION

V.5.1. Lignin degradation potential of *N. ephratae* gut microbiome

The capacity of termite *N. ephratae* gut microbiome to degrade lignin was assessed using as substrate two types of technical lignins FU01 and GV03, raw wheat straw (WS) and a lignin-enriched wheat straw (LWS) obtained by enzymatic exhaustion of its polysaccharide content (see Materials and Methods section). The lignin content measured in the experiments inoculated with the termite gut microbiome was compared to those obtained in non-inoculated control experiments and expressed as a percentage of lignin removal (Figure 1).

No significant lignin removal was observed when technical lignins (GV03 or FU01) were used as substrate (Figure 1.a.). Further characterization of these lignins by thioacidolysis (Rolando et al., 1992) showed that GV03 and FU01 displayed different yields in lignin monomers (H+G+S, $\mu mol.g^{-1}$) of 89±5 and 101±7, respectively, and S/G ratio of 0.81±0.01 and 1.02±0.00, respectively (Figure 1.b.). However, compared to the non-inoculated control experiments and irrespective the incubation period, no significant changes on the monomer fractions was observed due to termite gut microbiome activity.

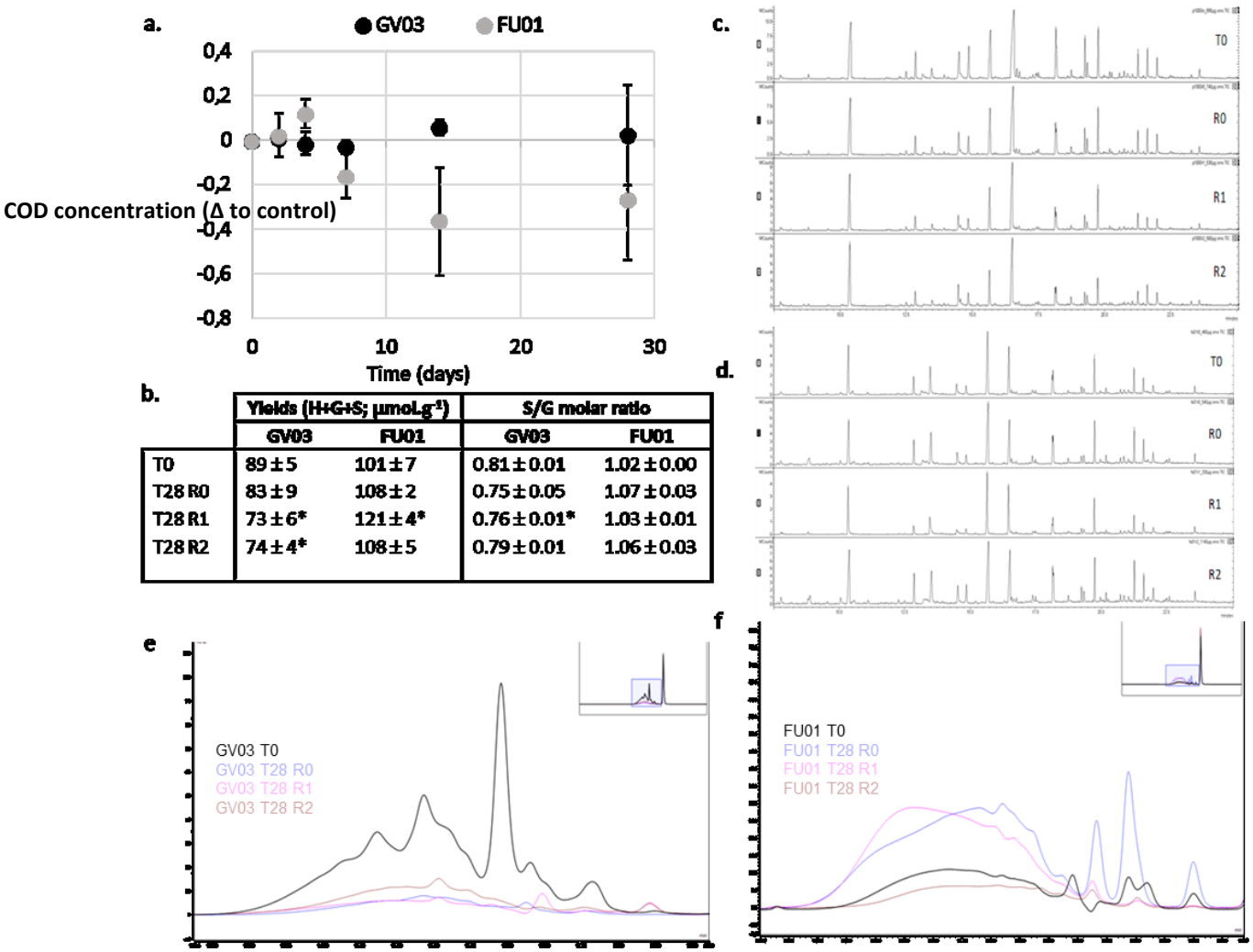
Pyrolysis-GC-MS analyses (Voxeur et al. 2017) performed in the solid fraction of GV03 and FU01 samples (Figure 1 c and d) further indicates that the gut microbiome treatment did not induce significant structural changes in the lignin polymer fraction. The apparent evolution of the pic at 18.25 min in the chromatograms was mainly an artefact due to the influence of the quantity of sample injected. Other peaks did not

change, irrespective the incubation time. Thus, no significant modification of the solid fraction of technical lignins was induced by the microbial activity.

Nevertheless, in depth analysis of the liquid fraction of these technical lignin experiments by HPSEC enable to shown some activity of the gut microbiome. Although the chromatograms relative to GV03 (Figure 1 e) indicated high qualitative and quantitative variability within replicates, the disappearance of the peak at 16.4 min and the modification of the peaks between 17 and 18 min for all the 28 days-samples was observed. This result suggests that some phenolic monomers of C_6C_3 and C_6C_1 types were degraded by the termite gut microorganisms. Similar observation was made with FU01 (Figure 1 f), for which changes in the oligomers profiles (13-16 min) were evident.

Further characterization by LC-MS analysis of the ethyl acetate extracts of the supernatant of these experiments, allowed to detect only phenolic monomers (Supplementary Figure 1). The identified monomers corresponded to those observed in the analysis of the ethyl acetate extracts of GV03 (*p*-OH benzaldehyde, syringic acid, vanillin, *p*-coumaric acid, syringaldehyde, ferulic acid, acetosyringone) and FU01 directly suspended in the culture medium (mainly *p*-coumaric acid and ferulic acid). Although *p*-coumaric acid and ferulic acid and vanillin totally disappeared from GV03 in the non-inoculated experiments, indicating abiotic degradation, no monomers were detected in the gut-microbiome inoculated experiments. This suggested that acetosyringone, syringaldehyde and syringic acid from GV03 were degraded by the termite-gut microorganisms. In the case of FU01, were *p*-coumaric acid and ferulic acid were degraded abiotically, as in GV03, no other monomer could be detected with or without inoculation. Thus, in that case, no specific effect of the gut microbiome could be evidenced.

Overall, chemical analysis showed no-significant effect of the termite-gut microorganisms on the technical lignins studied. It was however found that this microbiome had the ability to metabolize aromatic compounds such as acetosyringone, syringaldehyde and syringic acid.



	Yields (H+G+S; µmol.g^{-1})		S/G molar ratio	
	GV03	FU01	GV03	FU01
T0	89±5	101±7	0.81±0.01	1.02±0.00
T28 R0	83±9	108±2	0.75±0.05	1.07±0.03
T28 R1	73±6*	121±4*	0.76±0.01*	1.03±0.01
T28 R2	74±4*	108±5	0.79±0.01	1.06±0.03

Figure 1: a. Variation of DCO concentration against the blank with time (g/l) b. Thioacidolysis yields (total H+G+S; µmol.g^{-1}) and S/G molar ratio; c. Comparison of the Pyrolysis-GC-MS chromatograms relative to GV03; d. Comparison of the Pyrolysis-GC-MS chromatograms relative to FU01; e. Comparison of the HPSEC chromatograms relative to GV03 (column 100 Å); f. Comparison of the HPSEC chromatograms relative to FU01 (column 100 Å). * indicates a significant difference to T0. No difference was found compared to T28R0 (t.test against T0, $p<0.05$).

V.5.2. Enriching the lignocellulolytic potential of *N. ephratae* gut microbiome by sequential batch culture process

Based in the previous observations, the enrichment of the microbial community fraction of *N. ephratae* gut microbiome able to degrade lignin was performed using either WS or LWS. Two series of independent experiments were performed using 15 or 28 days of incubation. For each enrichment cycle, 10% (v/v) of the previous culture, was used as inoculum for the new cycle and this process was repeated four more times based in previous observations of consortia stabilization (Lazuka et al., 2015; 2018).

Removal of cellulose, hemicellulose and lignin at the end of each enrichment cycle was determined by comparison to the non-inoculated control using acid hydrolysis and the Klason method (Figure 2).

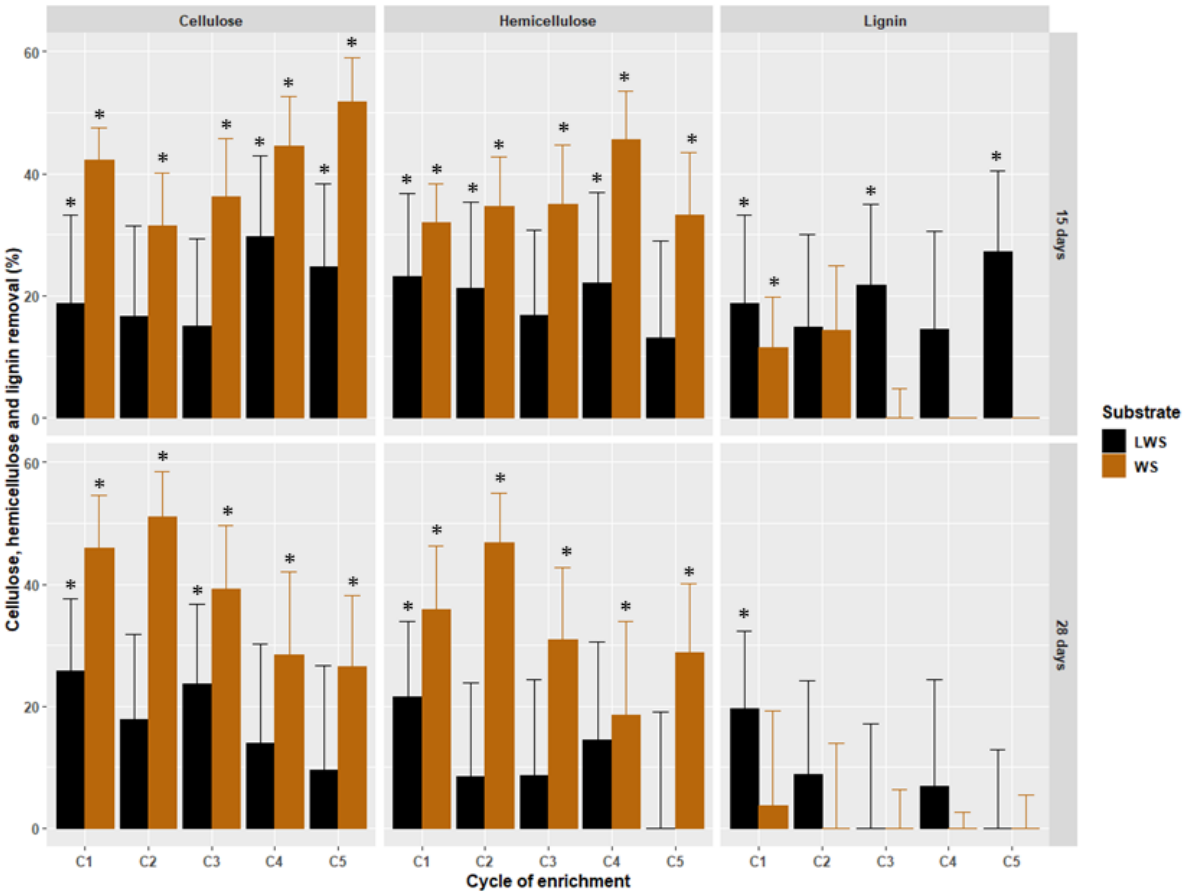


Figure 2: Cellulose, hemicellulose and lignin removal (%) across five cycles of enrichment (C1-C5) of microbial consortia derived from *N. ephratae* using WS or LWS as the substrate. * indicate significant differences compared to the control (t.test, $p<0.05$).

Cellulose was the polymer that was most effectively removed along the enrichment process with, for both incubation periods, a cellulose removal ranging from 10 to 30 % ($p>0.05$) for LWS, and from 28 to 52% ($p>0.05$) in the case of WS (Figure 3). Hemicellulose was also efficiently removed in these experiments, particularly in experiments incubated for 15 days, with an average hemicellulose removal of 18% for LWS and 35% for WS. Surprisingly, for both substrates, the hemicellulose removal levels were lower when experiments were incubated for 28 days, with average values of 10% for LWS and 30% for WS. For both incubation periods, the lower polysaccharides removal observed in LWS experiments compared to those of WS reflects the depletions of the most easily accessible polysaccharides in LWS induced by the enzymatic treatment. Lignin removal was mainly observed for LWS when experiments were incubated for 15 days, with a maximum of 25% lignin removal. Although significant lignin removal was observed when incubation was extended to 28 days (about 20% removal), this degradation capacity appeared to be lost in the successive culture transfers. In WS experiments, lignin degradation only occurred in experiments with the shortest incubation period. Interestingly, while consortia grown on WS displayed the highest cellulose and hemicellulose removal rates, they did not remove lignin from the substrate. Indeed, only in the first cycle of enrichment incubated 15 days a significant lignin removal was measured. It suggest that the consortium was selected for its capacity to hydrolyze readily available polysaccharides but that their presence did not favor the selection of lignin-degrading microorganisms.

When looking at trends across successive cycles, it appears that while the removal potential of consortia with 15 days cycles stayed unchanged or even increased with each cycle, consortia enriched with 28 days incubation periods tended to have a reduced lignocellulose degradation capacity. This seems especially true for LWS experiments. While they displayed similar polymer removals during the first cycle, the degradation capacity dwindled until becoming inexistent after the third enrichment cycle. The cellulose and hemicellulose removal of experiments incubated 28 days using WS was also lower compared to experiments incubated 15 days.

In order to obtain further information on the structural features of theses lignocellulosic substrates along the enrichment process, samples from the first and the fifth cycle of enrichment were analyzed by FTIR (Figure 3).

FTIR gives access to cellulose crystallinity through the calculation of the lateral order index (LOI). All consortia induced a decrease of LOI, indicating a decrease of the cellulose crystallinity and proving the effectiveness of these consortia on crystalline cellulose. It should be noted that LWS displayed a significantly higher LOI than WS (Supplementary Figure 4). This indicated that LWS polysaccharides may be more difficult to degrade as cellulose crystallinity has been linked to lignocellulose recalcitrance to degradation (Zoghlami & Paës 2019). Thus, the strongest decrease of LOI observed in LWS compared to WS might reflects a stronger capacity to remove crystalline cellulose of the consortia enriched in LWS; probably favored by the lower fraction of amorphous cellulose present in this substrate.

Regarding the lignin/holocellulose ratio, consortia enriched in WS showed an increase on this ratio compared to the original substrate. It reflects the preferential degradation of WS polysaccharides over lignin. In contrast, only a slight increase of this ratio was observed in LWS experiments, suggesting that both lignin and polysaccharides were partially removed from LWS. These observations are consistent with the wet chemical analysis data described above. Interestingly, LWS experiments incubated 28 days showed a significant increase in this ratio, in agreement with the chemical analysis that confirmed that lignin degradation capacity was lost during the enrichment process for this incubation period.

Finally, the S/G ratio was found to be slightly lower in the consortia obtained in LWS experiments, especially when a 15-days enrichment period was applied. No significant differences in S/G ratio were found for consortia grown on WS.

Based on these results, the microbial consortium obtained with LWS and a 15-days incubation period for the enrichment cycles was selected as the best consortium to degrade lignin in the conditions applied here.

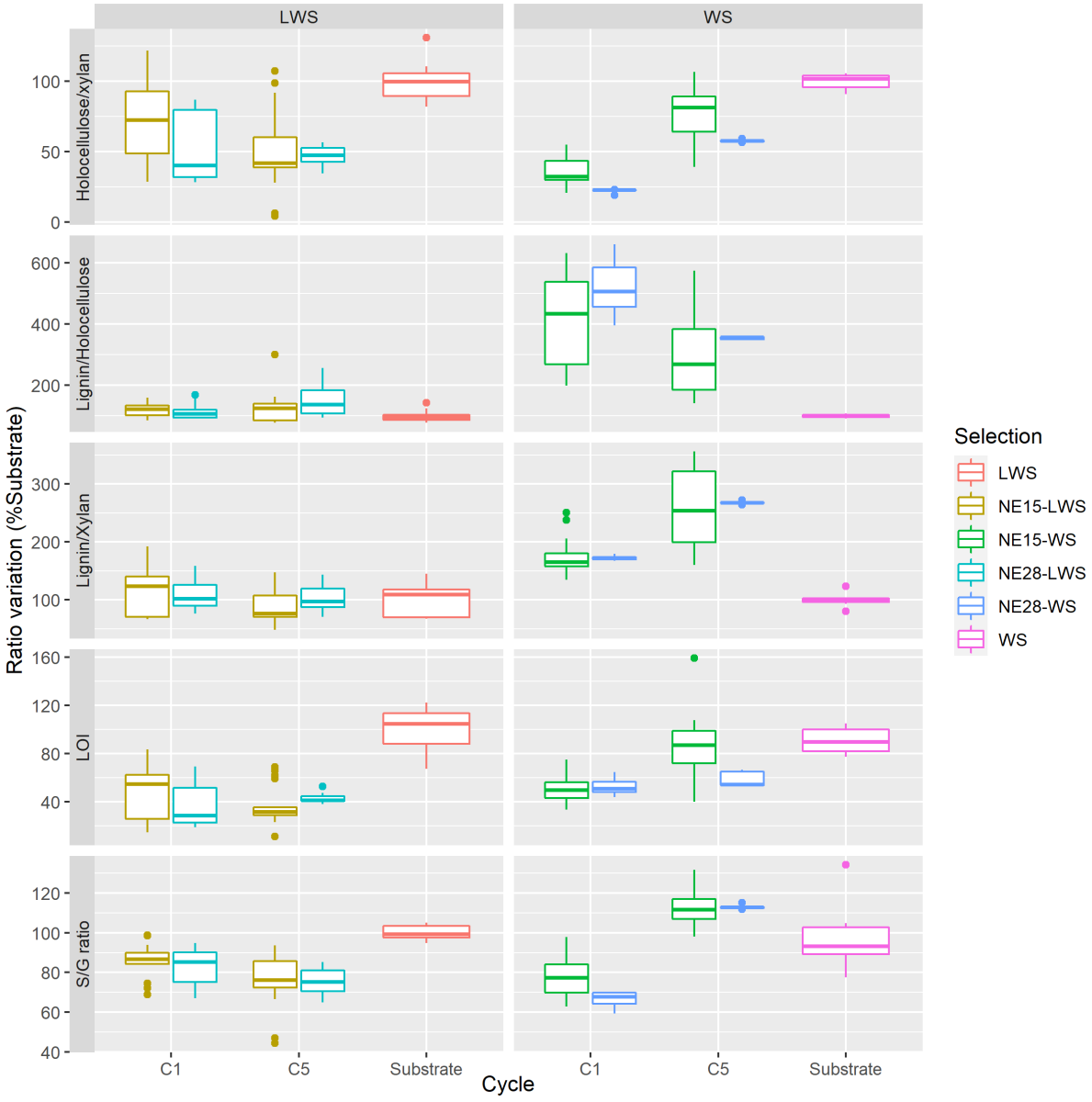


Figure 3: Variation in peak ratios (%substrate) quantified by ATR FTIR. Holocellulose/xylan (1225/1735cm-1), lignin/holocellulose (1512/1375 cm-1), lignin/xylan (1512/1730cm-1), LOI (Lateral Order Index, 1430/898cm-1) and S/G ratio (1327/1269cm-1).

V.5.3. Microbial consortia diversity and dynamics along the enrichment process

Microbial community diversity was determined in the initial gut inocula and the final-samples of each cycle of the consortia enrichments on LWS and WS using16S rRNA gene amplicon sequencing. The termite gut microbiome used as inocula for these experiments displayed the larges richness (270 OTUs) (Supplementary Table 2) and it was mainly composed by Spirochaetes and Fibrobacteres phyla (Supplementary figure 2) and a minor proportion (about 10% in total) of Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria. In contrast, the diversity of the consortia enriched in either LWS or WS, and irrespective the incubation period, displayed a lower richness (with 177-212 and 89-139 OUTs for LWS and WS, respectively) than the initial termite gut inoculum and have no discernable similarity with it. Both Spirochaetes and Fibrobacteres were lost since the first cycle of enrichment (Figure 4). The consortia enriched on both substrates showed a dominance of Proteobacteria (34-68%) and Bacteroidetes (13-39%), followed by Firmicutes, Actinobacteria and Gemmatimonadetes. This drastic shift in microbial diversity can be explained by the strong change imposed by the culture conditions, passing from a predominantly anaerobic termite gut system (Brune et al., 1995) to an aerobic one. Indeed, most Spirochaetes and especially those from the *Treponema* genus that were prevalent in *N. ephratae* gut microbiome are anaerobic (Smibert 1981). The same observation can be made for Fibrobacteres and *Bacteroides* (Thomas et al. 2011), the main representative genus of the Bacteroidetes phylum found in *N. ephratae* gut. This shift may also reflect the adaptation of the termite gut microbiome to the lignocellulosic substrates, favoring the selection of populations involved in lignocellulose degradation under the imposed culture conditions. Indeed, previous studies have shown that even when maintaining the conditions as close as possible to initial gut conditions, major changes were observed when culturing the termite gut microbiome under laboratory conditions, including the loss of Fibrobacteres and Spirochaetes (Lazuka et al. 2018). Similar changes in community diversity were also reported when enriching lignocellulolytic consortia from different inocula origins and using different lignocellulosic substrates (Eichorst et al. 2013; Allgaier et al. 2010)

Changes of the community composition were less drastic during the further steps of enrichment and the observed richness at the end of the enrichment process was not significantly different from the first cycle (Supp. table 2). The dominant phyla enriched in both substrates were Proteobacteria (55-68%) and Bacteroidetes (16-18%) followed by Firmicutes and Actinobacteria.

We compared the evolution of the abundance of genera across the enrichment process for NE15-LWS, NE28-LWS, NE15-WS and NE28-WS (Figure 4). While on LWS-cultivated consortia, the composition was similar with 55-68% of Proteobacteria, 16-18% Bacteroidetes and lower abundance of Firmicutes and Actinobacteria, it was different from WS-cultivated consortia. In the latter, Firmicutes formed up to 14% of the relative abundance and other phyla such as Gemmatimonadetes were found.

At the phylum level, NE15 consortium remained quite stable across the enrichment process.

At the family level, a progressive replacement of Sphingobacteriaceae by Weeksellaceae and Chitinophagaceae could be observed. An increase in Cellumonadaceae could also be found. Some changes were also found in the Proteobacteria profile with a decrease in Enterobacteriaceae, Moraxellaceae and Xanthomonadaceae, replaced by Pseudomonadaceae, Devosiaceae and Rhizobiaceae. On NE28, though the proportions of Proteobacteria differed, their profile and the evolution of this profile were similar. NE28 consortium evolution was also characterized by the increase in Bacteroidetes over time, Chitinophagaceae, Cytophagaceae and Flavobacteriaceae especially. Also, unlike NE15, an increase in Firmicutes was observed.

Unlike consortia grown on LWS, consortia grown on WS displayed a consistently higher abundance of Firmicutes. NE15-WS Bacteroidetes abundance diminished across cycles, due to the decrease in Weeksellaceae abundance. The profile of the Proteobacteria found also differed from NE15-LWS with a majority of Bulkholderiaceae and Kaistaceae after one cycle and an increase of the latter at the expense of the former across cycles. When during 28 days incubation cycles on WS, more Actinobacteria were found, though their population ultimately decreased after the second cycle. Like NE15-WS, more Firmicutes were found and remained in the

consortia after five cycles. Moraxellaceae was the main Proteobacteria family, followed by Bulkholderiaceae after the first cycle but across the enrichment, similarly to NE15-WS, Kaistaceae increased, as well as Rhizobiaceae.

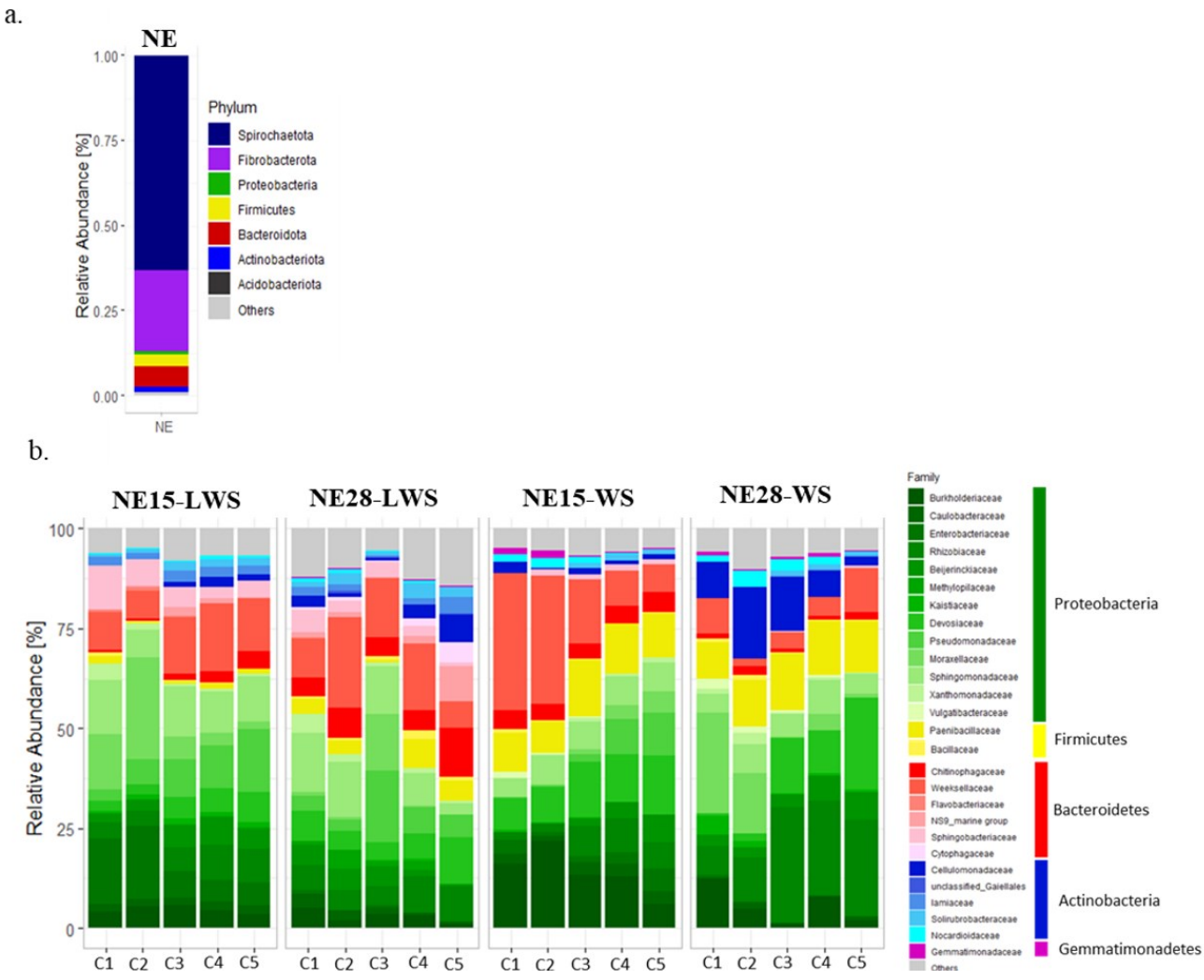


Figure 5: a. Taxonomic composition of the initial NE guts. B. Evolution of the microbial community composition during 5 cycles of enrichment of 6 microbial consortia cultivated on WS or LWS at the phylum and family levels.

V.5.4. Tracing the community members related to each substrates in the enriched consortia

In order to identify the community members distinctive of each lignocellulosic substrate, the diversity of the consortia enriched on LWS and WS at the end of the

enrichment process was compared by multivariate principal component analysis (PCA) on CLR-normalized abundance data at the family level (Figure 5). It appeared that the four consortia were distinct from one another. The first principal component (x- axis) discriminate the samples according to the substrate used for the enrichment while the second principal component (y- axis) discriminate consortia according to the length of cycle incubation. Figure 6 represents the ten families that contribute the most to the difference between consortia. WS enrichments were differentiated by members belonging to unclassified Solirubrobacterales and Steroidobacteraceae when a 28-days incubation cycle was used while Moraxellaceae were distinctive of the enrichment with 15-days incubation cycle. The OPB56, a family often classified as Chlorobi (Hiras et al. 2016) was found to be linked with enrichment on WS, irrespective the incubation time. Regarding the consortia enriched on LWS, the families Cytophagaceae, Cellvibrionaceae, Bacteroidetes_VC2.1.Bac22, Pseudomonadaceae and unclassified members of Flovobacteriales were the most discriminating families of this substrate. Interestingly, Pseudomonadaceae, in which several lignin degrading bacteria have been found (S. Lee et al. 2019) were linked to the enrichment with 15-days incubation cycle which was found to be the best lignin–degrading consortium, potentially highlighting its role in lignin degradation. Additionally, Enterobacteriaceae were linked to consortia with shorter incubation time, using both WS and LWS.

To consolidate these results we used hierarchical clustering to identify clusters of families that were linked to the consortia in the last cycle of enrichment (Figure 6).

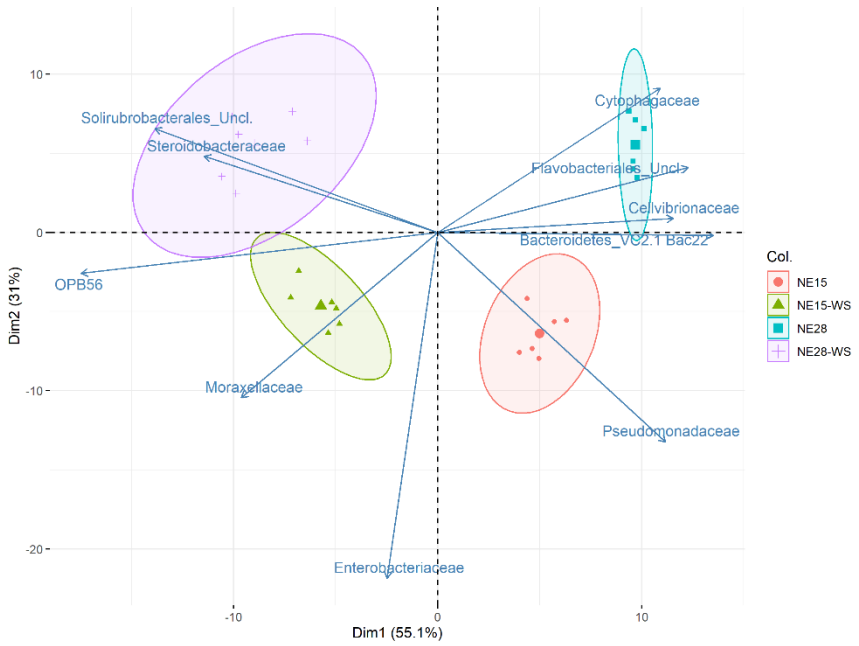


Figure 5: PCA biplot analyses based on 16S rRNA gene sequencing data of the 4 tested consortia along the principal components 1 and 2. Colors and shapes indicate the consortium (slightly larger point indicate the barycenter of the points), arrows indicates the families driving the variance.

Overall, consortia were separated from one another based on the selection process forming 4 distinct groups. We used data from 16S rRNA RNA and DNA sequencing but samples did not cluster along this parameter. A first cluster, composed by Enterobacteriaceae and Moraxellaceae (although to a lesser extend) was associated with consortia with shorter time of incubation (NE15-WS and NE28-WS). This might indicate that these bacteria are only able to grow on the free sugars available at the beginning of the incubation and that their abundance tends to decrease with time. A second cluster was characterized by its slightly lower abundance in NE28. It included Weeksellaceae, Sphingomonadaceae, Chitinophagaceae, Burkholderiaceae, Rhizobiaceae, Devosiaceae and Paenibacillaceae. This cluster is characterized by the diversity of phyla it incompassed with the presence of Bacteroidetes, Proteobacteria and Firmicutes. It showed the core families present in the selection. Though abundances may differ the most abundant taxa are conserved no matter the selection pressure applied. Pseudomonadaceae formed a cluster on its own and was associated

with NE15-WS and NE15-LWS consortia. It was found to be one of the families most associated with NE15-LWS. As several *Pseudomonas* strains have been found to produce dye-decolorizing peroxidases (Ravi et al. 2017; Yang et al. 2018; Salvachúa et al. 2020), known to be a lignin-modifying enzyme, it may play a role in the overall better lignin degradation in NE15-LWS consortium. A fourth cluster was identified with overall high abundances in all consortia. This cluster was not specific to any consortium and thus may not be involved in the differences observed in lignin degradation between the consortia. A fifth cluster composed by Solirubrobacterales Uncl. and OPB56 was highly specific to the use of WS as the substrate. It is interesting to notice that Solirubrobacterales Uncl. were absent of the most lignin degrading consortium as it was suggested that some Solirubrobacterales identified in forest soil could degrade lignin (Wilhelm et al. 2019). A sixth cluster included families that were specific to the use of LWS as the substrate. This cluster included for instance Microbacteriaceae that have been known to contain lignin-modifying or lignin-degrading bacteria (Bredon et al. 2018; Wilhelm et al. 2019). A seventh cluster including Spiromosaceae, Microspecaceae, Gallionellaceae and Steroidobacteraceae was specific to NE15-WS and NE15-LWS. An eighth cluster composed by Nocardiaceae and Kaistiaceae was specific to NE15-LWS and NE28-WS. Finally, a last cluster was identified which contained families with low relative abundance in NE15-LWS.

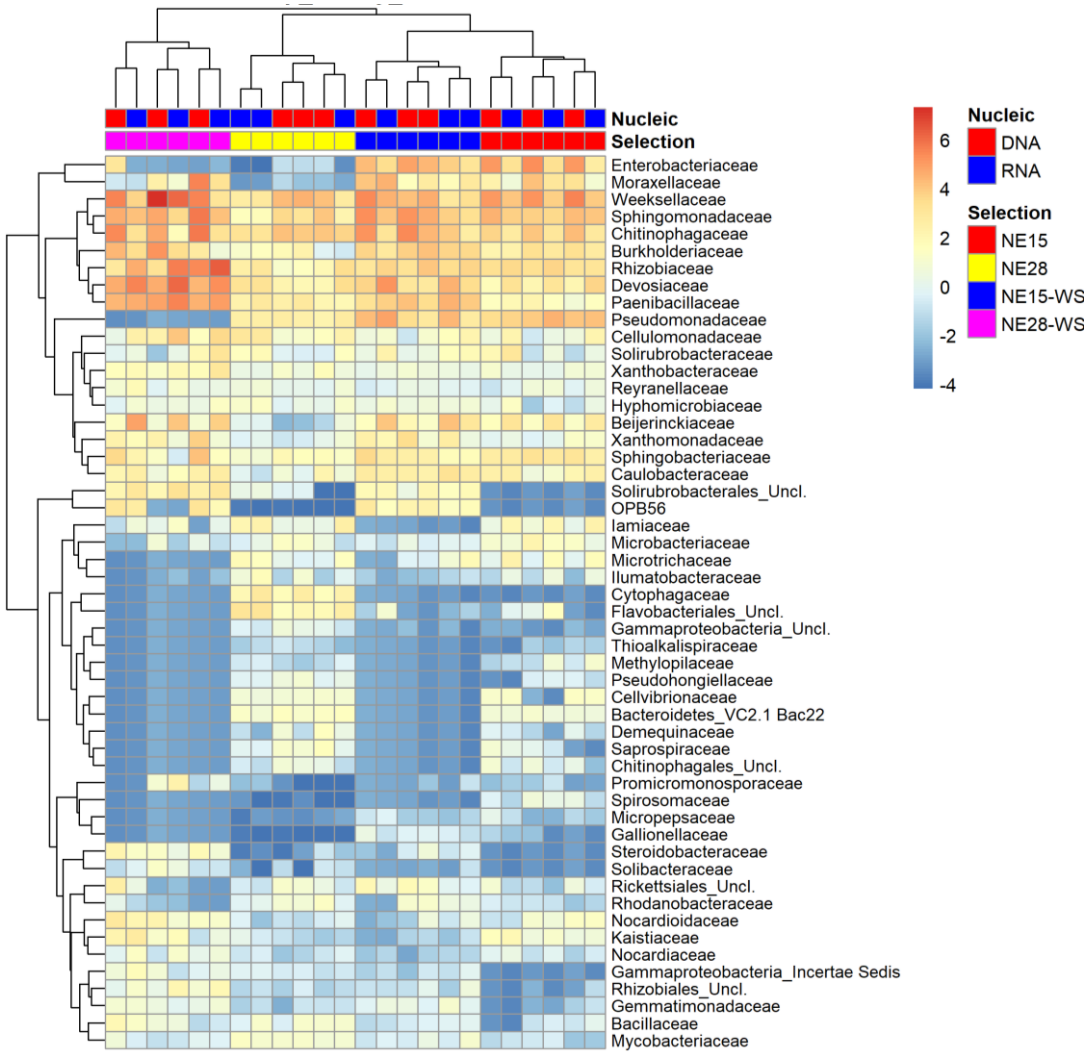


Figure 6: Heatmap and hierarchical clustering on CLR abundances of families with abundance > 1% across samples taken from the last cycle of enrichment. Samples selected in different conditions were clustered using hierarchical clustering and the dendrogram is displayed at the top of a heatmap visualizing the ward distance between samples.

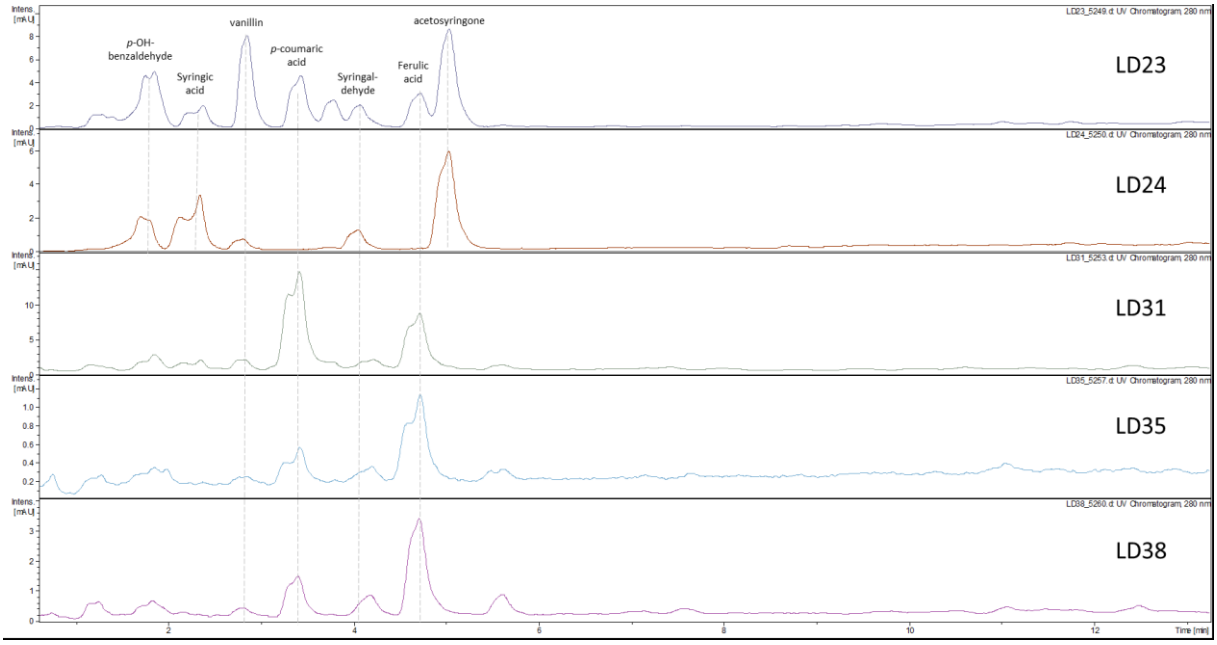
V.6. CONCLUSIONS

Different approaches were taken to assess the lignin degradation capacity of the termite gut microbiome of *N. ephratae* and to successfully enrich a termite derived lignolytic consortium. Though the termite gut microbiota was not able to directly digest technical lignins as sole carbon source, the use of lignin-enriched wheat straw

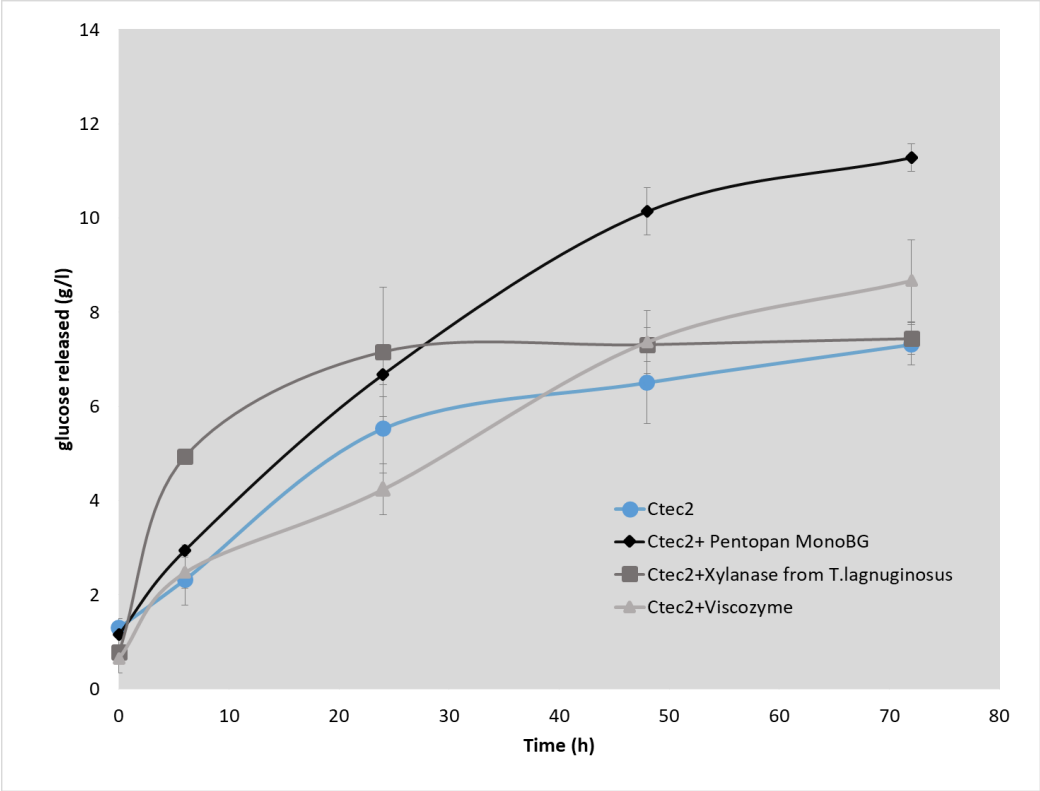
(LWS) was used to select and enrich a lignin-degrading consortium. Four different conditions were compared to select the best lignin-degrading consortium and it was found that using a 15-days incubation cycle with LWS as substrate amounted to the most important lignin loss. In addition, 16S rRNA sequencing was used to gain information on the bacterial diversity of the different consortia and its evolution across cycles of enrichment. Pseudomonadaceae were found to be the Familly the most associated with the best lignin-degrading consortium and is therefore hypothesized to play a major role in the better performance of LWS-15 in term of lignin degradation. Other families were cauterized in order to understand their potential role in the consortium.

Overall, this work represent a first step in selectively enhancing the lignin-degrading capacity of a termite gut microbiota. However, though this work offers a glimpse into the comprehension of lignin degradation by termite gut microbial consortia, efforts should be made into both a better characterization of the structural modifications induced by the consortium and an identification of the bacteria and enzymes responsible for its ligninolytic activity.

V.7. SUPPLEMENTARY FIGURES AND TABLES



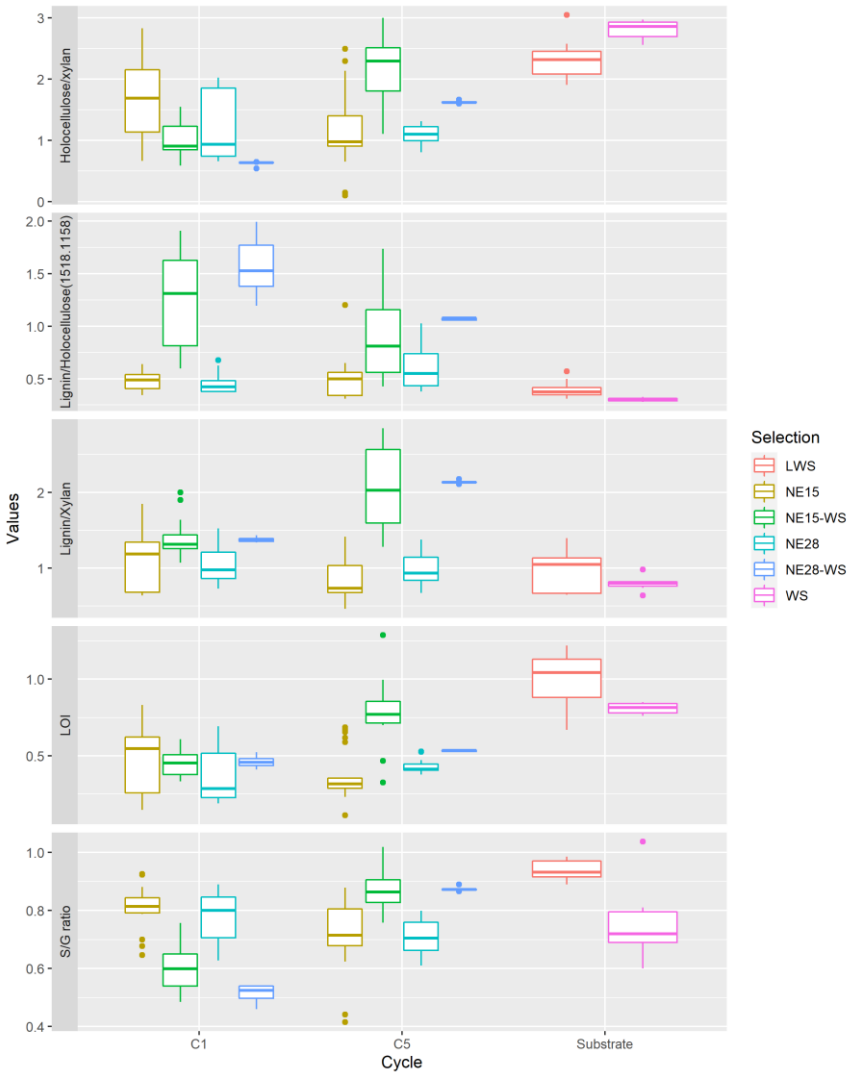
Supplementary Figure 1: ***Comparison of the LC-MS chromatograms relative to GV03 (LD23, LD24), FU01 (LD31) and wheat straw samples (LD31, LD35, LD38).***



Supplementary Figure 2: Kinetics of glucose (eqv) released from wheat straw by the four tested enzymatic cocktails.

Supplementary Table 1: Total reducing sugar release after a 48h saccharification assay (g/100 g). Statistical difference is marked by * ($p < 0.05$, $n = 3$, Student's t-test).

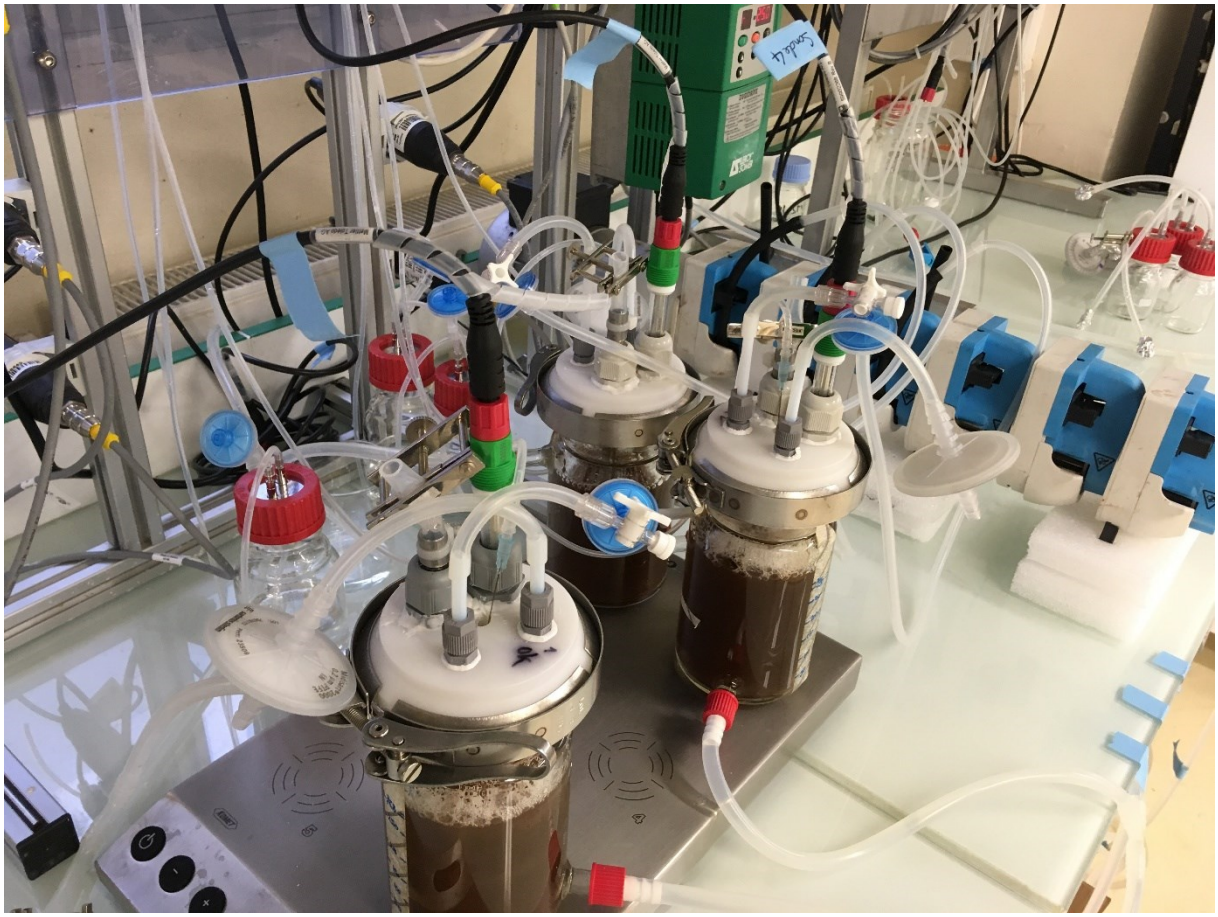
WS	LWS
31.4 ± 4.0	10.4 ± 1.9*



Supplementary figure 3: Variation in peak ratios (%substrate) quantified by ATR FTIR. Holocellulose/xylan (1225/1735cm-1), lignin/holocellulose (1512/1375 cm-1), lignin/xylan (1512/1730cm-1), LOI (Lateral Order Index, 1430/898cm-1) and S/G ratio (1327/1269cm-1).

Supplementary Table 2: α-diversity indexes of the termite gut consortium and after the first and fifth cycle of enrichment process: Observed = number of observed OUTs, Shannon= Shannon index, InvSimpson= Inverse Simpson Index). Groups that are significantly different are marked with abc ($p < 0.05$, $n = 3$, Student's t-test).

	Observed		**Shannon**		**InvSimpson**	
NE gut	270±16 [a]		4.5±0.4		45±3.0	
	C1	**C5**	**C1**	**C5**	**C1**	**C5**
WS 15	197±9 [a]	155±23 [bc]	2.1±0.6	3.1±0.1	3.7±2.1	10.6±0.9
WS 28	119±16 [b]	125±24 [c]	2.9±0.8	2.5±0.8	11.6±9.8	8.1±5.1
LWS 15	186±23 [a]	192±22 [b]	3.2±0.3	3.0±0.0	13.1±3.6	8.4±0.6
LWS 28	189±11 [a]	183±7 [b]	3.3±0.3	3.3±0.1	12.6±6.5	9.8±1.4



Supplementary data: Photograph of the homemade bioreactors for technical lignin degradation

VI. CHARACTERIZATION OF A LIGNIN-DEGRADING MICROBIAL CONSORTIUM DERIVED FROM THE GUT OF THE TERMITE NASUTIERMES EPHRATAE

VI.1. CHARACTERIZATION OF A LIGNIN-DEGRADING MICROBIAL CONSORTIUM DERIVED FROM THE GUT OF THE TERMITE *NASUTIERMES EPHRATAE*

*Louison Dumond[1], Vincent Chassagnac[1], Laurent Cézard[2], Amel Majira[2], Betty Cottyn[2], Stéphanie Baumberger[2], David Sillam-Dussès[3], Edouard Miambi[4], Pui Ying Lam[5], Yuki Tobimatsu[5] ,Gijs van Erven[6,7], Mirjam Kabel[6], Claire Hoede[8], Vincent Lombard[9,10,11], Bernard Henrissat[9,10,11], Guillermina Hernandez-Raquet[a]**

1 Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.

[2] Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000, Versailles, France

[3] Faculté des Sciences et Technologie, Université Paris Est Créteil, Département ECOEVO, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES, Paris), 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France

[4] Université Paris 13 - Sorbonne Paris Cité, Laboratoire d'Ethologie Expérimentale et Comparée, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

[5] Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto, Japan.

6 Wageningen University & Research, Laboratory of Food Chemistry, Bornse Weilanden 9, 6708 WG, Wageningen, The Netherlands.

[7] Wageningen Food & Biobased Research, Bornse Weilanden 9, 6708 WG, Wageningen, The Netherlands

[8] INRAE, UR 875 Unité de Mathématique et Informatique Appliquées, PF Genotoul Bioinfo, Auzeville, CS 52627, 31326 Castanet Tolosan cedex, France

[9] CNRS UMR 7257, Aix-Marseille University, F-13288, Marseille, France.

[10] USC 1408 AFMB, INRAE, F-13288, Marseille, France.

[11] Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

VI.2. ABSTRACT

Lignocellulose degradation by the termite gut microbiota has shown a lot of promises. However, its potential to degrade lignin remains largely untouched. Furthermore, little is known about the temporal evolution of a termite-derived microbial consortium in aerobic conditions and how it could affect lignoellulose degradation.

In this study we combined state-of-the-art analytical methods such as 2D HSQC NMR and ^{13}C-IS py-GC-MS with "shotgun" metagenomics sequencing to characterize a lignin-degrading microbial consortium derived from the gut of *Nasutitermes ephratae* termite. This consortium, named NE15 was cultivated in bioreactors on lignin-enriched wheat straw (LWS). At regular intervals, lignocellulose composition and lignin structure of the substrate was analyzed and the consortium was sequenced.

We demonstrated that NE15 consortium is capable of removing up to 20% of the total lignin and alsmost all soluble oligomeric lignin. We also showed that this removal was accompanied by a decrese in C-C linkages in the lignin and the production of HPV/HPS units. Metagenomics analyses allowed us to reconstruct 174 MAGs and to identify genes and bacteria with a putative capacity to degrade lignin. Finally we highlighted the highly synergistic nature of the consortium with a cooperation between MAGs with polysachharides degradation and MAGs with lignin degradation potential.

VI.3. INTRODUCTION

The necessity to limit our impact on climate change has sparked interest in developing biorefineries based on renewable, bio-based resources for the production of clean energy, fuels and commodity chemicals. Lignocellulose, and especially wheat straw is a sensible resource for biorefinery since it is available in large quantities from agricultural or forestry residues and does not compete for land use against feed

crop(Bentsen et al., 2014). Lignocellulose presents a complex polymeric structure of 30-50% of cellulose, 20-30% of hemicelluloses and 15-30% of lignin (Lapierre et al., 1995; del Río et al. 2012) interwoven with each other. This complexity is at the base of its recalcitrance to bioconversion. In particular, lignin provides plants with mechanical strength and protection against abiotic and biotic aggressions from the environment (Liu et al., 2018). As a result, lignin represents also the main limiting factor to polysaccharide accessibility to enzymes and microorganisms in biorefinery approaches. Lignin is also the most abundant natural renewable source of aromatic compounds for chemical industries. Therefore, there is a need to find efficient solutions to valorize lignin and unlock cellulose and hemicellulose(Kamimura et al. 2019).

Lignin is an aromatic polymer mainly constituted of three phenylpropane units, *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units synthesized from three monolignols, namely, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol(Ralph et al., 2019). In monocotyledon such as wheat, lignin also contains γ-*p*-coumaroylated G and S units, and flavone-derived tricin units(Ralph et al., 2019). In addition, (di)ferulates (FA), mainly ester-linked to hemicelluloses (glucuronoarabinoxylans, GAX), are partially bonded to lignin(Ralph et al., 2019; Karlen et al. 2016).

In recent years, bacterial degradation of lignin has gained attention as it represents a potential solution to this problem. Indeed, although it was recently argued that bacteria do not possess enzymes homologues to well-known lignin degrading enzymes found in fungi such as lignin peroxidase (LiP), manganese peroxidase (MnP) and versatile peroxidase (VP)(Silva et al. 2021; Davis et al. 2013), bacteria have shown to possess another type of peroxidase, known as dye-decolorizing peroxidase (DyP).(Y. Sugano 2009) DyP peroxidase have been identified in various Proteobacteria and Actinobacteria(de Gonzalo et al. 2016) and have demonstrated oxidative properties on model substrate and technical lignins(Rahmanpour & Bugg 2015; Rahmanpour et al. 2016; Lončar et al., 2016). Lignin-modifying laccases have also been identified in *Streptomyces* (Majumdar et al. 2014) and *Bacillus*(Sondhi et al. 2015). Furthermore, bacterial β-etherases have also demonstrated their effectiveness in breaking down

lignin linkages(Masai et al. 1989; Voß et al. 2020; Kontur et al. 2019). Along these, various bacterial enzymes classified as auxiliary activity (AA) in the CAZy database have been proposed for their synergistic action on lignin or oxidation through the Fenton reaction (Silva et al. 2021) .

In nature, however, lignin degradation is rarely performed by single individuals but rather through the concerted effort of complex microbial communities or microbial consortia. Ecosystems such as soil (Brink et al. 2019), the digestive tract of phytophagous insects (Zhou et al. 2017; Couger et al., 2020) or mammals (Ufarté et al. 2018) have provided valuable information about potential lignin-degrading bacteria and their enzymes. Using microbial consortia issued from these environments may provide a far superior metabolic diversity than single species and can help deciphering the cooperative effects of bacteria in lignin degradation.

Amongst phytophagous insects, termites play a crucial role in the decomposition of lignocellulose in natural environments (König et al., 2013). Termites are well known for their high efficiency to degrade plant cell walls (Brune 2014). Strong efforts have been realized to understand the *in vivo* functioning of the gut microbiome of higher termites, including *Nasutitermes* sp., and its relationship with the termite host.(Warnecke et al. 2007; Burnum et al. 2011) Thanks to the use of culture-independent methods such as metagenomics, metatranscriptomics or metaproteomics, several studies discovered the rich bacterial diversity and the vast repertoire of carbohydrate active enzymes (CAZymes) involved in lignocellulose degradation encoded by the symbiotic bacteria associated to these digestive systems and their response to changes in diet composition(Calusinska et al. 2020; Marynowska et al. 2020). In contrast, the way termites and their microbiome overcome the lignin barrier is still poorly understood. While previous studies proposed that termite were not able to degrade lignin, but only mono or dimeric model compounds,(Kato et al., 1998; Ke et al., 2011) the development of powerful analytical tools for lignin analysis, especially the use of multi-dimensional nuclear magnetic resonance (NMR) techniques(Kim & Ralph 2010; Mansfield et al. 2012) or pyrolysis coupled to gas chromatography with mass spectrometric detection (py-GC-MS), have shed a new light onto lignin degradation by termites and their gut microbiome(Hongjie Li et al. 2017; Tarmadi et al. 2018; Dumond et al. 2021).

Nevertheless, deciphering the functioning of the lignin degrading communities is challenging do to the tremendous complexity of these ecosystems. A powerful strategy to overcome the complexity of termite gut microbiome is to simplify the community and enrich the functionally relevant fraction by culture under a defined selection pressure. Such approach enable to minimize the ecosystem's complexity and increase the community fraction expressing a particular functional potential(DeAngelis et al. 2010). This enrichment approach has been successfully applied to environments such as soil(Moraes et al. 2018),(de Lima Brossi et al. 2016) compost, cow rumen,(Lazuka et al. 2015) or termite gut microbiomes(Lazuka et al. 2018). Therefore, the lignin degrading potential of termite gut microbiome could be better assessed by selecting and enriching a microbial consortium with high lignin degrading properties that could be relevant for the biorefinery.

With this aim, in a previous study we have enriched a lignolytic consortium derived from *Nasutitermes ephratae* gut microbiome using lignin rich wheat straw and aerobic conditions. In the current study, using state-of-the-art analytical technics, including 2D HSQC NMR and ^{13}C-IS py-GC-MS and whole metagenome sequencing, we expand our previous work characterizing in detail the dynamic of lignin degradation and reconstructing the metagenomic species involved in lignin transformation of this consortium. Here, we analyze the temporal succession of metagenomic species and genes coding for functions related to plant cell wall deconstruction and provide new insights into the underlying cooperation between bacteria to achieve lignocellulose and lignin degradation.

VI.4. MATERIALS AND METHODS

VI.4.1. Substrate

Wheat straw of the variety Koreli (WS), was milled to 2mm and enzymatically pretreated to be used as substrate. For WS enzymatic pretreatment, Cellic Ctec2® (Novozymes; Sigma Darmstadt, Germany) and Pentopan Mono BG® (Merck, Darmstadt, Germany**)** were used to remove the polysaccharide fraction of WS.

The enzymatic pre-treatment was carried out in a 2 L BIOSTAT® A+, (Sartorius, Germany) bioreactor containing 100 g of WS in acetate buffer (0.1M, pH 4.8; 45°C). Stirring was fixed at 300 rpm during 2 hours until a homogeneous suspension was obtained. Next, 200 ml of the enzymatic cocktail (40 U Cellic Ctec2® + 96 U Pentopan Mono BG® per gWS) were added and the stirring set at 250 rpm. After 48h incubation, the solid fraction was settled, removed and dried overnight (55°C). This pre-treated WS was then milled to 0.5 mm and the enzymatic pre-treatment repeated again. The final solid wheat straw, hereafter named lignin-rich wheat straw (LWS) was rinsed with distilled water, dried (55°C) and stored at ambient temperature until use. Before use in the bioreactor experiments, LWS was sterilized by dry autoclaving at 121°C, 1.1 bar for 20 min.

VI.4.2. Termite gut microbiome enrichment

The details of the enrichment process is available in Chapter V. Briefly, the enrichment process of the termite gut consortia was conducted in 250 ml baffled flasks using LWS as sole carbon source. The flasks containing 2.2 g of substrate (20g/l) were autoclaved (20 min, 121°C, 1.1 bar) prior to adding the mineral medium (MM). MM consisting in (per litre) : 3.21g KH_2PO_4, 1.57 g K_2HPO_4, 0.4g de NH_4Cl, 0.5g NaCl, 0.16g $MgCL_2$.6H2O, 0.09g $CaCl_2$.2H2O, 250µL of V7 vitamin solution and 1ml of trace micronutrient solution containing (per litre): 300 mg H_3BO_3, 1.1g $FeSO_4.7H_2O$, 190mg $CoCL_2.6H_2O$, 50 mg $MnCL_2.4H_2O$, 42 mg $ZnCL_2$, 24 mg NiCL2.6H2O, 18 mg $NaMoO_4.2H_2O$ and 2 mg $CuCL_2.2H_2O$. The mineral medium was filter sterilized with a 0.2µm Steritop device.

Gut samples originating from *Nasutitermes ephratae* (NE) were thawed at 4°C, centrifuged (7,197 x g, 10min, 4°C); the saline solution was eliminated and replaced by the same volume of MM. Fifty guts suspended in 10 ml MM were used as inoculum for 100 ml of medium.

Each culture was conducted in triplicate for 14 days at 30°C, 120 rpm and compared to a non-inoculated control flask. After 14 days, 10 ml of the flask suspension was used to inoculate a new flask containing the substrate and MM. The remaining volume

was used to characterize the substrate degradation. This process was done over 5 successive cycles of cultivation. The sixth cycle was used to generate bacterial biomass. The biomass generated in the three triplicates of this sixth cycle was pooled and homogenized constituting the NE15 consortium. Aliquots (70mL) of NE15 were snap frozen in liquid nitrogen and kept at -80°C until use as inocula for further bioreactor experiments.

VI.4.3. Bioreactor experiments for LWS degradation by NE15 consortium

The degradation of LWS by the termite gut-derived NE15 consortium was carried out in 2 L BIOSTAT® A+, (Sartorius, Germany) bioreactors using MM supplemented with LWS (20g.L^{-1}). The bioreactors were autoclaved (121°C, 20 min, 1.1 bar) before the medium was added. Sterilized MM and LWS were added to the bioreactors under sterile conditions.

Bioreactors were inoculated with 70 ml of frozen samples of NE15 thawed at 4°C.

The bioreactors temperature was maintained at 30°C and the pH was controlled at 7. These conditions were chosen to stay close to the conditions prevailing in the termite gut of *N. ephratae* (Brune, 2014). Stirring was fixed to 250 rpm. Aerobic conditions were ensured by the continuous injection of 0.5 VVM of filtered air (0.2µm, Sartorius). Bioreactor experiments were conducted in triplicate and compared to a control reactor without inoculation.

150 ml samples were taken at 0, 8, and 15 days of incubation. These samples were aliquoted according to the required analysis.

VI.4.4. Volatile solid quantification

Right after sampling, 120 ml of culture broth were divided in 4-aluminium dishes (beforehand dried and weigh) 48h at 55°C. The samples were dried at 55°C for 48h to determine volatile solid content (VS) corresponding to the total dry matter.

VI.4.5. Cell wall extraction and grinding

Extractive-free cell wall residues (CWR) were obtained by sonication-solvent extraction of 2g of VS samples (Mansfield et al. 2012). CWR was obtained after drying at room temperature overnight and for 24 h at 55°C. CWR was then milled in a ball mill MM 400 (Retsch).

VI.4.6. Polysaccharide analysis and Klason lignin determination

The chemical composition of LWS was determined on the original LWS samples and the digested samples collected from the bioreactors at the end of the incubation period. Their composition in cellulose and hemicellulose was determined using the sulfuric acid hydrolysis method described by Lazuka et al. (2015) (Supplementary Information). Sugar composition was determined on an Ultimate 3000 Dionex HPLC with refractive index detector (Thermo Scientific) equipped with a BioRad Aminex HPX 87H affinity column(Lazuka et al. 2015). The insoluble residue was washed with distilled water and dried at 105 °C overnight to determine Klason lignin content.

To estimate the absolute extent of removal of the different lignocellulose components by the NE consortium in the bioreactor experiments, we computed the total remaining mass of polymers (m_p) per liter as such [Equation 1]:

$$[Equation\ 1]\ m_p = w_p * m_{vs} * \frac{m_{CWR}}{m_t}$$

considering the mass percentage of said polymers (w_p), total mass of volatile solids (VS) remaining in the bioreactors (m_{VS}) corrected by weight loss induced by extraction ($\frac{m_{CWR}}{m_t}$).

VI.4.7. 2D HSQC NMR

CWR samples were swelled in DMSO-d_6/pyridine-d_5 (4:1, v/v) for the gel-state NMR analysis as described previously(Tarmadi et al. 2018). NMR spectra were acquired using the Avance III 800US system (Bruker Biospin, Billerica, MA, USA) equipped with a cryogenically cooled 5 mm TCI gradient probe (Bruker Biospin). 2D

HSQC NMR experiments were carried out using the standard Bruker implementation ('hsqcetgpsp.3') with the parameters described in literature(Kim & Ralph 2010; Mansfield et al. 2012). Data processing and analysis were performed using TopSpin 4.0 software (Bruker Biospin) as described previously.(Tobimatsu et al. 2013; P. Y. Lam et al. 2017) The central DMSO-d_6 solvent peaks (δ_C/δ_H: 39.5/2.49 ppm) were used as internal references. For volume integration analysis of the aromatic contour signals, C_2–H_2 correlations from G, G′ and F, C2–H2/C6–H6 correlations from S, S′ and P, and C2′–H2′/C6′–H6′ correlations from T were manually integrated and S, S′, P and T integrals were logically halved. For the polysaccharide anomeric signals, C1–H1 correlations from Gl, X, X′, X″, X‴, A and U were integrated. For the lignin side-chain signals, Cα–Hα correlations from I, II, III, III′ and IV″, Cγ–Hγ correlations from IV‴ and Cγ–Hγ correlations from IV and IV′ were integrated, and III, III′, IV, and IV‴ integrals were logically halved. The obtained integration values were normalized on a $½S_{2/6} + ½S'_{2/6} + G_2 + G'_2 + ½H'_{2/6} = 100$ basis for each spectrum.

VI.4.8. Quantitative py-GC-MS with ^{13}C lignin as internal standard

Analytical pyrolysis coupled to gas chromatography with high-resolution mass spectrometric detection (Exactive Orbitrap, Thermo Scientific, Waltham, MA, USA) was performed as previously described(van Erven et al. 2019). To each CWR sample (~80 µg), 10 µL of a ^{13}C wheat straw lignin internal standard (IS) solution (1 mg.mL^{-1} ethanol/chloroform 50:50 v/v) was added and dried prior to analysis. All samples were prepared and analyzed in triplicate. Lignin-derived pyrolysis products were monitored in full MS mode on the most abundant fragment per compound (both nonlabeled and uniformly ^{13}C labeled). Pyrograms were processed by TraceFinder 4.0 software. Lignin contents and relative abundances of lignin-derived pyrolysis products were calculated as described previously (van Erven et al. 2019).

VI.4.9. Analyses of the supernatant extracts

VI.4.9.1. Sample preparation

The supernatant (80mL) of the inoculated and control bioreactors were extracted with 20 mL ethyl acetate (EA) after pH adjustment to 3–4 (2M HCl aqueous solution). The aqueous layer was further extracted by 2x20 mL EA. The combined EA extracts were dried over $MgSO_4$ and dried under reduced pressure bellow 45 °C. The extracts were dissolved in THF and filtered (0.45 µm acrodisc GHP filter).

VI.4.9.2. HPSEC analysis

HPSEC analysis of the ethyl acetate extracts was performed using two different columns: 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, mixed-C pore type) and 600 × 7.5 mm PL-gel column (Polymer Laboratories, 5 µm, 100Å). THF (1 mL. min^{-1}) was used as eluent and detection was performed at 280 nm. All the chromatograms were normalized in time scale according to the peak of toluene, injected as standard. Only chromatograms obtained with the 100 Å are shown here since they were found more informative.

VI.4.9.3. LC-MS analysis

For LC-MS analysis of the ethyl acetate extracts, aliquots of the extracts were evaporated to dryness and dissolved in acetonitrile and filtered (0.45µm, GHP Acrodisc, Pall Gelman, Merck, Molsheim, France). Samples were injected into an UHPLC apparatus (Thermo Fisher Scientific) coupled to an electrospray ionization mass spectrometer (ESI-MS) and photodiode array (PDA) co-detection. UHPLC analysis was performed using a C18 column (2.7 µm, 50 mm x 2 mm I.D.mm; Highpurity, Thermo Fisher Scientific), a 12–95 vol.% aqueous acetonitrile, 1% HCOOH gradient (Millipore, ST Quentin-en-Yvelines, France) for 30 min and a 1 $mL.min^{-1}$ flow rate. Negative-ion ESI-MS spectra (120–2000 m/z) were acquired using a quadrupole–time-of-flight (Q-TOF) spectrometer (Impact II, Bruker, Leipzig, Germany) setting needle voltage at 4 kV and capillary temperature at 350°C.

VI.4.10. Enzymatic activities

As described Lazuka et al., (2018), 5 ml samples were used for enzyme activity measurement. The samples were centrifuged and xylanase and endoglucanase (CMCase) activities were measured using 1% w/v xylan beechwood (Sigma) and 1% w/v carboxymethyl cellulose (CMC) (Sigma) on both the supernatant and the sonicated pellet. One unit of CMCase or xylanase activity (UA, unit of activity) was defined as the amount of enzyme that produces 1 µmol of reducing sugars per minute.

Laccase, Lignin peroxidase (LiP) and Manganese peroxidase (MnP) activity measurement were conducted on 2,6-dimethoxy phenol (2,6-DMP) with the adjunction of the necessary cosubstrates, i.e Hydrogen peroxide for LiP, hydrogen peroxide and MnSO4 for MnP. The assays were conducted in the bioreactors conditions in a phosphate buffer at pH 7 and 30°C to have access to the enzymes activity in the conditions of the bioreactor. One unit of laccase, LiP or MnP activity (UA, unit of activity) was defined as the amount of enzyme that oxides 1 µmol of 2,6-dimethoxy phenol.

VI.4.11. DNA extraction

DNA was extracted from twelve samples (1.5 ml) taken from the duplicate bioreactors (6 each) throughout the incubation time. Samples were centrifuged (13,000 g, 5min, 4°C), the supernatant was removed and the pellet was snap frozen in liquid nitrogen and stored at -80°C until DNA extraction. Total DNA was extracted using an RNeasy PowerMicrobiome kit (Qiagen) following the manufacturer's instruction but omitting the final DNAse step. Cell lysis was performed with a FastPrep (MP Biomedicals) (2 x 30s at $4 m.s^{-1}$). DNA was purified using an AllPrep DNA/RNA Mini kit (Qiagen) following the manufacturer's instructions. DNA concentration was measured by Nanodrop 1000 spectrophotometer (ThermoScientific) measuring absorbance at 260 nm and 280 nm. DNA purity was assessed by electrophoresis on a 0.8% agarose gel. The purified DNA was also quantified by Qubit fluorimetric measurement (ThermoScientific).

VI.4.12. Shotgun metagenomics sequencing, de novo assembly and annotation

For shotgun metagenomics, DNA libraries for sixteen individual samples were prepared following Illumina NovaSeq instructions using True Seq DNA-PCR free kit v2 (Illumina, San Diego, CA, USA). Libraries were pooled in identical quantities and paired-end sequenced using one line of the Illumina NovaSeq platform to obtain 150 bp reads. Shotgun sequencing was performed at the GenoToul Genomics and Transcriptomics facility (GeT, Auzeville, France).

Shotgun sequencing resulted in 410 M paired reads in the dataset (range: 19-31 million per sample) (Supplementary data 1). All samples were submitted to quality control using FastQC (Andrews, 2010). As all samples displayed a minimal Phred quality score of 20, no remaining adapter, no overrepresented sequences were found, the k-mer profiles were normal and the desired length of 150bp they were kept for genome reconstruction.

For genome reconstruction reads form all samples were used for *de novo* co-assembly using MEGAHIT v1.1.3 with default parameters (D. Li et al. 2016). Co-assembly produced 708764 contigs (size ranging from 200 to 2.7 Mbp, N50=16175, L50 = 12143).

Assembly statistics were obtained with Assemblathon.pl. Paired-end reads were mapped onto the contigs using BWA-MEM v0.7.17 with default parameters (Heng Li 2013) and mapping statistics were compiled using the Samtools suite v1.8 (Heng Li et al. 2009).

Open reading frames (ORFs) were searched on contigs and structural genes were annotated from the assembly using Prodigal v2.6.3 run in a metagenomics mode (-p meta).(Hyatt et al. 2010)

VI.4.12.1. Metagenome assembled genome (MAG) reconstruction

For MAGs reconstruction, contigs were filtered using minimum contig length of 1500 bp and minimal abundance of one RPKM (reads per kilobase per million mapped reads) on at least two different samples. It resulted in 57,135 contigs that were used to

generate a contigs database using Anvi'o v5.4 (Eren et al. 2015). Contigs were binned using 3 different binning algorithm, namely CONCOCT v1.0.0 (Alneberg et al. 2014), MetaBAT 2 v2.12.1 (Kang et al. 2019) and MaxBin 2 v2.2.7 (Wu et al., 2016). The three independent binnings were then aggregated using DAS Tool (Sieber et al. 2018) to obtain a set of metagenome assembled genomes (MAGs). Each MAG was manually curated using Anvi'o v6.2 interactive based on differential coverage and nucleotide composition. SCG were used to assess the completeness and contamination of the resulting MAGs using Anvi'o v6.2.

We classified the resulting MAGs as quality (>90% completeness, <5% redundancy), medium quality (≥50% completeness, <10% redundancy) and low quality (other MAGs) in agreement to proposed classification(Bowers et al. 2017). Only MAGs of medium or high quality were kept for analysis, resulting in 174 MAGs.

VI.4.12.2. Taxonomic and phylogenomic analysis

Taxonomic affiliation of the MAGs was performed with CAT and BAT in Bin Annotation Tool v5.0.3 (BAT) mode (von Meijenfeldt et al. 2019) using –r=10 –f=0.3 as criteria.

A MAGs phylogenomic tree was constructed with Phylophlan3 using the Phylophlan database of proteins and parameters set to accurate/high diversity (Segata et al. 2013). The best consensus tree over 100 bootstraps was obtained from RAxML-ng v0.9.0 (Kozlov et al., 2019). The phylogenomic tree was drawn using the Interactive Tree Of Life (iTOL v5) (Letunic et Bork 2019).

VI.4.12.3. Functional annotation of MAGs: CAZymes and pfam

Functionnal annotation was separated into two parts.

CAZyme annotations were performed on the identified. A custom script which aims at accommodating the modularity of CAZymes was used to compare sequences to the full-length sequences stored in the CAZy database using BlastP (version 2.3.0+). If sequences are aligned with a sequence already in the CAZy database with 100% coverage, >50% amino acid sequence identity and E-value ≤ 10−6, they are assigned

to the same family or subfamily (or the same families if the found sequence contains more than one CAZy module). If not, the sequences have to undergo a similarity search that involves two steps. On the one hand, using BlastP, they are compared to a library of modules (instead of full sequences that can contains multiple modules). On the other hand, a HMMER3 search against a curated collection of hidden Markov models based on each of the CAZy module is performed. The script assigns a family when both BlastP E-value<10^{-4} and it aligns with >90% overlap.

Predicted proteins with a potential lignin degrading activity were identified using pfam modules. A set of 31 modules were selected as important for lignin degradation and aromatics metabolism (Moraes et al. 2018). Pfam were predicted using Hmmer v3.2.1on Pfam33.1 release. Only modules with E-values < 1 e10^{-5} were considered.

VI.4.12.4. MAGs temporal dynamics along LWS degradation

To estimate the abundance of the reconstructed MAGs at each sampling point, raw reads were queried against to a database constructed with the MAGs qualified as al least medium-quality using bwa-mem with default parameters. Unmapped reads and reads mapped to more than one location were removed with samtools view (-q =10). The relative abundance of each MAG for each sampling point was computed as the fraction of reads in each sample mapping to the respective MAG normalized on the size of that bin (fraction of reads per nucleotide in bin).

Abundance of the functional traits of MAGs along incubation time were explored and statistical differences between different traits was assessed using the R package vegan.

VI.4.13. Statistical analysis

Subsequent data analysis was performed on R v4.0.2. Principal component analysis (PCAs) and Partial Least Square Discriminant Analysis (PLSDAs) were done with the MixOmics v6.6.2(Rohart et al. 2017). Heatmap were generated through pheatmap and bootstrap was performed using pvclust (Suzuki & Shimodaira 2006). Networks analyses were performed using NetCoMi (Peschel et al. 2020). All analysis were performed using the Tidyverse v1.3.0.

Statistical differences between non-inoculated control and digested samples were calculated with t.test, $p<0.05$, n= number of samples

VI.5. RESULTS

VI.5.1. Lignocellulose degradation by NE15 consortium in aerobic bioreactors

In a previous work (see Chapter V) we have selected a microbial consortium derived from *N. ephratae* gut microbiome (NE15) for its ability to degrade LWS lignin.

Here, a detailed characterization of the lignocellulose degradation and more specifically of the lignin degradation capacity of this consortium was determined on triplicate bioreactors using LWS as sole carbon source. The results were compared to those obtained in a uninoculated control bioreactor.

The analysis of the uninoculated control displayed an average of 11% of cellulose removal in abiotic conditions after 8 days. No significant hemicellulose nor lignin removal was found in this control bioreactor (Figure 1). A detailed characterization of polymer removal is provided in Supplementary data 2.

The inoculated bioreactors demonstrated the ability of NE15 to degrade lignocellulose. Indeed, after 8 days, 22% of the cellulose, 19% of hemicellulose and 16% of lignin were removed (Figure 1), which is significantly different from the control. After 15 days of incubation, an average 29% of the cellulose, 31% of hemicellulose and 20% of lignin were removed from the bioreactors. These removal values were statistically different from the control and corresponded to the removal rates found in the experimental flasks during the enrichment process (Chapter V).

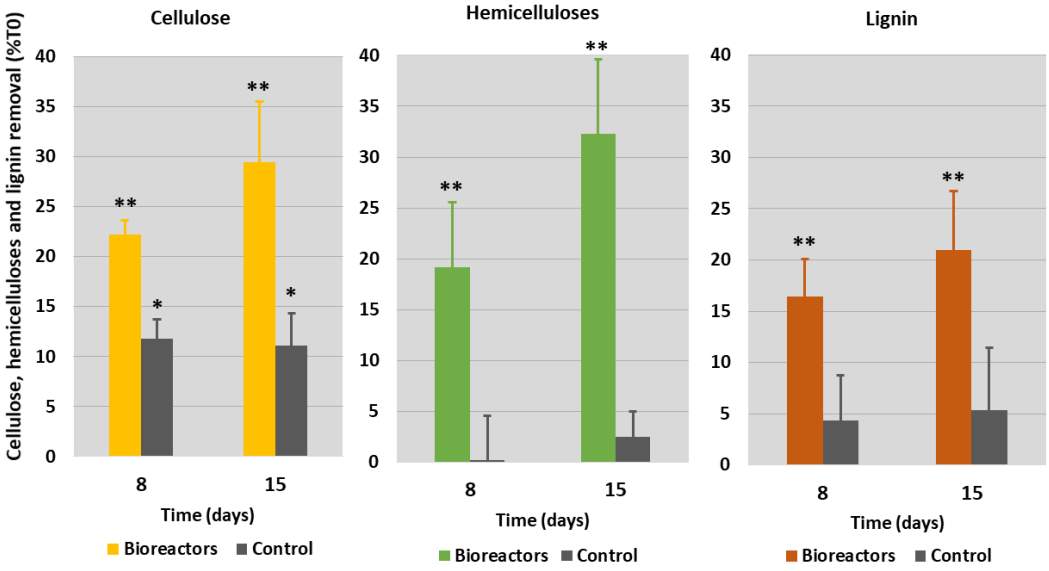


Figure 1: Average cellulose, hemicellulose and lignin removal in the bioreactors compared to the non-inoculated control. Significant differences between inoculated and non-inoculated samples are indicated by ** ($p<0.05$). Significant differences between non-inoculated samples and T0 are indicated by * ($p<0.05$).

The chemical composition of LWS was also analyzed in detailed before and after incubation (Table 1) and compared to the control. Overall, no statistical difference was observed between undigested LWS and the control after 15 days of degradation.

Overall, only minor compositional changes were observed. After 15 days, the samples had an average of 14.9% less cellulose, 10.1% less hemicellulose and 13.3% more lignin. The composition was not significantly different after 8 days of incubation. When considering ^{13}C-IS py-GC-MS results, no significant change could be observed. These results are in line with expectations, as all three polymers were removed to a fairly similar extent, although a slight increase in lignin in the overall composition would be expected.

Table 1: Compositional analysis of LWS samples before and during digestion by the selected consortium from *N. ephratae*. Values are expressed as means ± SD g per 100 g cell wall residues (CWR). Values in brackets indicate the percentage change compared to the original undigested LWS. Statistical differences are marked with * ($p < 0.05$, $n = 3$, Student's *t*-test).

	LWS	Control	NE-inoculated bioreactor		
		15d	0d	8d	15d
Cellulose (g /100 g CWR)					
Total	37.7 ± 1.0	36.3 ± 2.9 (-3.7%)	37.2 ± 1.3 (-1.3%)	35.5 ± 0.4 (-5.8%)	32.8 ± 0.8* (-14.9%)
Hemicelluloses (g /100 g CWR)					
Total	33.8 ± 1.1	35.1 ± 1.3 (+3.8%)	33.0 ± 0.9 (+0.6%)	33.1 ± 0.5 (-2.1%)	30.4 ± 0.7 (-10.1%)
Lignin (g /100 g CWR)					
by Klason	28.4 ± 0.7	28.6 ± 2.4 (+0.1%)	29.8 ± 2.1 (+4%)	31.8 ± 0.8 (+11.1%)	32.3 ± 0.8* (+13.3%)
by $^{13C\text{-}IS}$ py-GC-MS	27.2 ±1.6	26.4	26.9 ± 1.1 (-1.1%)	25.4 ± 0.2 (-6.6%)	26.1 ± 0.4 (-3.9%)

VI.5.1.1. Structural characterization of digested LWS residues by 2D HSQC NMR and ^{13}C-IS py-GC-MS

To investigate the lignin-unlocking mechanisms of the selected termite gut microbiome, lignocellulose structures of digested LWS samples were further characterized and compared to the non-inoculated control.

2D HSQC NMR was employed to obtain detailed information on the chemical structures of the wheat straw cell walls digested by the termite gut microbiome. To picture the modifications that occurred in the lignin, HSQC spectra of the CWR samples were collected using the dissolution/swelling method in a DMSO-d_6/pyridine-d_5 solvent system. The NMR spectrum of LWS displayed lignin and polysaccharides signals that can be found in grass cell walls. In particular, the aromatic sub-region (Figure 2a, Supporting Table 1) showed resolved signals from typical S, G and H lignin aromatic units (S, S′, G, G′ and H) and grass-specific cell wall components such as tricin (T) and *p*-coumarates (P)(Karlen et al. 2018) which have been proven to be bound to lignin(Karlen et al. 2018; W. Lan et al. 2015; P. Y. Lam et al. 2019). Ferulates (F) which are known to be bound to hemicellulose were also found in LWS.

On the one hand, the polysaccharide anomeric sub-region (Figure 2b) displayed signals from glucans (Gl), unacetylated (X) and acetylated (X′, X″, and X‴) xylans,

arabinan (A) and glucuronan (U). These signals are associated with cellulose and hemicellulose. On the other hand, the oxygenated aliphatic and aldehyde sub-regions displayed signals from various inter-monomeric linkage types, such as β-O-4 (I), β-5 (II) and β-β (III and III′) linkages, as well as polymer end-units, such as cinnamyl alcohol (IV), cinnamaldehyde (IV′), benzaldehyde (IV″) and HPV/HPS (IV‴) end-units, in the lignin polymer backbones (Figure 3a and b, and Supporting Table 1).

The compositional changes in the LWS cell walls induced by the consortium NE15 were deduced from the comparison between the control and digested cell wall of the volume integration of the HSQC contour signals. Lignin aromatic and polysaccharide anomeric signals were integrated and normalized based on the sum of the total of major lignin aromatic signals ($S_{2/6}$ + $S'_{2/6}$ + G_2 + G'_2 + $H_{2/6}$) (Figure 2c). Glucan (Gl) and xylan (X, X′, X″, and X‴) dominated the signals detected in all the cell wall spectra (Figure 2a). These signals were used to deduce changes in the composition of LWS in cellulose and arabinoxylan, respectively. However, it must be noted that glucan signals from crystalline cellulose are underestimated by the current gel-state NMR method due to its incomplete gelation in the solvent.(Kim & Ralph 2010; Mansfield et al. 2012)

Over time, a decrease of glucan signals was observed as well as a slight decrease from non-acetylated and mono-acetylated xylan (X, X′ and X″) and arabinan (A) signals (Figure 2c). This result further corroborates the preferential degradation of cellulose and hemicellulose over lignin by NE15 consortium as determined earlier by the wet-chemical analyses (Figure 1 and Table 1). Unlike glucan (Gl) and xylan (X, X′ and X″) signals, arabinans (A) and uronic acid (U) signals appeared not to diminish. Compared to the control reactor, ferulate (F), p-coumarate (P) and tricin signals were slightly weaker at the beginning of the bioreactor experiments (T0) but remained unchanged thereafter.

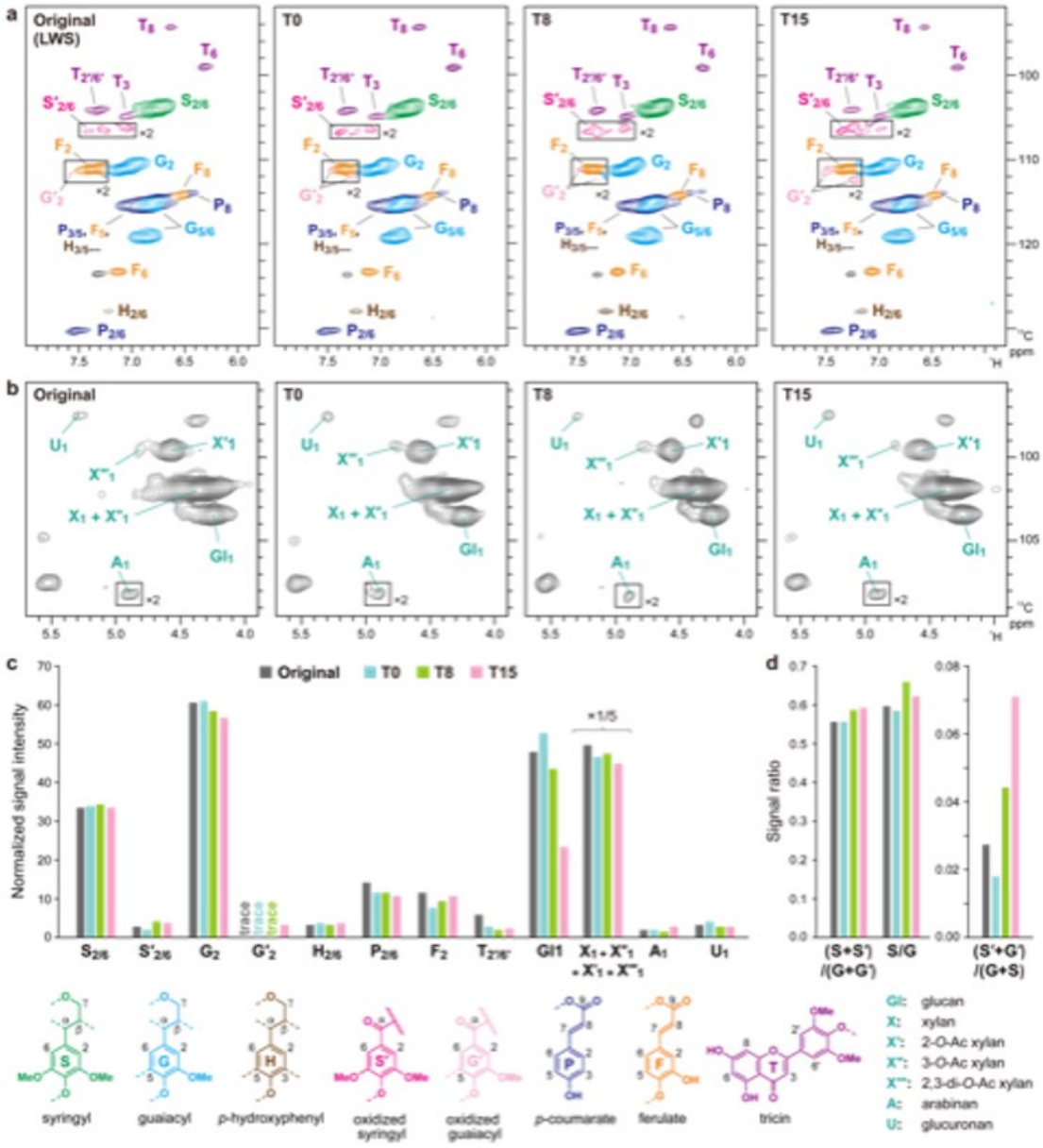


Figure 2: Partial short-range 1H-13C correlation (HSQC) NMR spectra of LWS cell walls after 15 days of incubation on the non-inoculated control and after digestion by NE15 for approximately 2 hours (T0), 8 days (T8) and 15 days (T15). Aromatic (a) and polysaccharide anomeric (b) sub-regions are shown. Signal assignments are listed in Supporting Table S1. Boxes labelled ×2 represent regions with the scale vertically enlarged by 2-fold. In (c) and (d) normalized signal intensity values of major aromatic and anomeric signals expressed on a ½S2/6 + ½S′2/6 + G2 + G′2 + ½H′2/6 = 100 basis (c), and ratios of S (½S2/6 and ½S′2/6) and G (G2 and G′2) aromatic signals (d) are

shown. Data labelled ×1/5 indicate that the reported values are divided by a factor 5 for visualization purposes.

It was found, in agreement with py-GC-MS data (described below), that the ratio of total lignin S and G aromatic units based on the NMR signal ratio (S + S′)/(G + G′) as well as the ratio of non-oxidized S and G units based on the S/G signal ratio increased with time in the bioreactors (Figure 2c). This increase was due to both a decrease in G units and an increase in Cα-oxidized S (S′) signals. The Cα-oxidized G signals (G′) also apparently increased with time, as only traces were found before 15 days (Figure 2c). As a result, the signal ratio (S′ + G′)/(S + G) increased by 4-fold between T0 and T15 in the spectrum of LWS cell wall samples.

Volume integration analysis for the signals from the inter-monomeric linkages and end-units in lignin polymer appearing in the oxygenated aliphatic and aldehyde sub-regions of the HSQC spectra brought interesting results (Figure 3). The intensities of I, II, III and III′ signals, normalized based on S + S′ + G + G′ + H, generally decreased between the original LWS cell wall spectrum and digested samples suggesting a slight degradation of the polymer (Figure 3d). Overall, the main inter-monomeric linkage i.e. β-O-4 (I) remained largely the same. However, it was found that β-β (III and III′) linkages decreased with time, suggesting a possible preference of the consortium to break down this type of C-C linkages. The decrease in β-β (III and III′) linkages with time was compensated by a clear increase in alcohol signals from cinnamyl alchohol (IV) and the appearance of aldehyde signals from cinnamaldehyde (IV′) and benzaldehyde (IV″) end-units and hydroxypropiovanilone (HPV) and hydroxypropiosyringone (HPS) end-units (IV‴; δC/δH, Cβ-Hβ at 41.5/3.14 and Cγ-Hγ at 57.0/3.85)[36] after 8days of digestion by NE15 (Figure 3c). These data corroborated the increase in Cα-oxidized lignin aromatic signals (S′ and G′) described previously (Figure 2c) and provides pieces of evidence of the action of NE15 consortium on polymeric lignin.

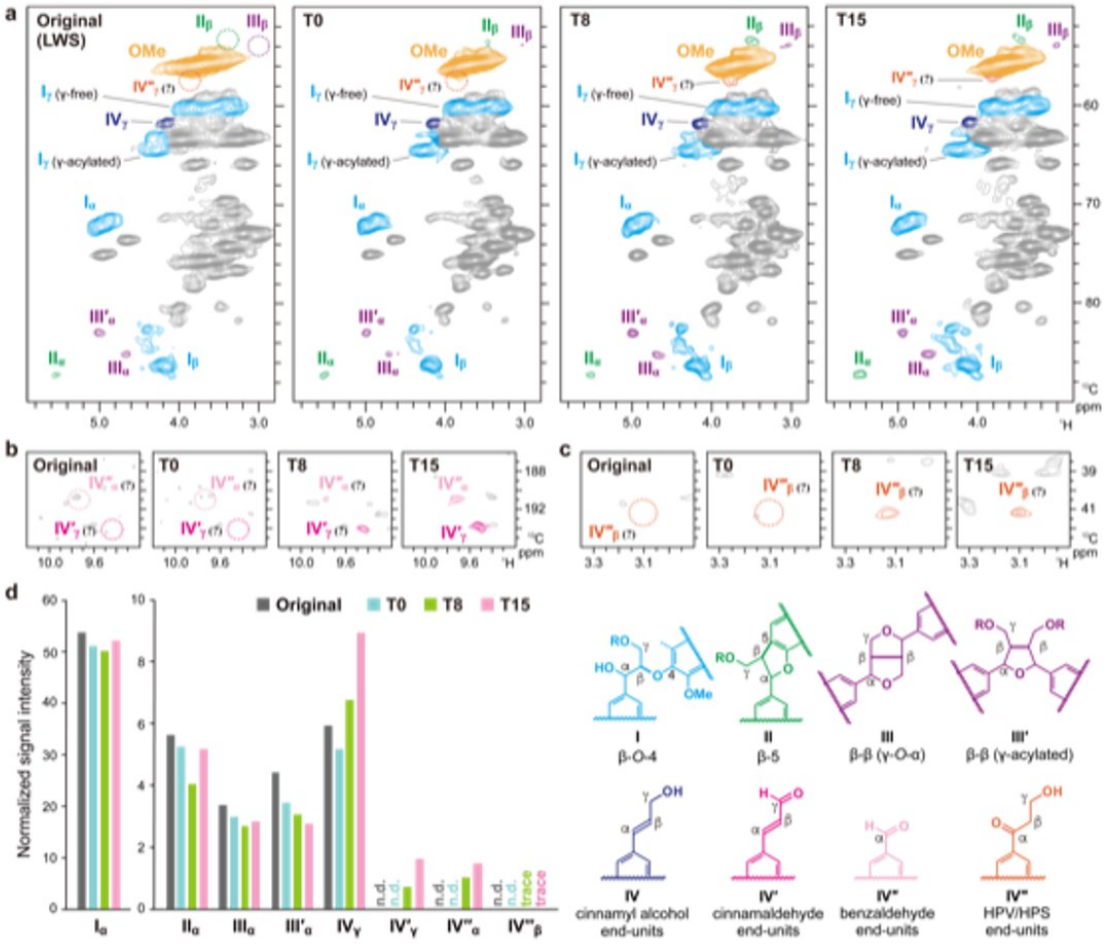


Figure 3: Partial short-range 1H-13C correlation (HSQC) NMR spectra of LWS cell walls in the non-inoculated control (Original) after 15 days of incubation and in the NE15-inoculated bioreactor along the incubation time: T0 (about 2h), T8 (8 days) and T15 (15 days). Oxygenated aliphatic (a and c) and aldehyde (b) sub-regions are shown. Signal assignments are listed in the Supporting Table S1. In (d), normalized signal intensity values expressed on a ½S2/6 + ½S′2/6 + G2 + G′2 + ½H′2/6 = 100 basis (see Figure 2), are shown. n.d., not detected.

^{13}C-IS py-GC-MS was used to corroborate and complement NMR analyses of lignin structure changes. The different lignin-derived pyrolysis products of the ~~original and~~ digested LWS were classified according to their structural moieties and summarized in Table 2. ^{13}C-IS py-GC-MS analysis of LWS along NE15 digestion displayed a different subunit composition than the control (after 15 incubation days), as all digested samples displayed significantly lower H and G subunit proportions and higher G unit. This entails that an increase of the S/G ration was observed. However, most

differences were already observed at T0, taken approximately 2 hours after inoculation and the 15 days of degradation do not appear to have impacted S/G ratio or only marginally.

Similarly, while differences between the control and digested samples could be found when looking at the structure in a more detailed way, no difference could be observed after the first couple hours of digestion.

Vinyl products were found to decrease with inoculation (~8-11%) (Table 2). This was due to a decrease in 4-vinylphenol (up to ca 13%) and 4-vinylguaiacol (up to ca 15%). As these products are mainly derived from p-coumarate and ferulate, this indicates a removal of these molecules, which is in agreement with the NMR results (Figure 2c).

Cγ-oxidized pyrolysis products increased slightly, although this increase was only statistically significant for T0. This is also related to a slight but non-significant increase in PhCγ products. These results, as previously proposed can be correlated to the appearance of HPV/HPS end-units in NMR (Dumond et al. 2021).

Finally, a 13-23% increase on unsubstituted moieties was displayed. Overall, lignin characterization by 2D HSQC NMR and ^{13}C-IS py-GC-MS demonstrated that NE15 degraded cellulose and hemicellulose preferentially over lignin. Although minor, differences were observed between control and inoculated bioreactors providing evidence of lignin modification by NE15. Most notable, an ability to degrade β-β linkage rather than β-O-4 was observed as well as a conversion to aldehydes end-units.

Table 2: ^{13}C-IS py-GC-MS relative abundance of residual lignin compounds of LWS after the digestion by the NE15 microbial consortium and compared to the non-inoculated control. Values are expressed as means ± SD. Statistical differences are marked with * ($p < 0.05$, $n = 3$, Student's *t*-test).

	Control at 15 days	Digested LWS		
		Incubation duration (days)		
		0	8	15
Lignin units (abundance in mol%)				
H	**9.1 ± 0.3**	8.5 ± 0.2*	8.5 ± 0.1*	8.2 ± 0.1*
G	**60.9 ± 0.4**	57.6 ± 0.5*	57.3 ± 0.6*	56.7 ± 1.1*
S	**30.0 ± 0.6**	33.9 ± 0.7*	34.2 ± 0.5*	35.1 ± 1.2*
S/G	**0.49 ± 0.01**	0.59 ± 0.02*	0.60 ± 0.01*	0.62 ± 0.03*
Structural moieties (abundance in mol%)				
Unsubstituted	**5.5 ± 0.09**	6.3 ± 0.13*	7.0 ± 0.34*	6.6 ± 0.66
Methyl	**3.7 ± 0.19**	3.7 ± 0.22	3.9 ± 0.17	3.7 ± 0.07
Vinyl	**36.8 ± 0.35**	33.2 ± 0.84*	33.9 ± 0.40*	32.8 ± 0.68*
4-VP	**7.1 ± 0.30**	6.5 ± 0.22	6.4 ± 0.06*	6.2 ± 0.11*
4-VG	**25.6 ± 0.28**	22.1 ± 0.69*	22.6 ± 0.19 *	21.7 ± 0.69*
C_α-ox	**3.9 ± 0.04**	3.9 ± 0.14	3.8 ± 0.27	3.7 ± 0.13
C_β-ox	**2.1 ± 0.03**	2.1 ± 0.09	2.1 ± 0.08	2.1 ± 0.05
C_γ-ox	**43.8 ± 0.65**	46.7 ± 0.50*	45.0 ± 0.58	46.9 ± 1.08
Miscellaneous	**4.2 ± 0.11**	4.1 ± 0.18	4.4 ± 0.01	4.3 ± 0.18
PhC3	**48.9 ± 0.62**	51.7 ± 0.48	50.3 ± 0.57	52.1 ± 1.10

4-VP = 4-Vinylphenol
4-VG = 4-Vinylguaiacol
PhC3 = Phenols with intact α, β, and γ carbon side chain, excluding diketones

VI.5.1.2. Characterization of the liquid fraction of the bioreactors by HPSEC, LC-MS and NMR

Lignin removal from the solid fraction could result from the solubilization of lignin residues.(Dumond et al. 2021) In order to verify the aftermath of lignin removal in the bioreactor experiments we analyzed the liquid fraction after ethyl acetate extraction. HPSEC, LC-MS and NMR (Figure 4) analysis of samples taken at the end of incubation (15 days) from the inoculated and non-inoculated bioreactors showed, firstly, that all monomers (putatively vanillin and ferulic acid) were degraded over this period, regardless of bioreactor inoculation (Supplementary data 3). LC-MS analysis of theses samples allowed detection only of phenolic monomers. No other compounds were detected in any of the samples (chromatograms not shown). The monomers identified were mainly Pc and Fe. Pc, Fe, all the other extractives totally disappeared after 15 incubation days even without any inoculation and no monomer could be detected anymore with inoculation. Such degradation resulted mainly from abiotic processes. However, additional analysis of samples taken at day 8, showed that the degradation kinetics of these monomers was faster with NE15-inoculation as no monomer was found in samples from the NE15-inoculated reactors while these compounds were still detected in the control reactor.

While most monomers were degraded after 15 days, oligomers were still found in the control reactor (Figure 4a). Indeed, other molecules with a retention time between 14 and 16 minutes, which corresponds to oligomers with a degree of polymerization of 2 to 6, were found (Supplementary data 4). These oligomers appeared to be mostly degraded in the inoculated bioreactors.

In order to better characterize this fraction, the average molecular weight of the compounds present in ethyl acetate extracts was quantified (Figure 4b). Overall, the molecular weight (both Mw and Mn) was found to be decreased by a factor 2 in NE15-inoculated bioreactors, suggesting that small oligomers and in particular, soluble lignin oligomers, may have been removed or depolymerized.

In order to identify these oligomers, HSQC NMR of the ethyl acetate extracts was performed. This analysis clearly detected low Mw lignin fragments in the control samples but neither lignin nor sugar fragments could be detected in NE15-inoculated

bioreactors after 15 days of incubation (Figure 4c). Only a tiny amount of OMe signals was found and most of this signal consisted in aliphatic signals. This result suggests that most if not all soluble lignin oligomers and polysaccharides were removed by the NE15 consortium and only fatty acids and other EPS remained.

Overall, we found that the NE15 microbial consortium is able to remove up to 20% polymeric lignin from LWS, and most of the oligomeric lignin. Dedicated experiments performed using both monomers as substrate further confirmed this result (see Annex).

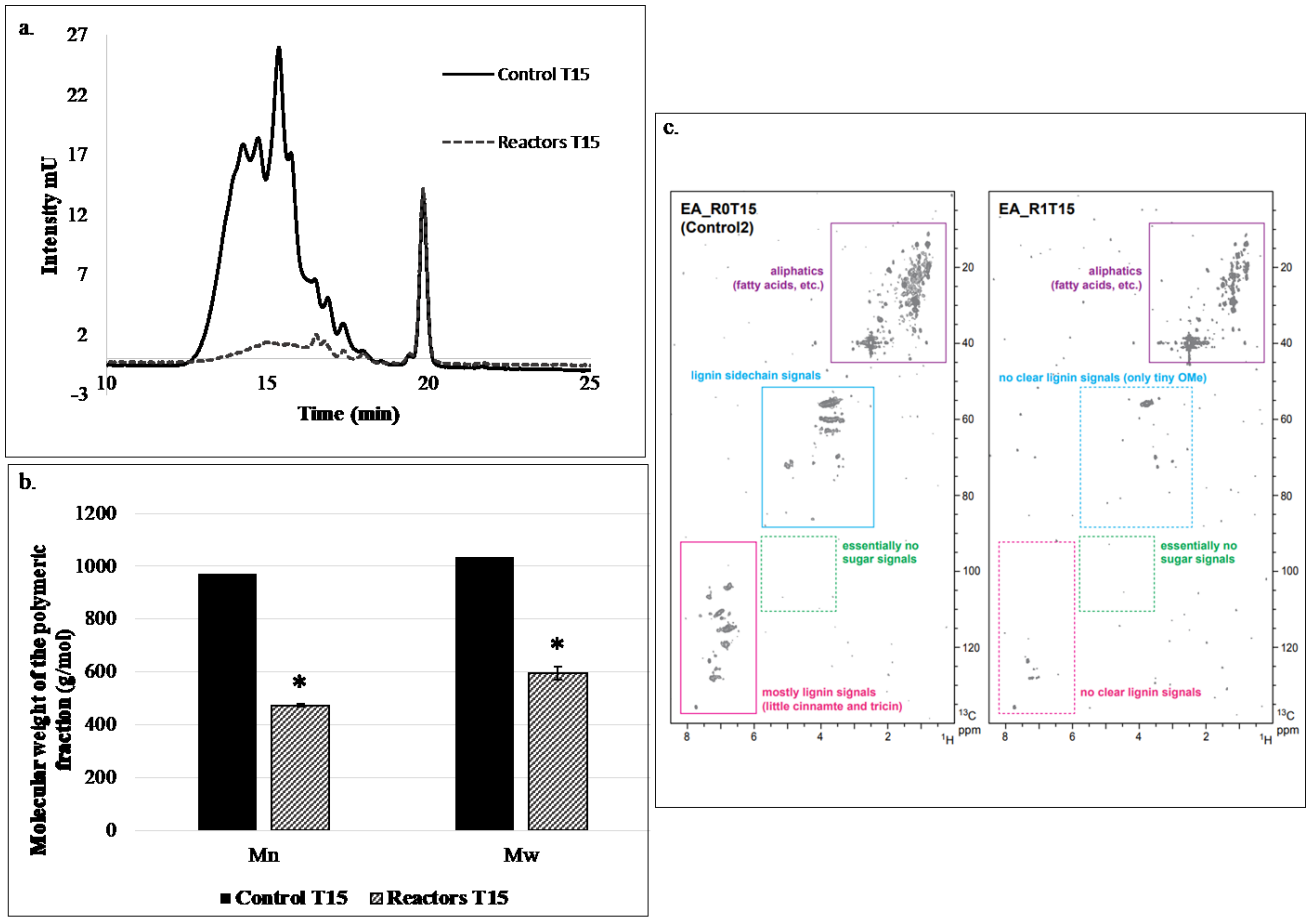


Figure 4: a. Spectrogram from HPSEC analysis of ethyl acetate extracts from supernatant after 15 days of incubation in the non-inoculated (Control T15) and NE15-inoculated bioreactors (Reactors T15). **b.** Mass weight (Mn and Mw) of ethyl acetate extracts. **c.** Partial short-range 1H-13C correlation (HSQC) NMR spectra of ethyl acetate extracts (control on the left, NE15-inoculated bioreactors on the right).

VI.5.2. Characterization of the lignocellulolytic potential of NE15 consortium

VI.5.2.1. Dynamics of enzyme activities

In order to better understand the enzymatic functions of NE15 consortium, key enzymatic activities linked to lignocellulose degradation were monitored over the 15 days of incubation in the bioreactor experiments. Enzyme activities linked to cellulose and hemicellulose degradation, both CMCase (endoglucanase activity) and xylanase activities (Figure 5 a) displayed a similar profile. Their activity increased over the first five days of incubation and then reached a plateau. Regarding the laccase activity, it increased over 8 days before reaching a plateau. Although low, the quantification of this laccase activity suggests some lignin degrading activity of the NE15 consortium. No LiP or MnP activity was found in the bioreactors. This indicates that either there are no peroxidases in this consortium or they are not active under the conditions of the bioreactors conditions.

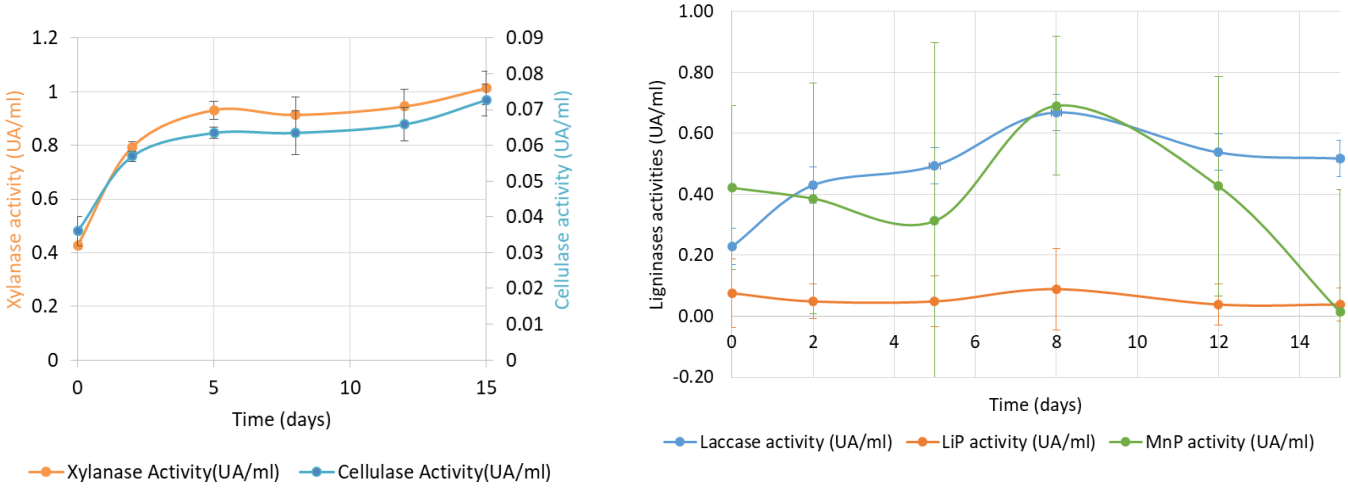


Figure 5: Lignocellulolytic enzyme activities of the NE15 consortium. **a.** Xylanase and cellulase (CMCase) activity (UA/ml reactor), **b.** Laccase, lignin peroxidase (LiP) and Manganese peroxidase (MnP) activities of the consortium (UA/ml reactor).

VI.5.2.2. Shotgun metagenomics analysis

Metagenomics allowed us to provide more detailed information on the functional potential related to polysaccharides and lignin degradation of the NE15 consortium.

Co-assembly of metagenomic shotgun sequencing data from LWS degradation experiments by NE15, followed by contigs filtering, resulted in a final dataset of 57,135 contigs with at least 1,500 bp length. This assembly had a N50 of 70,436 nt and a L50 of 3,424 nt. The length of the largest contig was 2,787,544 bp with 2,154 contigs longer than 100,000 bp. Mapping the reads to the assembled contigs resulted in 89% incorporation of read into filtered contigs.

ORF search in the contigs resulted in prediction of 1,067,026 proteins. As we were interested in lignocellulolytic related proteins, we searched for CAZyme and aromatic degradation-related proteins in the NE15 metagenome.

CAZyme annotation of the predicted proteins of the NE15 metagenome showed that this consortium encoded a wide diversity of CAZymes (Table 3) that were notably involved in cellulose and hemicellulose degradation. Shotgun sequencing revealed 24,931 CAZyme-related sequences mainly classified in the GT familly (39 %) followed by GH (37%) and CBM (10%). 1069 AAs were found, representing 4% of the total number of CAZymes. AA are known to be involved in lignin degradation.(Levasseur et al. 2013) The percentage of AA found in the NE15 consortium is similar to that found on lignocellulosic biomass (2-4%)(Montella et al. 2017) or in an apple pomace adapted compost (4.8%).(M. Zhou et al. 2017) Surprisingly, NE15 provided 2 fold less CAZymes/Mbp than a consortium derived from *N. ephratae* selected under anaerobic conditions using raw wheat straw as carbon source and 3 times less GHs/Mbp (see Chapter 4), making it less advantageous for cellulose/hemicellulose bioconversion. In contrast, NE15 provided more than 6 times more AA/Mbp, which may imply a higher potential for lignin deconstruction.

Table 3: Number of identified CAZy modules in the metagenome

CAZy modules	Number of Cazymes	Number of CAZymes/Mbp
AA	1069	0.95
CBM	2463	2.2
CE	1797	1.6
GH	9327	8.3
GT	9688	8.6
PL	587	0.5

A deeper analysis of the functional profile of the NE15 metagenome demonstrated the presence of crucial enzymes for cellulose and hemicellulose degradation. 182 GHs were identified as cellulases (mainly GH5 and GH9) and 285 hemicellulases (mainly GH10 and GH26) (Table 4). A large number of AA were also discovered which were mainly oxidases that could participate in lignin degradation by producing hydrogen peroxide. The latter can be used as a substrate for peroxidase or oxidize lignin through Fenton mechanism.(Janusz et al. 2017) A number of member of the AA1 family, whose functions include laccases and ferroxidases and 97 putative LPMOs were also found. Like oxidases, LPMOs may act as auxiliary enzymes of lignin-degrading peroxidases(F. Li et al. 2019).

In addition to CAZymes, the searched for potential lignin degrading protein by foraging for Pfam modules related to lignin degradation, allowed to identify 520 laccase modules (Table 4). Ninety-one modules for DyP peroxidases were also found but no module for LiP or MnP were discovered. This is in agreement with enzymatic assays as we were unable to detect such activities in the bioreactor experiments. This is also in agreement with recent views about LiP and MnP enzymes being exclusive to fungi(Davis et al. 2013).

Finally, 1,520 modules for beta-etherases were found which are known for their lignin breakdown potential(E. Masai et al. 1989; Voß et al. 2020).

Overall, the functional profile of NE15 metagenome allowed us to identify enzymes related to cellulose and hemicellulose degradation but also to lignin degradation. The high number of lignin-degradation related proteins suggest their involvement in the modification of polymeric lignin and in the degradation of soluble lignin oligomers in the bioreactor experiments.

Table 4: Number of identified CAZy modules in the NE15 metagenome

Main activity	**CAZyme familly**	**Number of CAZymes**
Cellulases	GH 5, GH9, GH44, GH48	182
Hemicellulases	GH8, GH10, GH11, GH26, GH93, GH141	285
Auxiliary activity		1039
Laccases/Ferroxidases	AA1	121
Oxidases (Incl. AAO, VAO, glucose oxidase)	AA3, AA4, AA5, AA7	821
LPMOs	AA10, AA12	97
PFAM modules		**Number of modules**
DyP peroxidases		91
Laccases		520
Beta-etherases		1502
MnP peroxidases		0
LiP peroxidases		0

VI.5.2.3. Reconstruction of metagenomic assembled genomes (MAGs) in NE15 metagenome

In order to identify the key members of the NE15 consortium and to better understand their functional complementarities and potential interactions that lead to the removal of polysaccharides and lignin in the LWS bioreactor experiments, we reconstructed the genomes of the main constituent bacteria.

After manual curation of the MAGs generated by the binning process, we managed to reconstruct 174 MAGs that meet the minimum quality requirements of 50% completion and less than 10% redundancy. Of these MAGs, 112 were considered of high-quality MAGs, having >90% completion and <5% redundancy, and the other 62 were qualified as medium quality. A full summary of the MAGs is provided in the Supplementary file 3.

Taxonomy of the MAGs was obtained using CAT and BAT tools on the contigs and the MAGs, respectively. The phylogenomic tree of the MAGs and their mean abundance in the metagenome was drawn using phyloseq (Figure 6).

The generated phylogenomic tree was found to have good bootstrap values as most of the branches were placed with a value of 75/100 or higher. Only a few branches could not be placed with high bootstrap values such as those of Bacteroidetes and Actinobacteria/Firmicutes. Nevertheless, most MAGs were placed on the tree according to their taxonomy with good confidence and no MAG was misplaced.

In term of number of MAGs, the NE15 consortium was mainly constituted by Proteobacteria with 91 MAGs (52%) followed by Bacteroidetes and Actinobacteria, with 39 (22%) and 20 MAGs (11%), respectively. Other phyla represented included Verrucomicrobia (6 MAGs), Firmicutes (5 MAGs), Candidatus Saccharibacteria and Planctomycetes (2 MAGs each), Acidobacteria, Chlorobi, Gemmatimonadetes and Candidatus Melainibacteria (1 MAG each).

At the class level, Proteobacteria were dominated by Alphaproteobacteria flowed by Gammaproteobacteria and Betaproteobacteria. Bacteroidetes were mainly Chitinophagiia and Sphingobacteriia followed by Flavobacteriia and Cytophagia.

Firmicutes and Actinobacteria MAGs were from a unique class, respectively Bacilli and Actinobacteriia.

Unlike when enriched in anaerobic conditions(Lazuka et al. 2018), no bacteria was predominant in the consortium with a rather even distribution of species. 11 MAGs represented 50% of the total abundance (Supplementary data 5). This includes two Bacteroidetes MAGs affiliated to *Moheibacter* (Refined_137) and *Chitinophagaceae Uncl.* (Refined_154), two Actinobacteria (Refined_86 & Refined_87) both affiliated to *Micrococceae Uncl.* and seven Proteobacteria. These 7 MAGs provide a mix between Alphaproteobacteria affiliated to *Alterythrobacter, Brucellaceae Uncl.* and *Devosia* (Refined_38, Refined_7 and Refined_4), Betaproteobacteria affiliated to *Variovorax* (Refined_70) and Gammaproteobacteria affiliated to *Enterobacteriaceae Uncl.,* Citrobacter and *Acinetobacter* (Refined_60, Refined_56, Refined_151). Overall, the even species distribution of NE15 consortium might indicates that a cooperation between its may be necessary to achieve the degradation of the different substrate fractions; each of the community members providing useful enzymes.

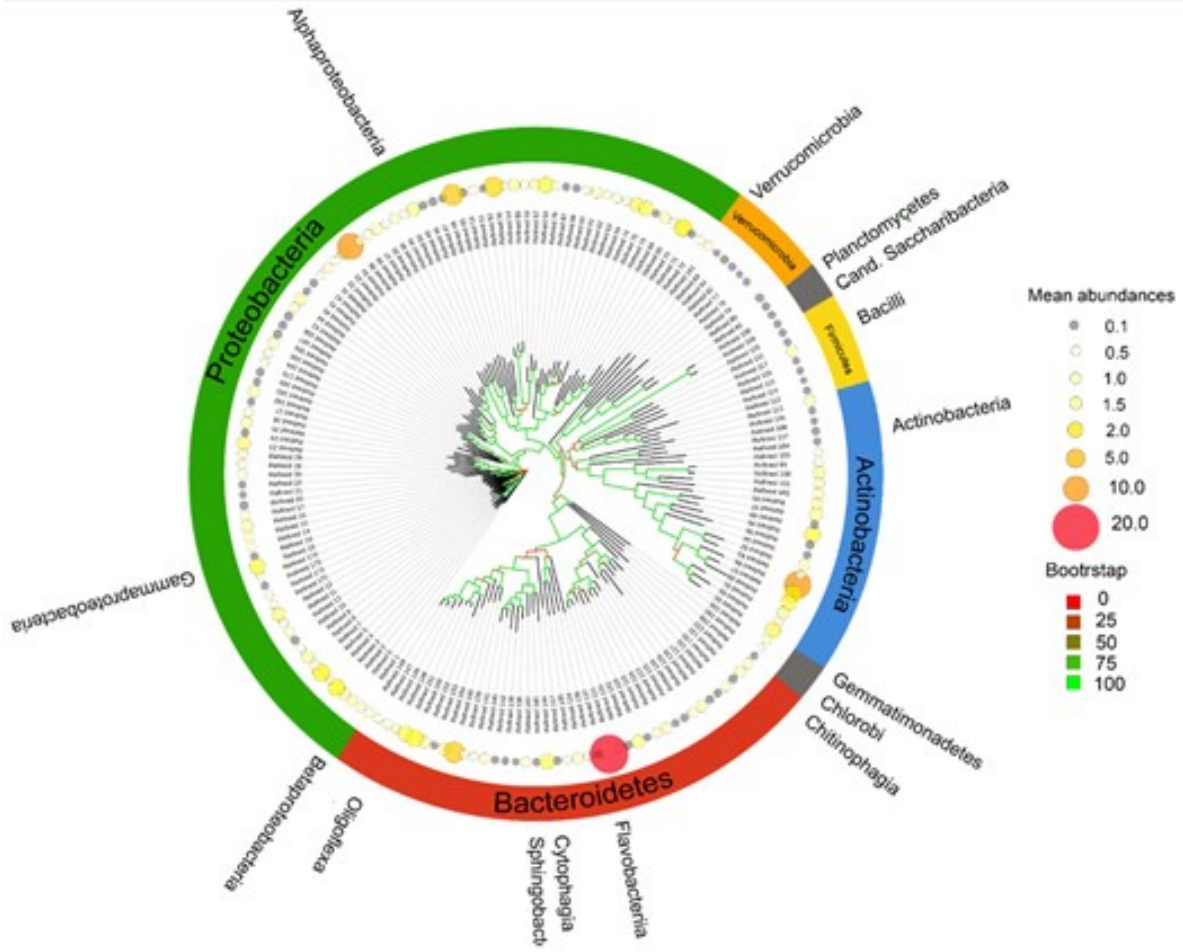


Figure 6: Phylogenomic tree of the selected MAGs with their average abundance in the metagenome. Taxonomic affiliation is provided as follow: Phylum in the colored circle, class as an outside crown. Mean abundances represents the average relative abundance in the metagenome of each MAG. Bootstrap was performed on 100 iterations using RaXML-ng.

VI.5.3. Lignocellulolytic potential of the most abundant MAGs in the termite-derived consortium

In order to shed light in how the NE15 consortium metabolize lignocellulose and to decipher the potential role of each MAG, we further characterized the 74 MAGs representing 90% of the total abundance (Supplementary 5). For each MAG, we searched for CAZymes and Pfam modules related to lignin degradation and to the

metabolism of aromatic compounds (Supplementary file 3). These modules have been classified according to their general function. GHs were regrouped as per their activity (see Chapter 4). The Pfam modules for aromatics metabolism were regrouped in accordance with the literature(Moraes et al. 2018). A normalized Z-score heatmap was performed to classify the 74 MAGs into groups using a cluster cutoff level of height =10, resulting in 4 distinct MAG clusters. The certainty of this classification was measured by running bootstrap (2,000 iterations) with pvclust (Suzuki & Shimodaira 2006) and the unbiaised p-values calculated for each branch are provied in the Supplementary data 6. On this basis, the uncertainty of cluster 1 and 3 (shown in green and yellow, respectively, in Figure 7) appeared to be high, but we chose to keep the separation at the same level rather than making the clusters purely homogeneous.

Cluster 1 (Figure 7, in green) grouped MAGs with a high content of genes involved in aromatics metabolism with a high number of Pfam modules implicated in metacleavage, DyP peroxidases, beta-aryl ether degradation and laccases (Supplementary data 7), such as the MAGs Refined_86 and 87, both affiliated to Micrococcaceae Uncl., (Actinobacteria) which are potentially interesting lignin degrader. The abundance of Micrococcaceae have already been reported to increase in lignin degrading microbial consortia(W. Zhang et al. 2021). Cluster 1 was also constituted by several Proteobacteria-related MAGs coding for laccases, affiliated to *Pseudixanthomonas*, *Rhizobiales*, *Pseudomonas*, *Enterobacteriaceae* and *Acinetobacter*. One of the most promising lignin degrader appears to be the MAG Refined_56 which was highly abundant in NE15 metagenome. *Acinetobacter* genus is known to be a powerhouse able to degrade various aromatics including lignin.(Ho et al. 2020; Vasudevan & Mahadevan 1991) In this cluster, two other MAGs affiliated to *Moheibacter* (Bacteroidetes) were fond. Among them, Refined_137 was the most abundant MAG in the bioreactor. It is interesting to notice that Bacteroidetes of this cluster coded for genes related to the aromatics meta cleavage pathway and laccases but also to hemicellulases.

Cluster 2 (blue in Figure 7), included MAGs with a specialization towards aromatics metabolism, coding particularly for beta-aryl ether degradation, the beta-ketoadipate pathway, ferulic acid and biphenyl compound degradation, beta-etherases and laccases. This cluster was exclusively composed by Actinobacteria and Proteobacteria.

Refined_38, affiliated to *Alterythrobacter* was characterized by a high number of biphenyl degradation Pfam modules and other peroxidase. Like *Moheibacter*, the role of *Altrerythrobacter* in lignin degradation is not proven, though it has proven to be able to degrade polyaromatic alkanes(Maeda et al. 2018; Teramoto et al. 2010).

Unlike the two precedent clusters, cluster 3 (yellow in Figure 7) was characterized by MAGs whose functionnal profile is skewed towards carbohydrates metabolism, particularly for pectin and chitin degradation. It was largely dominated by Bacteroidetes including Refined_154, affiliated to *Chitinophagaceae* uncl. *Chitinophagaceae* are known to be chitin and cellulose degraders(Rosenberg 2014).The composition of the cluster is completed by Proteobacteria (including Refined_151 affiliated to *Citrobacter*).Citrobacter have demonstrated an ability to degrade mono and polysacharides(Cortes-Tolalpa et al. 2020) .

Finally, cluster 4 (in red in Figure 7) comprised only by 5 MAGs with a strong focus on carbohydrate metabolism. These MAGs were characterized by a high content of phosphorylases, cellulase, hemicellulase, oligosaccharide degrading and debranching enzymes.

In order to better understand the role palyed by each cluster on the major functional properties of the NE15 consortium, we computed the contribution of each cluster to the temporal dynamics of the abudance of predicted proteins involved in lignocellulose degradation (Figure 8).

Firstly, the protein abundance of oligosaccharides degrading, oligosaccharaide-debranching and pectin-degrading enzymes displayed a similar trend, with an strong increase (by 150-450%) from day 0 to day 8. The MAGs in cluster 3 (yellow) and 4 (red) appear to contribute most to these activities, followed by the MAGs in cluster 1.

The abundance of cellulases and hemicellulases decreased during the first 2 days before increasing, showing a very similar dynamic which is in agreement with the measured CMCase and xylanase activities (Figure 5a). Not surprisingly, clusters 3 and 4 toghether contributed about 40% of the abundance of hemicellulases and 60% of cellulases. However, cluster 1 contributed 60% of the predicted hemicellulases

abundance. This is mostly due to the high number of hemicellulases genes in Refined_137 MAG.

Figure 7: Z-score normalized heatmap representing the prevalence of the different CAZymes modules and Pfam modules linked to aromatics metabolism in the 74 most abundant MAGs. MAGs were clustered using Ward's hierarchical clustering with Euclidean distance. The average abundance (Abund_avg) of the MAGs in the metagenome and their taxonomic affiliation are shown.

The contribution of cluster 3 and 4 to functions other than carbohydrates metabolism was much lower, as expected from the clustering. MAGs in clusters 1 (green) and 2 (blue) were the main contributors to activities related to aromatics metabolism. Interestingly, laccases and beta-etherases followed the same trend, i.e. an increase for the first 2 days before their pool abundance started to decrease. However, considering that laccase activity slightly increased until day 8 of incubation (Figure 5b), it is worth mentioning that the change in gene abundance may not be directly related to a change in enzyme production and activity, and that metatranscriptomics/metaproteomics data would be required to draw conclusions.

Interestingly, the abundance of predicted modules of dye-decolorizing peroxidases increased more than twofold in the first two days. This may indicate that DyP peroxidase are involved in the initial attack to LWS, facilitating the access to cellulose and hemicellulose. It reveals also the potential of DyP peroxidase to enhance lignin degradation in biotechnological processes(Falade et Ekundayo 2021).

Overall, the contribution of the MAGs to the temporal dynamics of functions relevant for lignocellulose deconstruction in the bioreactors shows that MAGs containing genes for oligosaccharide degradation/debranching, pectin degradation, laccase, bet-etherase and DyP-type peroxidase are the firsts to develop in the bioreactors, suggesting a potential role in the first days of incubation. On the other hand, MAGs with cellulase or hemicellulose activity appeared to be more active later. This could be explained by the need to remove some of the lignin as well as the polysaccharide-lignin linkages to access the cellulose or hemicellulose in a recalcitrant material such as LWS.

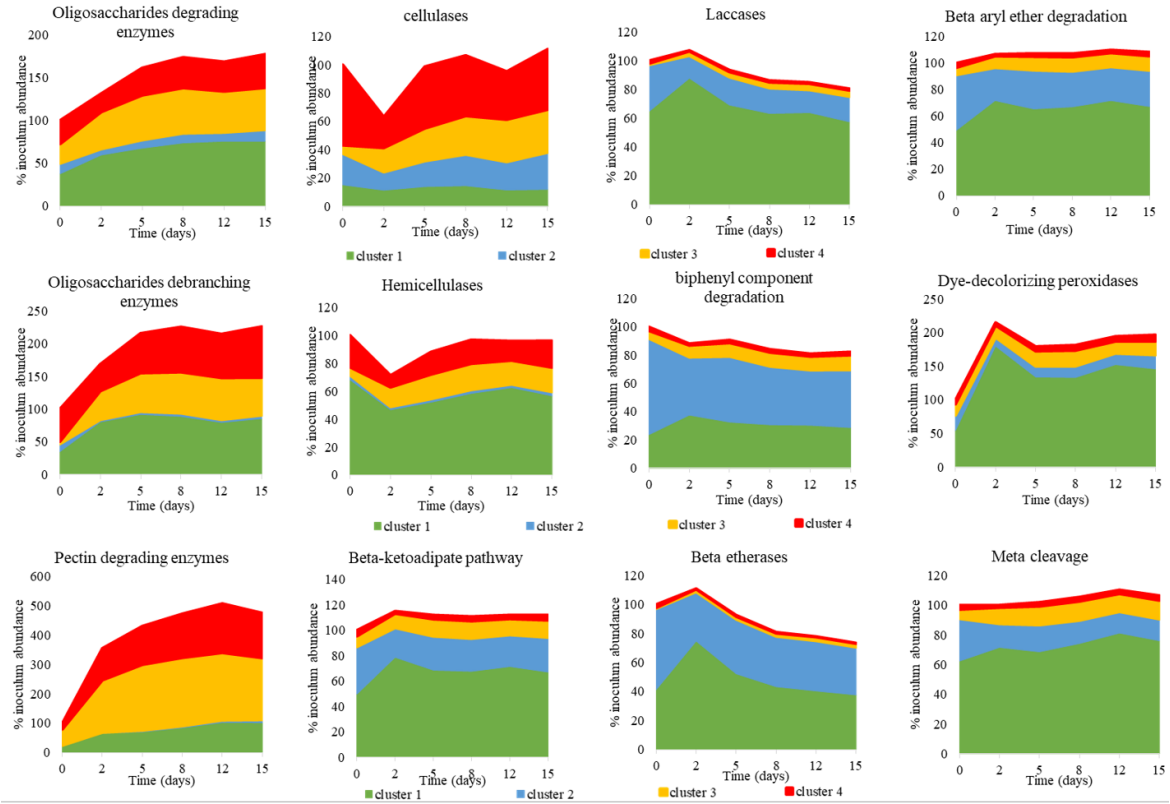


Figure 8: Evolution of the contribution of each cluster to the abundance of the predicted proteins from key functions. Abundances are expressed as % of clr abundance of the predicted proteins in the inoculum sample.

VI.5.4. Identification of networks of MAGs and functions

To further characterize the NE15 metagenome, we generated networks based on the temporal dynamics of the abundance (CLR-transformed data) of MAGs and predicted proteins (Figure 9).

The grouping of MAGs on their abundance profile made it possible to identify MAGs with potential synergistic effects. To exclude taxonomic biases in the construction of this network, the network map based on the temporal dynamics of MAGs abundance, colored according to their taxonomic affiliation is shown in the Supplementary file 8. This shows that the clusters were not built on the basis of taxonomy.

Two main clusters were identified (Figure 9a), one with 22 MAGs centered on Refined_8 and Refined_86, two MAGs from cluster 1, which appears to be antagonistic since their dynamics were negatively correlated. The other cluster, which contains 17 MAGs, is centered on Refined_60.

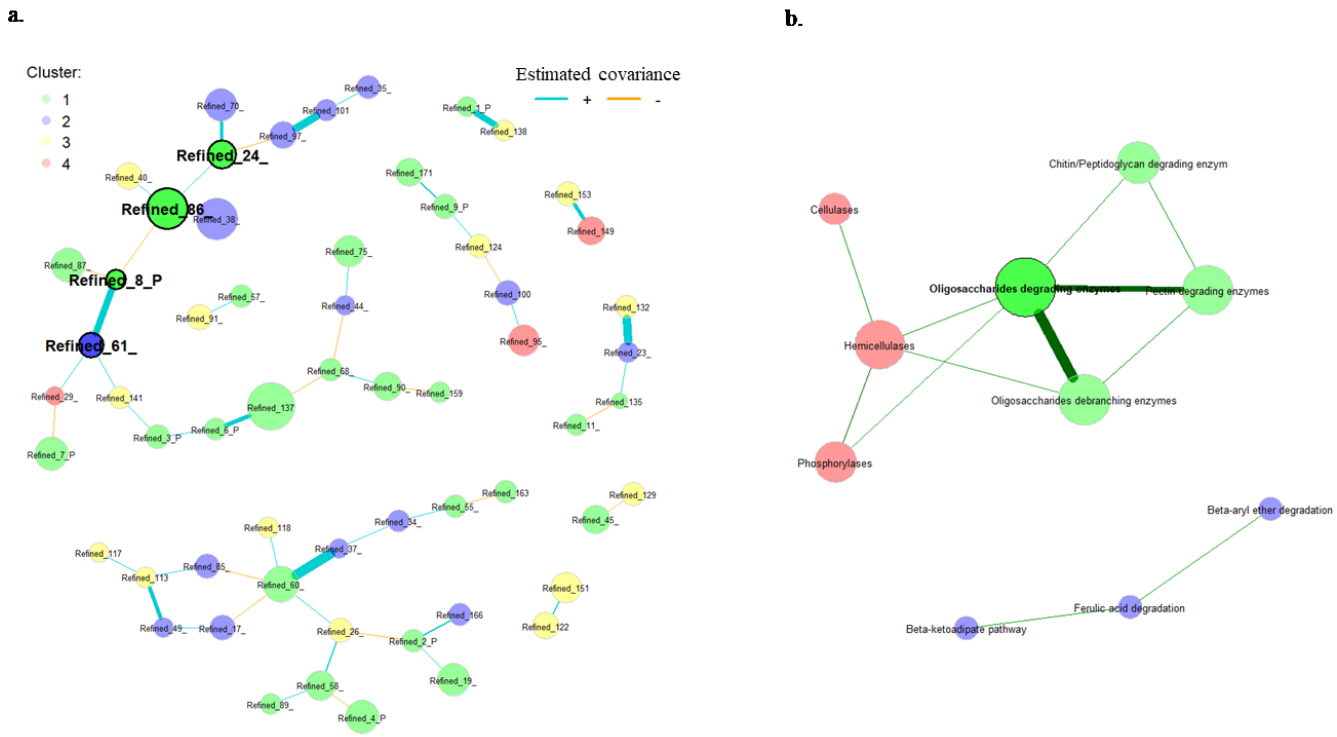


Figure 9: Network map inferred by SParse Inverse Covariance Estimation for Ecological Association Inference (SPIEC-EASI) (Kurtz et al. 2015) using NetCoMi, based on the temporal abundance of (**a**) MAGs and (**b**) predicted proteins**.**

Interestingly, all but two MAGs (Refined_151-Refined_122) displayed a direct covariance between MAGs from clusters related to aromatics metabolism (clusters 1 & 2) and to carbohydrates degradation (clusters 3 & 4). Furthermore, all MAGs from clusters 3 and 4, except Refined_45 and Refined_129, showed a synergistic relationship with at least one other MAG, suggesting that a synergistic effect for lignocellulose degradation could result from the combined action of functionally complementary MAGs.

The network analysis based on the abundance of predicted proteins supports the existence of two main functional classes in the NE15 bioreactor experiments. Indeed, a first cluster could be identified for carbohydrates metabolism, which shows that most functions covary, consistent with the strong specialization towards carbohydrate metabolisms of most of the MAGs involved in these activities (Figure 7). Although starch degradation activities are related to carbohydrates degradation, they were no part of this network (Figure 9b), suggesting that they were not essential for LWS degradation by the NE15 consortium. For functions related to the metabolism of aromatic compounds, only those related to beta-ketoadipate pathway, beta-aryl ether and ferulate degradation were clustered together.

VI.6. DISCUSSION

To characterize the potential lignolytic activity of a selected microbial consortium on lignin rich wheat straw (LWS) in aerobic bioreactors, we combined wet-chemistry, 2D HSQC NMR and ^{13}C-IS py-GC-MS measurement with metagenomics analysis.

By using wet chemistry coupled with advanced structural analysis, we provided new information on the lignin degrading and modifying potential of the NE15 consortium. Collectively, our results suggest that the polymeric lignin in LWS was partially modified and/or depolymerized by the termite-derived NE15 consortium when implemented in aerobic bioreactors. This resulted in a 21% lignin removal by the consortium within 15 days. Furthermore, we provide strong evidence that the soluble lignin-oligomers were almost entirely consumed by the consortium. As non-inoculated controls did not display such behavior, this indicates that removal of lignin in the bioreactors can only be explained by the degradation capacity of NE15.

Nevertheless, in line with previous observations,(Tarmadi et al. 2018; Dumond et al. 2021) NE15 consortium preferentially used polysaccharides, rather than lignin, as preferential carbon sources (Figure 1 & 2 and Table 1). However, compared to anaerobic conditions(Dumond et al. 2021), no particular difference was found in the degradation of cellulose versus suggesting that both polymers can be equally consumed by the consortium.

Beside the degrading potential, lignin-modifying potential of NE15 was also assessed. 2D HSQC NMR and ^{13}C-IS py-GC-MS. The total S lignin units were found to decrease slightly while G units increased inducing a slight increase of the S/G ratio though only marginally. However, 2D HSQC NMR demonstrated that oxidized S and G units (S' and G') clearly increased during the incubation, suggesting oxidative properties to the consortium.

Interestingly, the proportion of β-O-4 linkages was not modified by NE15. Instead C-C linkages, mostly β-β linkages diminished over the incubation time. This suggested that C-C linkages were more readily broken down than β-O-4 linkages (Figure 3). Such behavior has been observed in the microbiome of other termites such as *Odontermes*(Hongjie Li et al. 2017; Tarmadi et al. 2018).

Over the incubation time, a clear increase in cinnamyl alcohol, cinnamaldehyde and HPV/HPS end units was observed (Figure 3). The HPV/HPS products have been described for lignin degradation by the action of β-etherases, enzymes that degrade lignin in a reductive cascade.(Higuchi et al. 2018) The ability of β-etherases, particularly of glutathione-dependent-β-etherases cleaving β-aryl-ether lignin has been reported *Sphingobium* sp.(Gall et al. 2014) and demonstrated in *in-vitro* experiments(E. Masai et al. 1989; Gall et al. 2018) . Proteobacteria is the phylum which has thus far been found to harbour most of homologous glutathione-dependent-β-etherases genes. It has to be noted that it is also the case in this study as most MAGs harbouring β-etherases were described as Proteobacteria (Figure 7) (Kontur et al. 2019; Voß et al. 2020) .Overall, our results could suggest that bacteria harboring β-etherases genes may have been involved in the cleavage of β-O-4 linkages in the lignin polymer in the current bioreactor experiments.

We also demonstrated that oligomeric lignin compounds were completely metabolized by the consortium. This displayed how NE15, while able to degrade and modify lignin to some extent was much more efficient for the degradation of lignin-derived monomers and dimers. We hypothesize that lignolytic enzymes such as Dyp-type peroxidases and laccases that are found in NE15 could be involved in oligomeric lignin degradation.

Beside chemical analysis, an in-depth metagenomics analysis allowed us to gather evidence and propose hypothesis on the role of the various bacteria that constituted the

consortium. Numerous putative lignin-degrading proteins were found in the NE15 metagenome, including DyP peroxidase, laccase, β-etherases and enzymes from AA families associated with lignin depolymerization (Table 3 & Table 4). Interestingly, while DyP-like peroxidase were identified in the metagenome, no LiP or MnP-type peroxidase were found. This corroborates previous reports of the absence of genes encoding LiP or MnP in bacteria reported to have LiP or MnP-like activity(Davis et al. 2013) and the fact that LiP and MnP are exclusive to fungi(de Gonzalo et al. 2016). We found that DyP peroxidase are rather ubiquitous in members of the NE15 community, as modules were found not only in Actinobacteria and γ-Proteobacteria which carry the majority, but also in other Proteobacteria and Bacteroidetes (Figure 7). This is in line with the known wide distribution of DyP peroxidases in bacterial taxa.(Rai et al. 2021; Granja-Travez et al. 2020; Savelli et al. 2019) We also provide information supporting the existence of DyP peroxidase in organisms such as *Achromobacter*(Strachan, VanInsberghe, et Williams 2012) or *Leucobacter*(Ge, Ai, et Dong, s. d.) previously reported in only a few studies. However, although lignin degradation is one of the most studied application of these enzymes, their true role remains unclear and their function can vary considerably depending on their structure, as some may have hydrolase or oxidase functions. The physiological role of DyP peroxidases may also vary considerably, from lignin degradation or oxidation of recalcitrant phenolic compounds, to proteins or iron/copper transport, to involvement in the transition from vegetative mycelium to aerial hyphae development in certain Ascomycetes(Falade & Ekundayo 2021; Sugano & Yoshida 2021). Further research is needed to identify specific activities of the DyP peroxidases found in the NE15 consortium, which could have potential application in lignin upgrading or bioremediation.

By hierarchical clustering of the constructed MAGs, we were able to simplify the functional profile of the NE15 metagenome to a smaller number of variables (Figure 7). This allowed us to highlight the complementarity of the MAGs presents in our metagenome. Most interestingly, almost all MAGs identified as aromatics degraders interacted positively with MAGs labelled as carbohydrates degraders. This corroborates the advantageous synergistic properties of bacterial consortia over pure strains.

Based on the collective results on the temporal dynamics of NE15 along LWS transformation, we hypothesize that oligosaccharides degrading/debranching enzymes and DyP peroxidase/laccases played a crucial role in the first days of incubation by removing oligosaccharides decorations of cellulose and hemicellulose as well as by modifying the polymeric lignin through oxidative process, unblocking cellulose and hemicellulose. This may have favored the action of MAGs with cellulases or hemicellulases activity to access their substrate.

To have more specific information on the role of each identified MAGs, further analyses would be required. The use of metatranscriptomic or metaproteomic would be necessary to have more consequential information on the functional dynamics inside the consortium and would allow us to point directly at the most promising MAGs for further elucidation. A deeper characterization of the DyP peroxidase and other potential lignin degrading enzymes sequences would also be required to strengthen our comprehension of this system which would bring new information and contribute to their potential application within a lignocellulose-based biorefinery.

VI.7. ACKNOWLEDGMENT

The authors thank Dr. Hironori Kaji and Ms. Ayaka Maeno (ICR, Kyoto University) for their support in NMR experiments. This study was supported in part by PHC Sakura Program from MAEDI-MENESR and the Japan Society for the Promotion of Science (JSPS), JSPS KAKENHI grants (grant no. JP#16H06198 and JP#20H03044) and a research grant from Exploratory Research on Sustainable Humanosphere Science from RISH, Kyoto University. A part of this study was conducted using the facilities in the DASH/FBAS of RISH, Kyoto University, and the NMR spectrometer at the JURC of ICR, Kyoto University. This work has received funding from the Bio Based Industry Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement no. 720303 (EU-Zelcor, Zero waste ligno-cellulosic biorefineries by integrated lignin valorization). It was also supported by the French National Institute of Research for Agriculture, Food and Environment (INRAE), the Region Languedoc-Roussillon Midi-Pyrénées grant 31000553, the Carnot Institute 3BCAR—Insyme project and the PHC Sakura program from Campus

France under the project N° 45094PM. The authors thank the GeT Plage platform of INRAE Toulouse, E. Mengelle, M. Bounouba and S. Dubos for the technical support with bioreactor experiments.

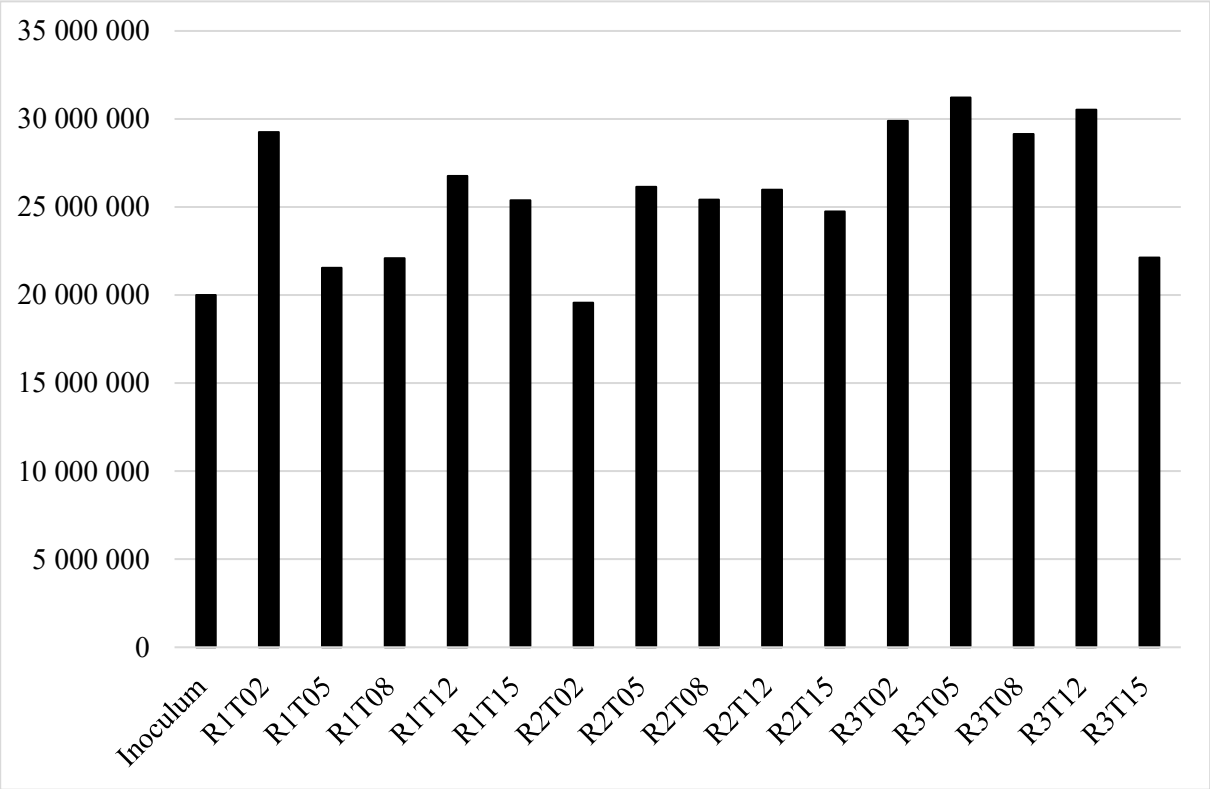
VI.7.1. Author Contributions

L.D. prepared the samples and carried out and analyzed the enrichment of the consortium, the culture in bioreactor, wet chemistry experiments and the analysis of the sequencing. VC participated in preparing samples for supernatant analysis. BC and SB performed and analyzed HPSEC and LC-MS experiments. DSD and EM provided expertise on termites and welcomed LD for termite gut extractions. CH provided guidance for metagenome sequencing analysis and participated in LD supervision. PYL and YT carried out and analyzed 2D HSQC NMR experiments. GvE and MK performed and analyzed py-GC-MS experiments. VL and BH provided expertise regarding the CAZy database and VL performed the CAZyme annotation. As project coordinator, GHR designed the study, participated on experimental design and contributed at all stages. The manuscript was written by LD under GHR supervision and with important intellectual contributions from all authors. All authors have given approval to the final version of the manuscript. GHR and YT participated in funding this research.

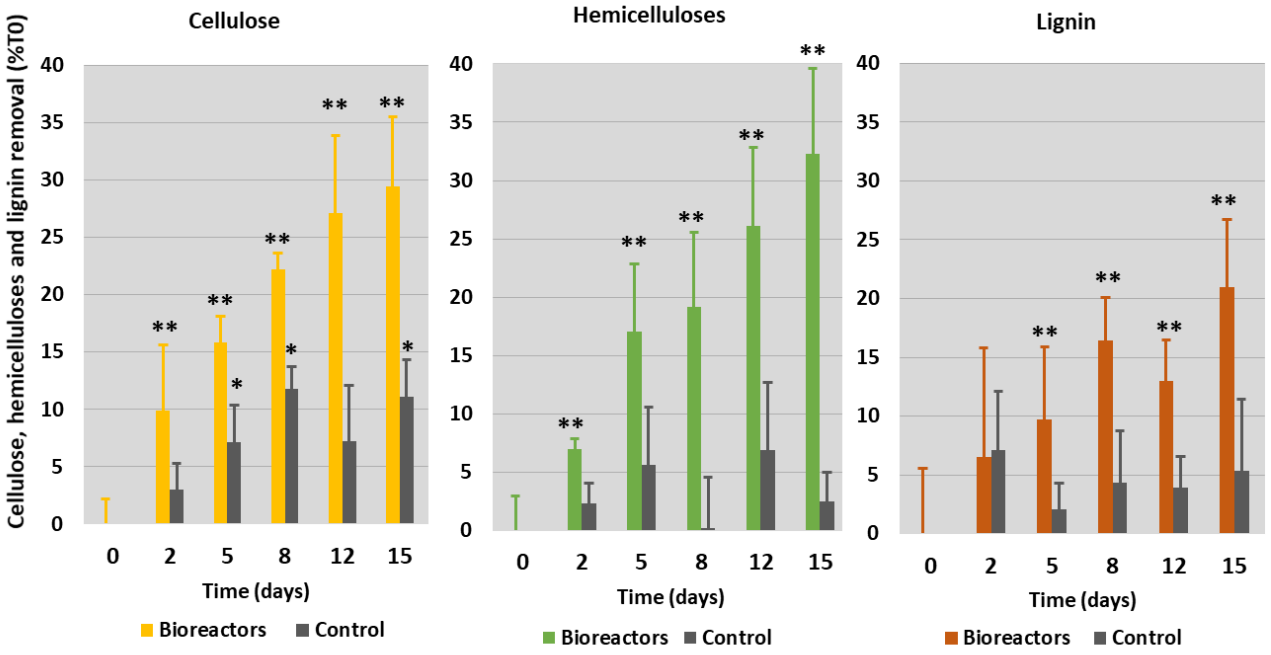
Conflicts of interest

The authors declare no competing financial interest.

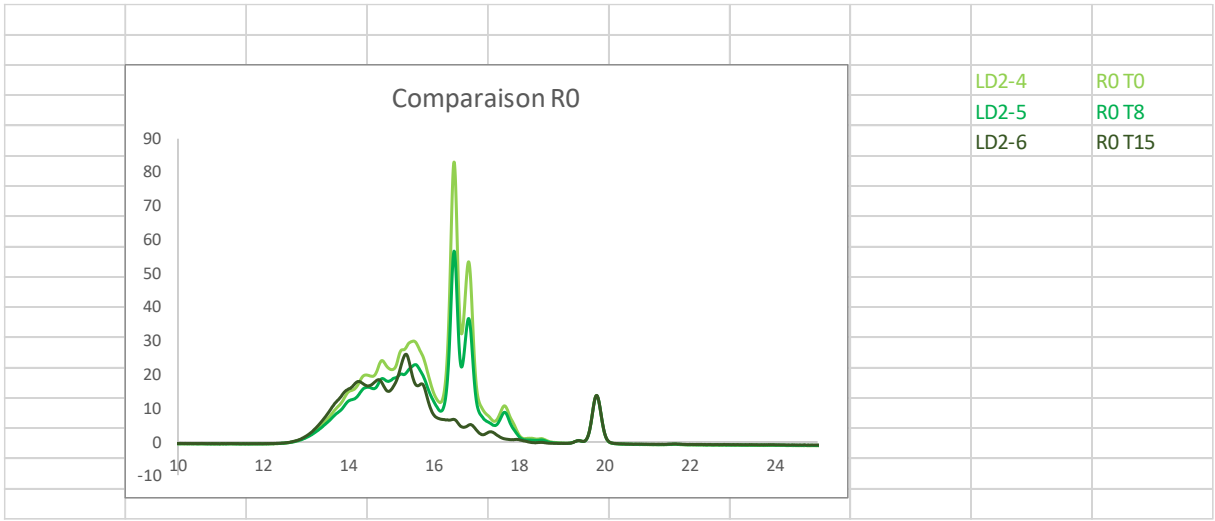
VI.8. SUPPLEMENTARY INFORMATION



Supplementary data 1: Number of sequence by sample



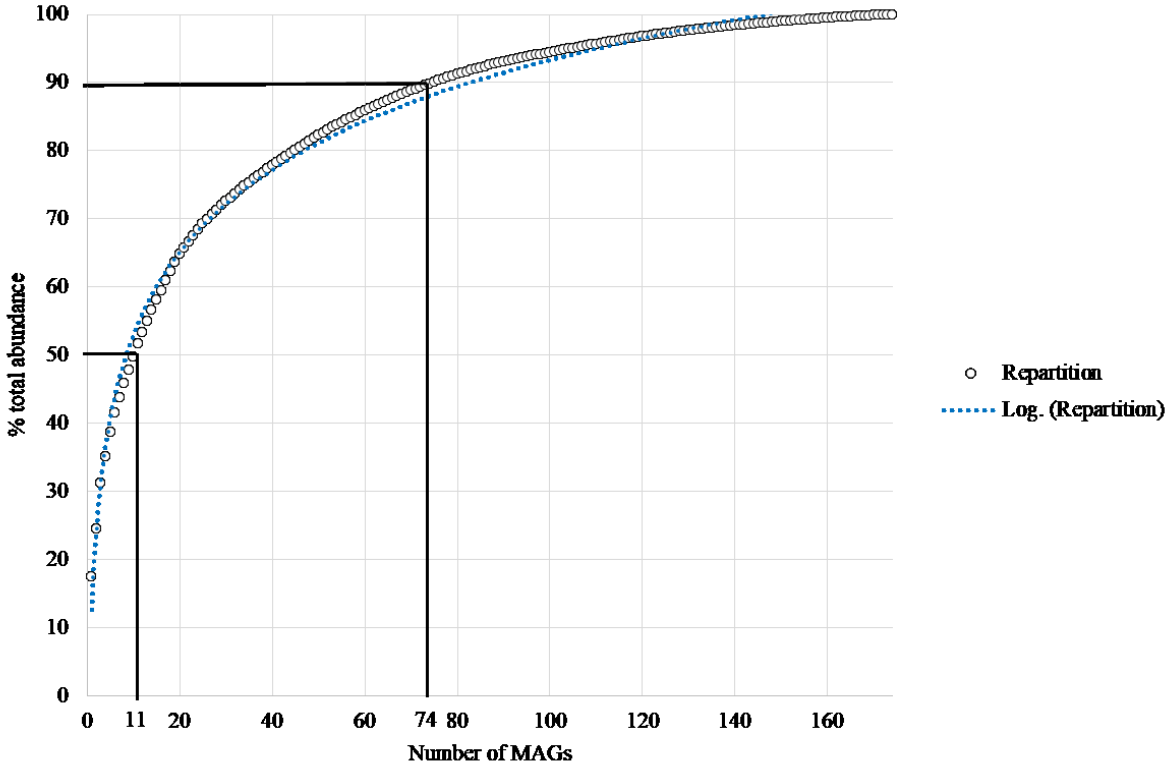
Supplementary data 2: Average cellulose, hemicellulose and lignin removal in the bioreactors compared to the non-inoculated control (Significant differences between inoculated and control samples is marked by ** ($p<0.05$), Significant differences between samples and T0 were marked by * ($p<0.05$))



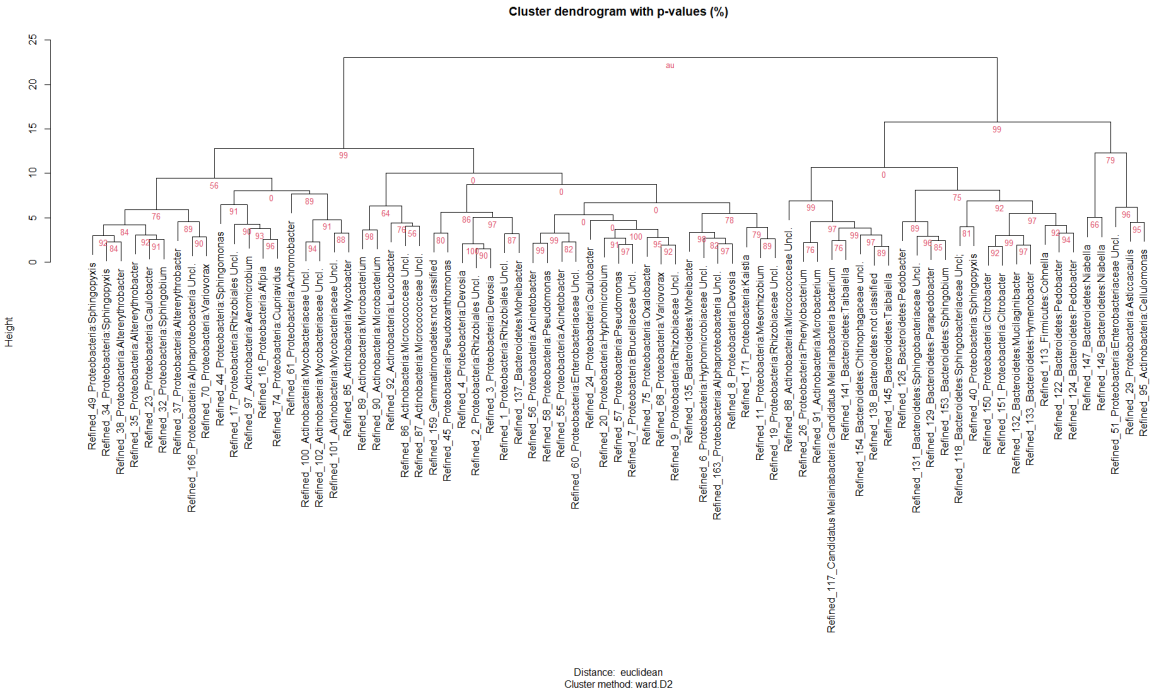
Supplementary 3: LC/MS spectra of control samples at T0, T8 and T15

component or DP	**Time**
vanilline	18
G ou AC	16.9
acide férulique	16.4
coniférine	16
DP2	
GG bb	16.1
GG 55	15.7
GG b5	15.8
DP3	14.9
DP4	14.6
DP6	14.1

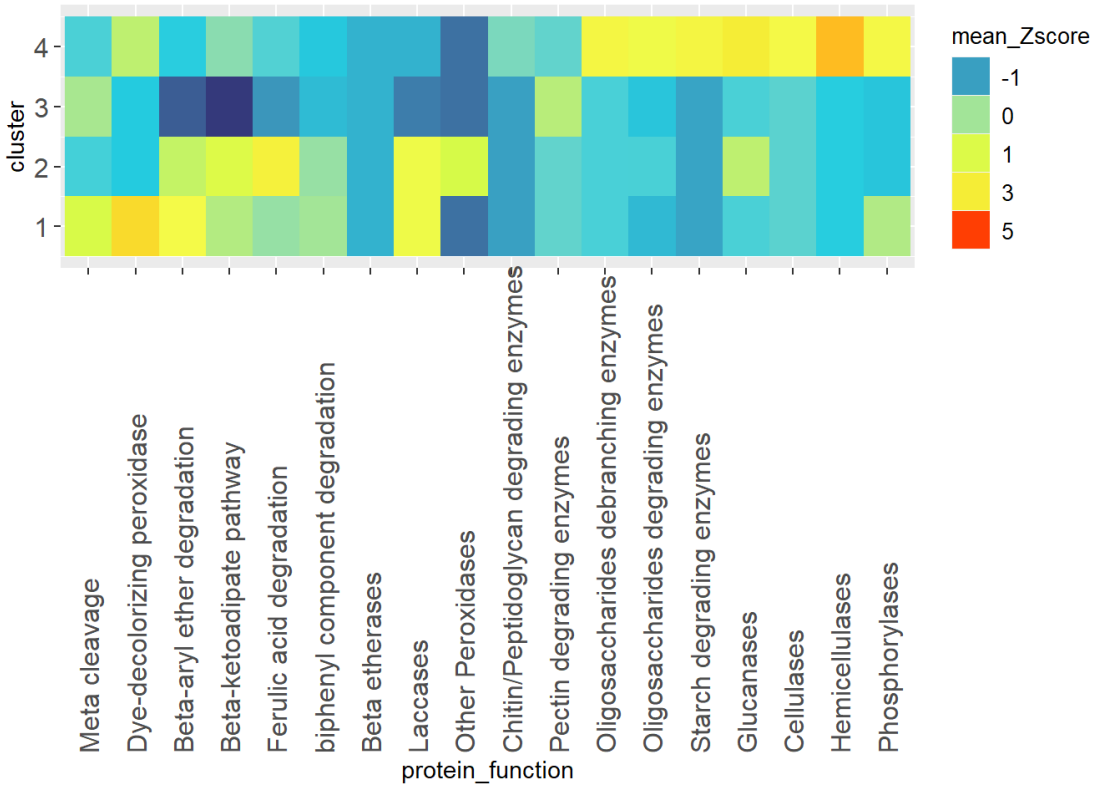
Supplementary data 4: Retention time of the different identified products



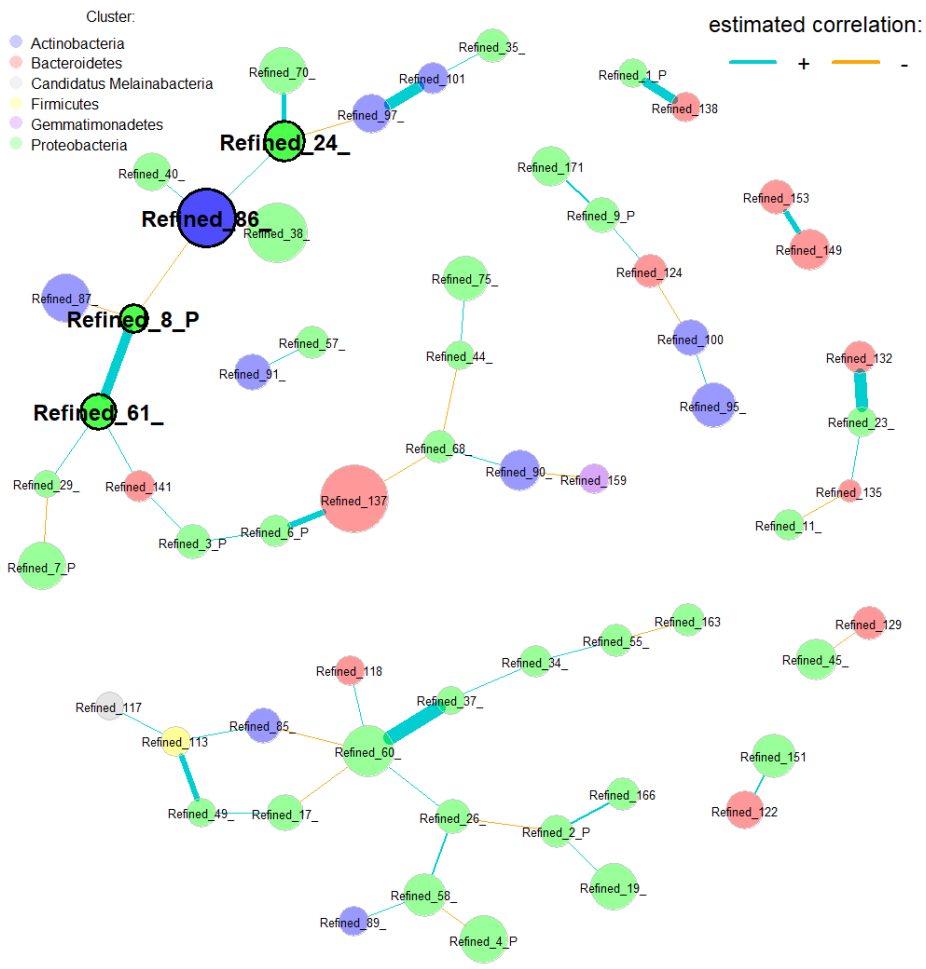
Supplementary5: Repartition of the MAGs abundance



Supplementary 6: AU (approximately unbiased) *p*-values for each of the cluster branches



Supplementary 7: Average Z-score of the clusters



Supplementary 8: SpiecEasi inferred network map based on MAGs abundance over time

VII.

CONCLUSION

La conversion de la lignocellulose en molécules d'intérêt est un enjeu majeur pour la bioraffinerie. En particulier, la valorisation de la lignine reste encore aujourd'hui un obstacle à la viabilité de ces bioraffineries. Dans ce travail de recherche, nous avons essayé de comprendre si et comment des consortia microbiens issus de termites supérieurs pouvaient dégrader la lignine et apporter des solutions à cette problématique. Nous avons identifiés deux points d'achoppement à dépasser. Les premiers travaux devaient nous permettre de démontrer : (a) que ce type de consortia microbiens a bel et bien une activité lignolytique et, (b) qu'il est bien possible d'enrichir un consortia lignolytique à partir du système digestif de termites supérieurs. Une fois ce questionnement résolu, notre attention s'est portée sur la caractérisation du fonctionnement des consortia obtenus. Cette deuxième étape, plus exploratoire par nature, avait pour objectif : (a) de reconstruire les génomes de bactéries potentiellement lignolytiques, (b) d'identifier les enzymes impliquées, (c) de comprendre les mécanismes d'interactions se déroulant dans cet écosystème et l'avantage éventuel induite par leur utilisation dans un procédé. Pour réaliser cela, nous avons allié des méthodes d'enrichissement fonctionnel des consortia à l'analyse fine compositionnelle et structurelle de la biomasse et l'analyse par métagénomique « shotgun » des communautés microbiennes enrichies.

Dans un premier temps, nous avons essayé de répondre à une question qui restait en suspens : Le microbiome intestinal des termites supérieurs est-il capable de dégrader la lignine ? En effet, depuis près d'un siècle, les hypothèses se sont suivies, les résultats n'arrivant pas à générer un consensus. De plus, les diverses études menées sur le sujet portaient sur les termites en condition réelle. Pour la première fois, nous avons pu démontrer la capacité lignolytique du seul microbiome intestinal. Pour cela, le microbiome intestinal de 4 espèces de termites supérieurs a été mis en culture dans des bioréacteurs anaérobies. Nous avons analysé la structure de la paille de blé avant et après incubation de ce microbiome, en utilisant des techniques de chimie analytique (hydrolyse acide, thioacidolyse) et les techniques modernes d'analyse de la lignocellulose que sont la pyrolyse couplée à la chromatographie gazeuse et à la spectrométrie de masse (13C-IS py-GC-MS) et la spectrométrie RMN en deux dimensions (2D HSQC NMR). Conformément à la littérature, nous avons pu montrer

que ces consortia microbiens issus du système digestif des termites présentaient une préférence pour la dégradation de l'hémicellulose et de la cellulose. Cependant, nous avons pu observer une diminution de la masse de lignine présente dans les bioréacteurs allant jusqu'à 37%, couplée à des modifications structurelles de cette dernière. L'analyse structurelle a notamment démontré que ces consortia dégradent la tricine et qu'ils sont responsables d'une oxydation des liaisons C-C plutôt que C-O, ce qui a déjà été observé chez certains termites supérieurs. Ces résultats, prometteurs reposaient sur des hypothèses quant aux acteurs de la dégradation. Si l'implication de bactéries utilisant des DyP peroxydases et des β-etherases a été proposée, la présence de gènes codant pour ces enzymes n'a encore jamais été montrée chez les bactéries issues d'intestins de termites. Ce point restait donc à éclairer.

De plus, ces résultats, en démontrant la capacité d'un consortium bactérien issu d'intestin de termite à dégrader la lignine, nous ont servi de point de départ pour essayer d'améliorer la capacité de dégradation de la lignine par des consortia microbiens issus de termites. Etant donné que la plupart des organismes connus dégradant la lignine sont aérobies, nous avons voulu explorer cette piste sur un écosystème, pourtant, majoritairement anaérobie mais comportant des zones aérobies.

Des études précédentes menées dans l'équipe ont démontré la possibilité d'enrichir un consortium microbien issu de termite en activités cellulolytiques et hemicellulolytiques, tout en favorisant la production d'acides gras volatils. Après enrichissement, deux cinétiques de dégradation ont été obtenues et une analyse de l'évolution temporelle du consortium a été effectuée grâce à des méthodes de metabarcoding sur le gène codant pour l'ARNr 16S. En prenant cela comme point de départ, nous avons développé un pipeline d'analyse de séquençage « shotgun » pour approfondir l'analyse fonctionnelle de cette diversité. Cela nous a permis de reconstruire (au moins à 50%) les génomes de 99 bactéries. Cela nous a permis de confirmer la diversité obtenue grâce au séquençage du 16S, étant donné la similitude des résultats. De plus, cela nous a permis de mettre en avant la sous-représentation des Spirochaetes engendrée par le séquençage amplicon. L'utilisation du séquençage shotgun nous a surtout permis de mettre en avant la spécialisation des membres constituant le consortium. Ainsi, les Bacteroidetes présentaient l'essentiel des protéines prédites avec une action lignocellulolytique et ont par conséquent été

identifiés comme la principale force derrière le très bon potentiel lignocellulolytique de ces consortia. A l'inverse, les Proteobacteria présentaient l'essentiel des gènes liés à la production d'acides gras volatils. Cela nous a permis de montrer l'intérêt de l'utilisation de consortia microbiens et l'importance des relations de coopérations dans un tel écosystème. La recherche des gènes liés à la dégradation de la lignine nous a également permis d'identifier des β-etherases dont le rôle avait été proposé dans le chapitre II.

D'une manière générale, l'étude d'un consortium microbien issu des termites et sa culture en milieu anaérobie nous a permis d'apporter de nouvelles preuves quant à la capacité du microbiome intestinal des termites de dégrader la lignine et nous a permis de mieux comprendre la spécialisation du microbiome et les interactions bactériennes à l'œuvre.

En règle générale, si la culture anaérobie permet préférentiellement la dégradation de la cellulose et de l'hémicellulose ainsi que la production d'acides gras volatils, la dégradation de la lignine est plus généralement associées avec des organismes aérobies (champignons filamenteux, bactéries aérobies). Bien que le système digestif des termites soit majoritairement anaérobie, il a été démontré qu'en réalité un gradient d'oxygénation existait. Nous avons donc voulu vérifier cette hypothèse. De plus, le choix du substrat a une influence majeure sur la sélection microbienne. Nous avons donc considéré différents types de substrats pour enrichir un consortium microbien lignolytique aérobie.

Pour cela nous avons d'abord essayé de cultiver le microbiome intestinal des termites en utilisant des lignines techniques comme substrat. Cela n'a malheureusement pas donné les effets escomptés et nous avons pu montrer que ce type de consortia ne pouvait pas utiliser des lignines techniques comme seul substrat carboné. C'est ainsi que nous avons fait le choix d'utiliser la paille de blé, utilisée précédemment pour les cultures anaérobies, et de produire un substrat enrichi en lignine et appauvri en polysaccharides par un traitement enzymatiques. Dans ce substrat produit, appelé LWS, nous avons modifié la composition initiale de la paille de blé de 42:35:23 (cellulose:hemicellulose:lignine) à 36:31:33 dans le nouveau substrat, en retirant les polysaccharides facilement hydrolysables. Afin d'enrichir les

consortia en activités d'intérêt, à savoir leur capacité à dégrader la lignine, le microbiome des termites ont été cultivés pendant 15 ou 28 jours sur ces substrats (paille de blé ou LWS), puis des cycles de repiquages successifs en batch ont été réalisés. Nous avons ainsi pu démontrer que lorsque les consortia sont enrichi sur paille de blé, il perd progressivement sa faculté de dégrader la lignine. De plus, une durée de 28 jours d'incubation a un effet néfaste sur sa capacité à dégrader la lignocellulose, quel que soit le substrat utilisé. Finalement, il résulte de ces travaux d'enrichissement que l'on peut sélectionner un consortium lignolytique en réalisant des cycles d'enrichissement de 15 jours sur un substrat prétraité, enrichi en lignine (LWS).

Une analyse de l'évolution de la diversité au cours des cycles d'enrichissements nous a permis d'identifier des pistes d'explications pour comprendre le rôle des différents membres de la communauté dans la dégradation de la lignocellulose. L'un des résultats les plus importants est que la variation de diversité entre les enrichissements est notamment marquée par le rôle prépondérant des *Pseudomonadaceae* dans l'enrichissement conduisant à la meilleure dégradation de la lignine.

Après avoir enrichi un consortium capable de dégrader la lignine, dénommé NE15, nous avons analysé l'évolution temporelle de la dégradation de LWS par ce consortium. Pour cela, nous avons utilisé la méthodologie de métagénomique « shotgun » développée sur le consortium anaérobie.

Par l'analyse fine de la composition et structure du substrat, nous avons réussi à montrer que le consortium NE15 était capable, non seulement de dégrader jusqu'à 21% de la lignine totale dans les bioréacteurs. De plus, la quasi-totalité de la lignine soluble était consommée par le consortium. Les analyses métagénomiques ont démontré la présence de nombreux gènes codant pour des protéines liées à la dégradation de la lignine. En particulier, de nombreuses peroxydases de type DyP, des β-etherases et des oxidases possiblement impliquées dans la dégradation de la lignine (AAO, VAO etc.) ont été observées dans le métagénome de NE15. D'une manière générale, la distribution de ces enzymes est en accord avec la littérature, à savoir principalement présentes dans les génomes reconstruits (MAGs) affiliés à des Actinobacteria, α-Proteobacteria et γ-Proteobacteria. En outre, les données récoltées permettent

d'illustrer l'avantage de l'utilisation de consortia microbiens par rapport à des cultures pures en mettant en lumière les synergies entre bactéries cellulolytiques et lignolytiques.

Ainsi, nous avons pu montrer que des communautés microbiennes issues de termites supérieurs ont la capacité de dégrader la lignine et qu'il est possible d'enrichir un consortium dans cette optique. Nous avons également réalisé une caractérisation fine de ces consortia, nous permettant à la fois d'identifier de nouvelles espèces et enzymes potentiellement impliqués dans la dégradation de la lignine mais aussi de mieux comprendre les mécanismes d'interaction microbienne dans de tels écosystèmes.

La reconstruction de génome effectuée dans ce travail de thèse a permis d'identifier nombre de bactéries possédant un potentiel pour la valorisation de la lignine. Il serait utile de déterminer si une partie de ces bactéries peut être cultivée individuellement. Leur isolement permettrait d'évaluer le potentiel individuel des MAGs identifiés pour des applications biotechnologiques. En partant de ce principe, il pourrait également être intéressant d'isoler des bactéries avec des activités spécifiques pour reconstruire des consortia microbiens synthétiques conservant la totalité des fonctions identifiées.

L'analyse métagénomique effectuée n'a pas seulement permis d'identifier des bactéries mais également un nombre important de gènes codant pour des enzymes putativement lignolytiques. La caractérisation plus fine de ces gènes et la production des enzymes dans des bactéries recombinantes permettraient d'apporter des réponses plus définitives. Par exemple, il serait intéressant de comparer les séquences des protéines prédites à un ensemble de séquences connues pour leur efficacité de dégradation pour voir si elles partagent une structure ou une histoire évolutive similaire. Cela nous permettrait d'enrichir notre connaissance de ces enzymes. Une autre approche serait de construire des banques de fosmides contenant les séquences issues du métagénome et de faire un criblage d'activité sur des substrats spécifiques. Cette approche a en effet été mise en place dans le cadre du projet Zelcor et il est prévu de comparer les résultats obtenus. Enfin une troisième méthode consisterait en l'expression des gènes d'intérêt dans des bactéries recombinantes afin de pouvoir

produire les enzymes pour étudier leur structure plus en détail. Ces différentes pistes de réflexion apporteraient toutes des éclaircissements bienvenus sur la fonction réelle des protéines identifiées.

Que ce soit en milieu aérobie ou anaérobie, nous avons essayé de décrypter les évolutions temporelles du métagénome. Cependant, les outils utilisés ne sont pas nécessairement les plus puissants pour ce type d'analyse. Un séquençage métatranscriptomique nous aurait éclairé sur l'expression des gènes à chaque étape et nous aurait permis d'identifier les bactéries qui ont un impact direct sur la conversion de la lignine et à quel moment l'activité se déclenche. Dans le même ordre d'idée, une analyse métaprotéomique nous aurait éclairés sur la production des enzymes d'intérêt et leurs dynamiques temporelles. Des analyses métaprotéomiques ont été appliquées au métagénome du consortium anaérobie et pourraient, à l'avenir, être envisagées pour le métagénome du consortium aérobie NE15.

Une autre piste d'amélioration du travail effectué au cours de cette thèse concerne l'utilisation du consortium enrichi NE15. En effet, puisque nous avons réussi à enrichir un consortium capable de dégrader au moins partiellement une lignine naturelle, un approfondissement de la caractérisation de ce consortium pourrait être entrepris sur d'autres substrats. La capacité de ce consortium à dégrader des lignines techniques a été esquissée en annexe sur des fractions de faible poids moléculaires mais mériterait un traitement plus exhaustif. Le niveau réel de dégradation de lignines techniques polymériques reste notamment à être déterminé. En outre, il serait intéressant de mesurer à quel point la diversité du consortium NE15 pourrait être impactée par la culture sur de tels substrats (lignines techniques, fractions extraites à l'acétate d'éthyle).

Enfin, puisque l'origine de ce travail provient de la volonté d'améliorer les performances de bioraffineries de la biomasse lignocellulosique, et bien que l'ensemble des travaux aient été réalisés à l'échelle du laboratoire, une étude critique de la viabilité environnementale par l'analyse de cycle de vie serait un atout majeur pour évaluer les points clés d'amélioration des procédés mis en place.

VIII.

REFERENCES

Abdul Rahman, Nurdyana, Donovan H. Parks, Inka Vanwonterghem, Mark Morrison, Gene W. Tyson, et Philip Hugenholtz. 2016. « A Phylogenomic Analysis of the Bacterial Phylum Fibrobacteres ». Frontiers in Microbiology 6: 1469.

Agler, Matthew T., Brian A. Wrenn, Stephen H. Zinder, et Largus T. Angenent. 2011. « Waste to bioproduct conversion with undefined mixed cultures: the carboxylate platform ». Trends in Biotechnology 29 (2): 70-78.

Ahmad, Mark, Joseph N. Roberts, Elizabeth M. Hardiman, Rahul Singh, Lindsay D. Eltis, et Timothy D. H. Bugg. 2011. « Identification of DypB from Rhodococcus Jostii RHA1 as a Lignin Peroxidase ». Biochemistry 50 (23): 5096-5107.

Ahmad, Mark, Charles R. Taylor, David Pink, Kerry Burton, Daniel Eastwood, Gary D. Bending, et Timothy D. H. Bugg. 2010. « Development of Novel Assays for Lignin Degradation: Comparative Analysis of Bacterial and Fungal Lignin Degraders ». Molecular BioSystems 6 (5): 815-21..

Ai, Binling, Xue Chi, Jia Meng, Zhanwu Sheng, Lili Zheng, Xiaoyan Zheng, et Jianzheng Li. 2016. « Consolidated Bioprocessing for Butyric Acid Production from Rice Straw with Undefined Mixed Culture ». Frontiers in Microbiology 7 (octobre).

Aita, Giovanna M., et Misook Kim. 2010. « Pretreatment Technologies for the Conversion of Lignocellulosic Materials to Bioethanol ». In Sustainability of the Sugar and Sugar−Ethanol Industries, 1058:117-45. ACS Symposium Series 1058. American Chemical Society.

Alessi, Anna M., Susannah M. Bird, Nicola C. Oates, Yi Li, Adam A. Dowle, Etelvino H. Novotny, Eduardo R. deAzevedo, et al. 2018. « Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies ». Biotechnology for Biofuels 11 (1): 166.

Allgaier, Martin, Amitha Reddy, Joshua I. Park, Natalia Ivanova, Patrik D'haeseleer, Steve Lowry, Rajat Sapra, et al. 2010. « Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community ». PLOS ONE 5 (1): e8812.

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z. Ijaz, Leo Lahti, Nicholas J. Loman, Anders F. Andersson, et Christopher Quince. 2014. « Binning Metagenomic Contigs by Coverage and Composition ». Nature Methods 11 (11): 1144-46.

Alonso, David Martin, Stephanie G. Wettstein, et James A. Dumesic. 2012. « Bimetallic Catalysts for Upgrading of Biomass to Fuels and Chemicals ». Chemical Society Reviews 41 (24): 8075-98.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, et D. J. Lipman. 1990. « Basic Local Alignment Search Tool ». Journal of Molecular Biology 215 (3): 403-10.

Álvarez, Consolación, Francisco Manuel Reyes-Sosa, et Bruno Díez. 2016. « Enzymatic hydrolysis of biomass from wood ». Microbial Biotechnology 9 (2): 149-56.

Amarasinghe, Shanika L, Matthew E Ritchie, et Quentin Gouil. 2021. « long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data ». GigaScience 10 (2).

Amarasinghe, Shanika L., Shian Su, Xueyi Dong, Luke Zappia, Matthew E. Ritchie, et Quentin Gouil. 2020. « Opportunities and challenges in long-read sequencing data analysis ». Genome Biology 21 (1): 30.

Artzi, Lior, Edward A. Bayer, et Sarah Moraïs. 2017. « Cellulosomes: Bacterial Nanomachines for Dismantling Plant Polysaccharides ». Nature Reviews Microbiology 15 (2): 83-95.

Atasoy, Merve, Ozge Eyice, Anna Schnürer, et Zeynep Cetecioglu. 2019. « Volatile Fatty Acids Production via Mixed Culture Fermentation: Revealing the Link between PH, Inoculum Type and Bacterial Composition ». Bioresource Technology 292 (novembre): 121889.

Auer, Lucas, Adèle Lazuka, David Sillam-Dussès, Edouard Miambi, Michael O'Donohue, et Guillermina Hernandez-Raquet. 2017. « Uncovering the Potential of Termite Gut Microbiome for Lignocellulose Bioconversion in Anaerobic Batch Bioreactors ». Frontiers in Microbiology 8: 2623.

Ayling, Martin, Matthew D. Clark, et Richard M. Leggett. 2019. « New Approaches for Metagenome Assembly with Short Reads ». Briefings in Bioinformatics, février.

Baaijens, Jasmijn A., Amal Zine El Aabidine, Eric Rivals, et Alexander Schönhuth. 2017. « De Novo Assembly of Viral Quasispecies Using Overlap Graphs ». Genome Research 27 (5): 835-48.

Bağcı, Caner, Sascha Patz, et Daniel H. Huson. 2021. « DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences ». Current Protocols 1 (3): e59..

Bajpai, Pratima. 2012. « Integrated Forest Biorefinery ». In Biotechnology for Pulp and Paper Processing, 375-402. Springer, Boston, MA.

Bastien, Géraldine, Grégory Arnal, Sophie Bozonnet, Sandrine Laguerre, Fernando Ferreira, Régis Fauré, Bernard Henrissat, et al. 2013. « Mining for Hemicellulases in the Fungus-Growing Termite Pseudacanthotermes Militaris Using Functional Metagenomics ». Biotechnology for Biofuels 6 (1): 78.

Bauer, Stefan, Hagit Sorek, Valerie D. Mitchell, Ana B. Ibáñez, et David E. Wemmer. 2012. « Characterization of Miscanthus giganteus Lignin Isolated by Ethanol Organosolv Process under Reflux Condition ». Journal of Agricultural and Food Chemistry 60 (33): 8203-12.

Beghini, Francesco, Lauren J McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, et al. 2021. « Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3 ». Édité par Peter Turnbaugh, Eduardo Franco, et C Titus Brown. eLife 10 (mai): e65088.

Bengtsson-Palme, Johan, Martin Hartmann, Karl Martin Eriksson, Chandan Pal, Kaisa Thorell, Dan Göran Joakim Larsson, et Rolf Henrik Nilsson. 2015. « Metaxa2: Improved Identification and Taxonomic Classification of Small and Large Subunit RRNA in Metagenomic Data ». Molecular Ecology Resources 15 (6): 1403-14.

Bentley, David R., Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, et al. 2008. « Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry ». Nature 456 (7218): 53-59.

Bentsen, Niclas Scott, Claus Felby, et Bo Jellesmark Thorsen. 2014. « Agricultural Residue Production and Potentials for Energy and Materials Services ». Progress in Energy and Combustion Science 40 (février): 59-73.

Berchem, Thomas, Olivier Roiseux, Caroline Vanderghem, Arnaud Boisdenghien, Guy Foucart, et Aurore Richel. 2017. « Corn Stover as Feedstock for the Production of Ethanol: Chemical Composition of Different Anatomical Fractions and Varieties ». Biofuels, Bioproducts and Biorefining 11 (3): 430-40.

Berlemont, R., N. Winans, D. Talamantes, H. Dang, et H.-W. Tsai. 2020. « MetaGeneHunt for Protein Domain Annotation in Short-Read Metagenomes ». Scientific Reports 10 (1): 7712.

Berlin, Alex. 2013. « Microbiology. No Barriers to Cellulose Breakdown ». Science (New York, N.Y.) 342 (6165): 1454-56.

Bertucci, Marie, Magdalena Calusinska, Xavier Goux, Corinne Rouland-Lefèvre, Boris Untereiner, Pau Ferrer, Patrick A. Gerin, et Philippe Delfosse. 2019. « Carbohydrate Hydrolytic Potential and Redundancy of an Anaerobic Digestion Microbiome Exposed to Acidosis, as Uncovered by Metagenomics ». Édité par Maia Kivisaar. Applied and Environmental Microbiology 85 (15).

Billings, Andrew F., Julian L. Fortney, Terry C. Hazen, Blake Simmons, Karen W. Davenport, Lynne Goodwin, Natalia Ivanova, et al. 2015. « Genome sequence and description of the anaerobic lignin-degrading bacterium Tolumonas lignolytica sp. nov. » Standards in Genomic Sciences 10 (novembre).

Boisvert, Sébastien, Frédéric Raymond, Elénie Godzaridis, François Laviolette, et Jacques Corbeil. 2012. « Ray Meta: Scalable de Novo Metagenome Assembly and Profiling ». Genome Biology 13 (12): R122.

Bokulich, Nicholas A., Sathish Subramanian, Jeremiah J. Faith, Dirk Gevers, Jeffrey I. Gordon, Rob Knight, David A. Mills, et J. Gregory Caporaso. 2013. « Quality-Filtering Vastly Improves Diversity Estimates from Illumina Amplicon Sequencing ». Nature Methods 10 (1): 57-59.

Bolger, Anthony M., Marc Lohse, et Bjoern Usadel. 2014. « Trimmomatic: a flexible trimmer for Illumina sequence data ». Bioinformatics 30 (15): 2114-20.

Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. « Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2 ». Nature Biotechnology 37 (8): 852-57.

Bowers, Robert M., Nikos C. Kyrpides, Ramunas Stepanauskas, Miranda Harmon-Smith, Devin Doud, T. B. K. Reddy, Frederik Schulz, et al. 2017. « Minimum Information about a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of Bacteria and Archaea ». Nature Biotechnology 35 (8): 725-31.

Bredon, Marius, Jessica Dittmer, Cyril Noël, Bouziane Moumen, et Didier Bouchon. 2018. « Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate ». Microbiome 6 (1): 162.

Brink, Daniel P., Krithika Ravi, Gunnar Lidén, et Marie F. Gorwa-Grauslund. 2019. « Mapping the Diversity of Microbial Lignin Catabolism: Experiences from the ELignin Database ». Applied Microbiology and Biotechnology 103 (10): 3979-4002.

Brune, A., E. Miambi, et J. A. Breznak. 1995. « Roles of Oxygen and the Intestinal Microflora in the Metabolism of Lignin-Derived Phenylpropanoids and Other Monoaromatic Compounds by Termites. » Applied and Environmental Microbiology 61 (7): 2688-95.

Brune, Andreas. 2014. « Symbiotic Digestion of Lignocellulose in Termite Guts ». Nature Reviews Microbiology 12 (3): 168-80..

———. 2018. « Methanogens in the Digestive Tract of Termites ». In (Endo)Symbiotic Methanogenic Archaea, édité par Johannes H. P. Hackstein, 81-101. Microbiology Monographs. Cham: Springer International Publishing.

Buchfink, Benjamin, Chao Xie, et Daniel H. Huson. 2015. « Fast and Sensitive Protein Alignment Using DIAMOND ». Nature Methods 12 (1): 59-60.

Bugg, Timothy D. H., Mark Ahmad, Elizabeth M. Hardiman, et Rahman Rahmanpour. 2011. « Pathways for Degradation of Lignin in Bacteria and Fungi ». Natural Product Reports 28 (12): 1883-96.

Burnum, Kristin E., Stephen J. Callister, Carrie D. Nicora, Samuel O. Purvine, Philip Hugenholtz, Falk Warnecke, Rudolf H. Scheffrahn, Richard D. Smith, et Mary S. Lipton. 2011. « Proteome Insights into the Symbiotic Relationship between a Captive Colony of Nasutitermes Corniger and Its Hindgut Microbiome ». The ISME Journal 5 (1): 161-64.

Butler, J. H. A, et J. C Buckerfield. 1979. « Digestion of lignin by termites ». Soil Biology and Biochemistry 11: 507-13.

Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, et Susan P Holmes. 2016. « DADA2: High resolution sample inference from Illumina amplicon data ». Nature methods 13 (7): 581-83.

Calusinska, Magdalena, Martyna Marynowska, Marie Bertucci, Boris Untereiner, Dominika Klimek, Xavier Goux, David Sillam-Dussès, et al. 2020. « Integrative Omics Analysis of the Termite Gut System Adaptation to Miscanthus Diet Identifies Lignocellulose Degradation Enzymes ». Communications Biology 3 (1): 1-12.

Campanaro, Stefano, Laura Treu, Panagiotis G. Kougias, Xinyu Zhu, et Irini Angelidaki. 2018. « Taxonomy of Anaerobic Digestion Microbiome Reveals Biases Associated with the Applied High Throughput Sequencing Strategies ». Scientific Reports 8 (1): 1-12.

Campanaro, Stefano, Laura Treu, Luis M. Rodriguez-R, Adam Kovalovszki, Ryan M. Ziels, Irena Maus, Xinyu Zhu, et al. 2020. « New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters ». Biotechnology for Biofuels 13 (1): 25.

Campbell, William George. 1930. « The chemistry of the white rots of wood ». Biochemical Journal 24 (4): 1235-43.

Cantarel, Brandi L., Pedro M. Coutinho, Corinne Rancurel, Thomas Bernard, Vincent Lombard, et Bernard Henrissat. 2009. « The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics ». Nucleic Acids Research 37 (suppl_1): D233-38.

Caro-Quintero, A., K. M. Ritalahti, K. D. Cusick, F. E. Löffler, et K. T. Konstantinidis. 2012. « The Chimeric Genome of Sphaerochaeta: Nonspiral Spirochetes That Break with the Prevalent Dogma in Spirochete Biology ». MBio 3 (3).

Case, Rebecca J., Yan Boucher, Ingela Dahllöf, Carola Holmström, W. Ford Doolittle, et Staffan Kjelleberg. 2007. « Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies ». Applied and Environmental Microbiology 73 (1): 278-88.

Cavaillé, Laëtitia, Estelle Grousseau, Mathieu Pocquet, Anne-Sophie Lepeuple, Jean-Louis Uribelarrea, Guillermina Hernandez-Raquet, et Etienne Paul. 2013. « Polyhydroxybutyrate Production by Direct Use of Waste Activated Sludge in Phosphorus-Limited Fed-Batch Culture ». Bioresource Technology 149 (décembre): 301-9.

Chao, Anne. 1984. « Nonparametric Estimation of the Number of Classes in a Population ». Scandinavian Journal of Statistics 11 (4): 265-70.

Chen, S., Z. Wen, W. Liao, C. Liu, R. L. Kincaid, J. H. Harrison, D. C. Elliott, M. D. Brown, et D. J. Stevens. 2005. « Studies into Using Manure in a Biorefinery Concept ». Applied Biochemistry and Biotechnology 124 (1-3): 999-1015

Cherubini, Francesco. 2010. « The biorefinery concept: Using biomass instead of oil for producing energy and chemicals ». Energy Conversion and Management 51 (7): 1412-21.

Chiu, Chun-Huo, et Anne Chao. 2016. « Estimating and comparing microbial diversity in the presence of sequencing errors ». PeerJ 4 (février).

Claesson, Marcus J., Qiong Wang, Orla O'Sullivan, Rachel Greene-Diniz, James R. Cole, R. Paul Ross, et Paul W. O'Toole. 2010. « Comparison of Two Next-Generation Sequencing Technologies for Resolving Highly Complex Microbiota Composition Using Tandem Variable 16S RRNA Gene Regions ». Nucleic Acids Research 38 (22): e200-e200.

Commichaux, Seth, Nidhi Shah, Jay Ghurye, Alexander Stoppel, Jessica A Goodheart, Guillermo G Luque, Michael P Cummings, et Mihai Pop. 2021. « A critical assessment of gene catalogs for metagenomic analysis ». Bioinformatics, no btab216 (avril).

Constant, Sandra, Hans L. J. Wienk, Augustinus E. Frissen, Peter de Peinder, Rolf Boelens, Daan S. van Es, Ruud J. H. Grisel, et al. 2016. « New Insights into the Structure and Composition of Technical Lignins: A Comparative Characterisation Study ». Green Chemistry 18 (9): 2651-65.

Cookson, L. J. 1987. « 14C-Lignin Degradation by Three Australian Termite Species ». Wood Science and Technology 21 (1): 11-25.

Cortes-Tolalpa, Larisa, Yanfang Wang, Joana Falcao Salles, et Jan Dirk van Elsas. 2020. « Comparative Genome Analysis of the Lignocellulose Degrading Bacteria Citrobacter freundii so4 and Sphingobacterium multivorum w15 ». Frontiers in Microbiology 11: 248.

Couger, Matthew B., Caroline Graham, et Babu Z. Fathepure. 2020. « Genome Sequence of Lignin-Degrading Arthrobacter Sp. Strain RT-1, Isolated from Termite Gut and Rumen Fluid ». Microbiology Resource Announcements 9 (3).

Cragg, Simon M, Gregg T Beckham, Neil C Bruce, Timothy DH Bugg, Daniel L Distel, Paul Dupree, Amaia Green Etxabe, et al. 2015. « Lignocellulose Degradation Mechanisms across the Tree of Life ». Current Opinion in Chemical Biology, Energy • Mechanistic biology, 29 (décembre): 108-19

Crawford, Don L., et Ronald L. Crawford. 1980. « Microbial degradation of lignin ». Enzyme and Microbial Technology 2 (1): 11-22.

Davis, Jennifer R., Lynne Goodwin, Hazuki Teshima, Chris Detter, Roxanne Tapia, Cliff Han, Marcel Huntemann, et al. 2013. « Genome Sequence of Streptomyces viridosporus Strain T7A ATCC 39115, a Lignin-Degrading Actinomycete ». Genome Announcements 1 (4): e00416-13.

DeAngelis, Kristen M., John M. Gladden, Martin Allgaier, Patrik D'haeseleer, Julian L. Fortney, Amitha Reddy, Philip Hugenholtz, et al. 2010. « Strategies for Enhancing the Effectiveness of Metagenomic-Based Enzyme Discovery in Lignocellulolytic Microbial Communities ». BioEnergy Research 3 (2): 146-58.

DeAngelis, Kristen M., Deepak Sharma, Rebecca Varney, Blake A. Simmons, Nancy G. Isern, Lye Meng Markillie, Carrie D. Nicora, et al. 2013. « Evidence Supporting Dissimilatory and Assimilatory Lignin Degradation in Enterobacter Lignolyticus SCF1 ». Frontiers in Microbiology 4.

Delannoy-Bruno, Omar, Chandani Desai, Arjun S. Raman, Robert Y. Chen, Matthew C. Hibberd, Jiye Cheng, Nathan Han, et al. 2021. « Evaluating Microbiome-Directed Fibre Snacks in Gnotobiotic Mice and Humans ». Nature 595 (7865): 91-95.

Delmont, Tom O., A. Murat Eren, Lorrie Maccario, Emmanuel Prestat, Özcan C. Esen, Eric Pelletier, Denis Le Paslier, Pascal Simonet, et Timothy M. Vogel. 2015. « Reconstructing rare soil microbial genomes using in situ enrichments and metagenomics ». Frontiers in Microbiology 6 (avril).

Desai, Aarti, Veer Singh Marwah, Akshay Yadav, Vineet Jha, Kishor Dhaygude, Ujwala Bangar, Vivek Kulkarni, et Abhay Jere. 2013. « Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data ». PLOS ONE 8 (4): e60204.

Deusch, Simon, Amélia Camarinha-Silva, Jürgen Conrad, Uwe Beifuss, Markus Rodehutscord, et Jana Seifert. 2017. « A Structural and Functional Elucidation of the Rumen Microbiome Influenced by Various Diets and Microenvironments ». Frontiers in Microbiology 8 (août).

Directive (EU) 2015/1513 of the European Parliament and of the Council of 9 September 2015 Amending Directive 98/70/EC Relating to the Quality of Petrol and Diesel Fuels and Amending Directive 2009/28/EC on the Promotion of the Use of Energy from Renewable Sources (Text with EEA Relevance). 2015. 239. Vol. OJ L. http://data.europa.eu/eli/dir/2015/1513/oj/eng.

Doi, R. H., et Y. Tamaru. 2001. « The Clostridium Cellulovorans Cellulosome: An Enzyme Complex with Plant Cell Wall Degrading Activity ». Chemical Record (New York, N.Y.) 1 (1): 24-32.

Domenek, Sandra, Abderrahim Louaifi, Alain Guinault, et Stéphanie Baumberger. 2013. « Potential of Lignins as Antioxidant Additive in Active Biodegradable Packaging Materials ». Journal of Polymers and the Environment 21 (3): 692-701.

Dou, Chang, Wilian F. Marcondes, Jessica E. Djaja, Renata Bura, et Rick Gustafson. 2017. « Can we use short rotation coppice poplar for sugar based biorefinery feedstock? Bioconversion of 2-year-old poplar grown as short rotation coppice ». Biotechnology for Biofuels 10 (juin).

Dumond, Louison, Pui Ying Lam, Gijs van Erven, Mirjam Kabel, Fabien Mounet, Jacqueline Grima-Pettenati, Yuki Tobimatsu, et Guillermina Hernandez-Raquet. 2021. « Termite Gut Microbiota Contribution to Wheat Straw Delignification in Anaerobic Bioreactors ». ACS Sustainable Chemistry & Engineering 9 (5): 2191-2202.

Dutilh, B. E., V. van Noort, R. T. J. M. van der Heijden, T. Boekhout, B. Snel, et M. A. Huynen. 2007. « Assessment of Phylogenomic and Orthology Approaches for Phylogenetic Inference ». Bioinformatics (Oxford, England) 23 (7): 815-24.

Edgar, Robert C. 2010. « Search and Clustering Orders of Magnitude Faster than BLAST ». Bioinformatics (Oxford, England) 26 (19): 2460-61.

2013. « UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads ». Nature Methods 10 (10): 996-98.

2016. « UCHIME2: Improved Chimera Prediction for Amplicon Sequencing ». https://www.biorxiv.org/content/10.1101/074252v1.

Eichorst, Stephanie A., Patanjali Varanasi, Vatalie Stavila, Marcin Zemla, Manfred Auer, Seema Singh, Blake A. Simmons, et Steven W. Singer. 2013. « Community Dynamics of Cellulose-Adapted Thermophilic Bacterial Consortia ». Environmental Microbiology 15 (9): 2573-87.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, et Tom O. Delmont. 2015. « Anvi'o: an advanced analysis and visualization platform for 'omics data ». PeerJ 3 (octobre).

Erven, Gijs van, Roelant Hilgers, Pieter de Waard, Erik-Jan Gladbeek, Willem J. H. van Berkel, et Mirjam A. Kabel. 2019. « Elucidation of In Situ Ligninolysis Mechanisms of the Selective White-Rot Fungus Ceriporiopsis subvermispora ». ACS Sustainable Chemistry & Engineering 7 (19): 16757-64

Erven, Gijs van, Nazri Nayan, Anton S. M. Sonnenberg, Wouter H. Hendriks, John W. Cone, et Mirjam A. Kabel. 2018. « Mechanistic Insight in the Selective Delignification of Wheat Straw by Three White-Rot Fungal Species through Quantitative 13C-IS Py-GC-MS and Whole Cell Wall HSQC NMR ». Biotechnology for Biofuels 11: 262.

Erven, Gijs van, Ries de Visser, Donny W. H. Merkx, Willem Strolenberg, Peter de Gijsel, Harry Gruppen, et Mirjam A. Kabel. 2017. « Quantification of Lignin and Its Structural Features in Plant Biomass Using 13C Lignin as Internal Standard for Pyrolysis-GC-SIM-MS ». Analytical Chemistry 89 (20): 10907-16..

Erven, Gijs van, Ries de Visser, Pieter de Waard, Willem J. H. van Berkel, et Mirjam A. Kabel. 2019. « Uniformly 13C Labeled Lignin Internal Standards for Quantitative Pyrolysis−GC−MS Analysis of Grass and Wood ». ACS Sustainable Chemistry & Engineering 7 (24): 20070-76.

Erven, Gijs van, Jianli Wang, Peicheng Sun, Pieter de Waard, Jacinta van der Putten, Guus E. Frissen, Richard J. A. Gosselink, et al. 2019. « Structural Motifs of Wheat Straw Lignin Differ in Susceptibility to Degradation by the White-Rot Fungus Ceriporiopsis subvermispora ». ACS Sustainable Chemistry & Engineering 7 (24): 20032-42.

Escudié, Frédéric, Lucas Auer, Maria Bernard, Mahendra Mariadassou, Laurent Cauquil, Katia Vidal, Sarah Maman, Guillermina Hernandez-Raquet, Sylvie Combes, et Géraldine Pascal. 2017. « FROGS: Find, Rapidly, OTUs with Galaxy Solution ». Bioinformatics (Oxford, England), décembre.

« European Biorefinery Joint Strategic Research Roadmap Star-coliBRi StRatEgic taRgEtS foR 2020 – collaBoRation initiativE on BioREfinERiES (PDF Download Available) ». s. d. Consulté le 9 février 2018. https://www.researchgate.net/publication/257896346_European_Biorefinery_Joint_Strategic_Research_Roadmap_Star-coliBRi_StRatEgic_taRgEtS_foR_2020_-_collaBoRation_initiativE_on_BioREfinERiES?ev=auth_pub.

Falade, A. O., et T. C. Ekundayo. 2021. « Emerging Biotechnological Potentials of DyP-Type Peroxidases in Remediation of Lignin Wastes and Phenolic Pollutants: A Global Assessment (2007-2019) ». Letters in Applied Microbiology 72 (1): 13-23.

Falentin, Hélène, Lucas Auer, Mahendra Mariadassou, Géraldine Pascal, Olivier Rué, ERIC DUGAT-BONY, Céline Delbes, et al. 2019. « Guide pratique à destination des biologistes, bioinformaticiens et statisticiens qui souhaitent s'initier aux analyses métabarcoding ». Cahiers des Techniques de l'INRA 2019 (97): 1-23.

« FAOSTAT ». s. d. Consulté le 21 mars 2018. http://www.fao.org/faostat/en/#data/EL.

« FASTQC. A quality control tool for high throughput sequence data | BibSonomy ». s. d. Consulté le 22 août 2021. https://www.bibsonomy.org/bibtex/f230a919c34360709aa298734d63dca3.

Fichot, Erin B., et R. Sean Norman. 2013. « Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform ». Microbiome 1 (1): 10.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, et Weizhong Li. 2012. « CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data ». Bioinformatics (Oxford, England) 28 (23): 3150-52.

Gabrielle Benoît, Nguyen The Nicolas, Maupu Pauline, et Vial Estelle. 2012. « Life cycle assessment of eucalyptus short rotation coppices for bioenergy production in southern France ». GCB Bioenergy 5 (1): 30-42.

Gall, Daniel L., Wayne S. Kontur, Wu Lan, Hoon Kim, Yanding Li, John Ralph, Timothy J. Donohue, et Daniel R. Noguera. 2018. « In Vitro Enzymatic Depolymerization of Lignin with Release of Syringyl, Guaiacyl, and Tricin Units ». Applied and Environmental Microbiology 84 (3).

Gall, Daniel L., John Ralph, Timothy J. Donohue, et DanielR. Noguera. 2014. « A Group of Sequence-Related Sphingomonad Enzymes Catalyzes Cleavage of β-Aryl Ether Linkages in Lignin β-Guaiacyl and β-Syringyl Ether Dimers ». Environmental Science & Technology 48 (20): 12454-63..

Ge, Shimei, Wenjing Ai, et Xinjiao Dong. s. d. « High-Quality Draft Genome Sequence of Leucobacter sp. Strain G161, a Distinct and Effective Chromium Reducer ». Genome Announcements 4 (1): e01760-15.

Geib, Scott M., Timothy R. Filley, Patrick G. Hatcher, Kelli Hoover, John E. Carlson, Maria del Mar Jimenez-Gasco, Akiko Nakagawa-Izumi, Rachel L. Sleighter, et Ming Tien. 2008. « Lignin Degradation in Wood-Feeding Insects ». Proceedings of the National Academy of Sciences 105 (35): 12932-37.

Glenn, J. K., et M. H. Gold. 1985. « Purification and Characterization of an Extracellular Mn(II)-Dependent Peroxidase from the Lignin-Degrading Basidiomycete, Phanerochaete Chrysosporium ». Archives of Biochemistry and Biophysics 242 (2): 329-41.

Godin, Bruno, François Ghysel, Richard Agneessens, Thomas Schmit, Sébastien Gofflot, Stéphane Lamaudière, Georges Sinnaeve, et al. 2010. « Détermination de la cellulose, des hémicelluloses, de la lignine et des cendres dans diverses cultures lignocellulosiques dédiées à la production de bioéthanol de deuxième génération ». BASE, novembre.

Godon, Jean-Jacques, Laure Arcemisbéhère, Renaud Escudié, Jérôme Harmand, Edouard Miambi, et Jean-Philippe Steyer. 2013. « Overview of the Oldest Existing Set of Substrate-Optimized Anaerobic Processes: Digestive Tracts ». BioEnergy Research 6 (3): 1063-81.

Gonzalo, Gonzalo de, Dana I. Colpa, Mohamed H. M. Habib, et Marco W. Fraaije. 2016. « Bacterial Enzymes Involved in Lignin Degradation ». Journal of Biotechnology 236 (octobre): 110-19.

Goodwin, Sara, John D. McPherson, et W. Richard McCombie. 2016. « Coming of Age: Ten Years of next-Generation Sequencing Technologies ». Nature Reviews Genetics 17 (6): 333-51.

Granja-Travez, Rommel Santiago, Gabriela Felix Persinoti, Fabio M. Squina, et Timothy D. H. Bugg. 2020. « Functional Genomic Analysis of Bacterial Lignin Degraders: Diversity in Mechanisms of Lignin Oxidation and Metabolism ». Applied Microbiology and Biotechnology 104 (8): 3305-20. h

Green, T. R. 1977. « Significance of Glucose Oxidase in Lignin Degradation ». Nature 268 (5615): 78-80.

Greenblum, Sharon, Peter J. Turnbaugh, et Elhanan Borenstein. 2012. « 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 of the United States of America 109 (2): 594-99.

Grieco, Maria B., Fabyano A. C. Lopes, Louisi S. Oliveira, Diogo A. Tschoeke, Claudia C. Popov, Cristiane C. Thompson, Luna C. Gonçalves, et al. 2019. « Metagenomic Analysis of the Whole Gut Microbiota in Brazilian Termitidae Termites Cornitermes Cumulans, Cyrilliotermes Strictinasus,

Syntermes Dirus, Nasutitermes Jaraguae, Nasutitermes Aquilinus, Grigiotermes Bequaerti, and Orthognathotermes Mirim ». Current Microbiology 76 (6): 687-97.

Guidi, Werther, Frédéric E. Pitre, et Michel Labrecque. 2013. « Short-Rotation Coppice of Willows for the Production of Biomass in Eastern Canada ». Biomass Now - Sustainable Growth and Use.

Guo, Peng, Wanbin Zhu, Hui Wang, Yücai Lü, Xiaofen Wang, Dan Zheng, et Zongjun Cui. 2010. « Functional Characteristics and Diversity of a Novel Lignocelluloses Degrading Composite Microbial System with High Xylanase Activity ». Journal of Microbiology and Biotechnology 20 (2): 254-64.

Haitjema, Charles H., Sean P. Gilmore, John K. Henske, Kevin V. Solomon, Randall de Groot, Alan Kuo, Stephen J. Mondo, et al. 2017. « A Parts List for Fungal Cellulosomes Revealed by Comparative Genomics ». Nature Microbiology 2 (8): 1-8.

Hallmaier-Wacker, Luisa K., Simone Lüert, Sabine Gronow, Cathrin Spröer, Jörg Overmann, Nicky Buller, Rebecca J. Vaughan-Higgins, et Sascha Knauf. 2019. « A Metataxonomic Tool to Investigate the Diversity of Treponema ». Frontiers in Microbiology 10 (septembre).

Hamelinck, Carlo N, Geertje van Hooijdonk, et André PC Faaij. 2005. « Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term ». Biomass and Bioenergy 28 (4): 384-410.

Harman-Ware, Anne E., Cliff Foster, Renee M. Happs, Crissa Doeppke, Kristoffer Meunier, Jackson Gehan, Fengxia Yue, Fachuang Lu, et Mark F. Davis. 2016. « A thioacidolysis method tailored for higher-throughput quantitative analysis of lignin monomers ». Biotechnology Journal 11 (10): 1268-73

Hatfield, Ronald D., Hans-Joachim G. Jung, John Ralph, Dwayne R. Buxton, et Paul J. Weimer. 1994. « A Comparison of the Insoluble Residues Produced by the Klason Lignin and Acid Detergent Lignin Procedures ». Journal of the Science of Food and Agriculture 65 (1): 51-58.

He, Shaomei, Natalia Ivanova, Edward Kirton, Martin Allgaier, Claudia Bergin, Rudolf H. Scheffrahn, Nikos C. Kyrpides, Falk Warnecke, Susannah G. Tringe, et Philip Hugenholtz. 2013. « Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites ». PLOS ONE 8 (4): e61126.

Herbaut, Mickaël, Aya Zoghlami, Anouck Habrant, Xavier Falourd, Loïc Foucat, Brigitte Chabbert, et Gabriel Paës. 2018. « Multimodal analysis of pretreated biomass species highlights generic markers of lignocellulose recalcitrance ». Biotechnology for Biofuels 11 (février).

Hervé, Vincent, Pengfei Liu, Carsten Dietrich, David Sillam-Dussès, Petr Stiblik, Jan Šobotník, et Andreas Brune. 2020. « Phylogenomic Analysis of 589 Metagenome-Assembled Genomes Encompassing All Major Prokaryotic Lineages from the Gut of Higher Termites ». PeerJ 8 (février): e8614.

Heux, S., I. Meynial-Salles, M. J. O'Donohue, et C. Dumon. 2015. « White Biotechnology: State of the Art Strategies for the Development of Biocatalysts for Biorefining ». Biotechnology Advances 33 (8): 1653-70. Higuchi, Yudai, Shogo Aoki, Hiroki Takenami, Naofumi Kamimura, Kenji Takahashi, Shojiro Hishiyama, Christopher S. Lancefield, et al. 2018. « Bacterial Catabolism of β-Hydroxypropiovanillone and β-Hydroxypropiosyringone Produced in the Reductive Cleavage of Arylglycerol-β-Aryl Ether in Lignin ». Applied and Environmental Microbiology 84 (7).

Hilburg, Shayna L., Allison N. Elder, Hoyong Chung, Rachel L. Ferebee, Michael R. Bockstaller, et Newell R. Washburn. 2014. « A universal route towards thermoplastic lignin composites with improved mechanical properties ». Polymer 55 (4): 995-1003.

Hiras, Jennifer, Yu-Wei Wu, Stephanie A Eichorst, Blake A Simmons, et Steven W Singer. 2016. « Refining the phylum Chlorobi by resolving the phylogeny and metabolic potential of the representative of a deeply branching, uncultivated lineage ». The ISME Journal 10 (4): 833-45.

H. Isikgor, Furkan, et C. Remzi Becer. 2015. « Lignocellulosic Biomass: A Sustainable Platform for the Production of Bio-Based Chemicals and Polymers ». Polymer Chemistry 6 (25): 4497-4559

Ho, Margaret T., Michelle S. M. Li, Tim McDowell, Jacqueline MacDonald, et Ze-Chun Yuan. 2020. « Characterization and genomic analysis of a diesel-degrading bacterium, Acinetobacter calcoaceticus CA16, isolated from Canadian soil ». BMC Biotechnology 20 (1): 39.

Hofrichter, Martin, René Ullrich, Marek J. Pecyna, Christiane Liers, et Taina Lundell. 2010. « New and Classic Families of Secreted Fungal Heme Peroxidases ». Applied Microbiology and Biotechnology 87 (3): 871-97

Holm-Nielsen, Jens, et Ehiaze Augustine Ehimen. 2016. Biomass Supply Chains for Bioenergy and Biorefining. Woodhead Publishing.

Hongoh, Yuichi. 2010. « Diversity and Genomes of Uncultured Microbial Symbionts in the Termite Gut ». Bioscience, Biotechnology, and Biochemistry 74 (6): 1145-51.

———. 2011. « Toward the Functional Analysis of Uncultivable, Symbiotic Microorganisms in the Termite Gut ». Cellular and Molecular Life Sciences: CMLS 68 (8): 1311-25.

Hongoh, Yuichi, Moriya Ohkuma, et Toshiaki Kudo. 2003. « Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae) ». FEMS Microbiology Ecology 44 (2): 231-42.

Hopkins, D. W., J. A. Chudek, D. E. Bignell, J. Frouz, E. A. Webster, et T. Lawson. 1998. « Application of 13C NMR to Investigate the Transformations and Biodegradation of Organic Materials by Wood- and Soil-Feeding Termites, and a Coprophagous Litter-Dwelling Dipteran Larva ». Biodegradation 9 (6): 423-31.

Hu, Jiajun, Yiyun Xue, Hongcheng Guo, Min-Tian Gao, Jixiang Li, Shiping Zhang, et Yiu Fai Tsang. 2017. « Design and Composition of Synthetic Fungal-Bacterial Microbial Consortia That Improve Lignocellulolytic Enzyme Activity ». Bioresource Technology 227 (mars): 247-55.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. « EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses ». Nucleic Acids Research 47 (D1): D309-14.

Hug, Laura A., Brett J. Baker, Karthik Anantharaman, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, et al. 2016. « A New View of the Tree of Life ». Nature Microbiology 1 (5): 1-6.

Hunt, Martin, Astrid Gall, Swee Hoe Ong, Jacqui Brener, Bridget Ferns, Philip Goulder, Eleni Nastouli, Jacqueline A. Keane, Paul Kellam, et Thomas D. Otto. 2015. « IVA: Accurate de Novo Assembly of RNA Virus Genomes ». Bioinformatics (Oxford, England) 31 (14): 2374-76

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, et Loren J Hauser. 2010. « Prodigal: prokaryotic gene recognition and translation initiation site identification ». BMC Bioinformatics 11 (mars): 119.

Ioelovich, Michael. 2008. « CELLULOSE AS A NANOSTRUCTURED POLYMER: A SHORT REVIEW ». BioResources 3 (4): 1403-18.

J A Breznak, et and A. Brune. 1994. « Role of Microorganisms in the Digestion of Lignocellulose by Termites ». Annual Review of Entomology 39 (1): 453-87.

Janusz, Grzegorz, Anna Pawlik, Justyna Sulej, Urszula Świderska-Burek, Anna Jarosz-Wilkołazka, et Andrzej Paszczyński. 2017. « Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution ». FEMS Microbiology Reviews 41 (6): 941-62.

Johnson, Jethro S., Daniel J. Spakowicz, Bo-Young Hong, Lauren M. Petersen, Patrick Demkowicz, Lei Chen, Shana R. Leopold, et al. 2019. « Evaluation of 16S RRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis ». Nature Communications 10 (1): 5029.

Johnson, R. A., R. J. Thomas, T. G. Wood, et M. J. Swift. 1981. « Inoculation of the Fungus Comb in Newly Founded Colonies of Some Species of the Macrotermitinae (Isoptera) from Nigeria ». Journal of Natural History.

Jones, Ashley, Cynthia Torkel, David Stanley, Jamila Nasim, Justin Borevitz, et Benjamin Schwessinger. 2021. « High-Molecular Weight DNA Extraction, Clean-up and Size Selection for Long-Read Sequencing ». PLOS ONE 16 (7): e0253830.

Jong, Ed de, et Gerfried Jungmeier. 2015. « Chapter 1 - Biorefinery Concepts in Comparison to Petrochemical Refineries ». In Industrial Biorefineries & White Biotechnology, édité par Ashok

Pandey, Rainer Höfer, Mohammad Taherzadeh, K. Madhavan Nampoothiri, et Christian Larroche, 3-33. Amsterdam: Elsevier.

Kakirde, Kavita S., Larissa C. Parsley, et Mark R. Liles. 2010. « Size Does Matter: Application-driven Approaches for Soil Metagenomics ». Soil biology & biochemistry 42 (11): 1911-23.

Kamimura, Naofumi, Shingo Sakamoto, Nobutaka Mitsuda, Eiji Masai, et Shinya Kajita. 2019. « Advances in Microbial Lignin Degradation and Its Applications ». Current Opinion in Biotechnology, Food Biotechnology • Plant Biotechnology, 56 (avril): 179-86.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, et Mao Tanabe. 2016. « KEGG as a Reference Resource for Gene and Protein Annotation ». Nucleic Acids Research 44 (D1): D457-62.

Kang, Dingrong, Samuel Jacquiod, Jakob Herschend, Shaodong Wei, Joseph Nesme, et Søren J. Sørensen. 2020. « Construction of Simplified Microbial Consortia to Degrade Recalcitrant Materials Based on Enrichment and Dilution-to-Extinction Cultures ». Frontiers in Microbiology 10: 3010.

Kang, Dongwan D., Jeff Froula, Rob Egan, et Zhong Wang. 2015. « MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities ». PeerJ 3: e1165.

Kang, Dongwan, Feng Li, Edward S. Kirton, Ashleigh Thomas, Rob S. Egan, Hong An, et Zhong Wang. 2019. « MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies ». e27522v1. PeerJ Inc.

Karlen, Steven D., Heather C. A. Free, Dharshana Padmakshan, Bronwen G. Smith, John Ralph, et Philip J. Harris. 2018. « Commelinid Monocotyledon Lignins Are Acylated by P-Coumarate ». Plant Physiology 177 (2): 513-21.

Karlen, Steven D., Chengcheng Zhang, Matthew L. Peck, Rebecca A. Smith, Dharshana Padmakshan, Kate E. Helmich, Heather C. A. Free, et al. 2016. « Monolignol ferulate conjugates are naturally incorporated into plant lignins ». Science Advances 2 (10).

Katahira, Rui, Thomas J. Elder, et Gregg T. Beckham. 2018. « Chapter 1 A Brief Introduction to Lignin Structure », 1-20

Kato, Kinya, Shinya Kozaki, et Masanori Sakuranaga. 1998. « Degradation of Lignin Compounds by Bacteria from Termite Guts ». Biotechnology Letters 20 (5): 459-62.

Katsumata, Kyoko S., Zhenfu Jin, Keko Hori, et Kenji Iiyama. 2007. « Structural Changes in Lignin of Tropical Woods during Digestion by Termite, Cryptotermes Brevis ». Journal of Wood Science 53 (5): 419.

Ke, Jing, Dhrubojyoti D. Laskar, et Shulin Chen. 2013. « Varied lignin disruption mechanisms for different biomass substrates in lower termite ». Renewable Energy 50 (février): 1060-64.

Ke, Jing, Dhrubojyoti D. Laskar, Deepak Singh, et Shulin Chen. 2011. « In Situ Lignocellulosic Unlocking Mechanism for Carbohydrate Hydrolysis in Termites: Crucial Lignin Modification ». Biotechnology for Biofuels 4 (juin): 17

Ke, Jing, Deepak Singh, et Shulin Chen. 2011. « Aromatic Compound Degradation by the Wood-Feeding Termite Coptotermes Formosanus (Shiraki) ». International Biodeterioration & Biodegradation 65 (6): 744-56.

Kim, Hoon, et John Ralph. 2010. « Solution-State 2D NMR of Ball-Milled Plant Cell Wall Gels in DMSO-D6/Pyridine-D5 ». Organic & Biomolecular Chemistry 8 (3): 576-91.

Kim, Minseok, Mark Morrison, et Zhongtang Yu. 2011. « Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes ». Journal of Microbiological Methods 84 (1): 81-87.

Kim, S. J., et M. Shoda. 1999. « Decolorization of Molasses and a Dye by a Newly Isolated Strain of the Fungus Geotrichum Candidum Dec 1 ». Biotechnology and Bioengineering 62 (1): 114-19.

Knop, Doriv, Dana Levinson, Arik Makovitzki, Avi Agami, Elad Lerer, Avishai Mimran, Oded Yarden, et Yitzhak Hadar. 2016. « Limits of Versatility of Versatile Peroxidase ». Applied and Environmental Microbiology 82 (14): 4070-80.

Kobayashi, Hirokazu, et Atsushi Fukuoka. 2013. « Synthesis and Utilisation of Sugar Compounds Derived from Lignocellulosic Biomass ». Green Chemistry 15 (7): 1740-63.

König, Helmut, Li Li, et Jürgen Fröhlich. 2013. « The Cellulolytic System of the Termite Gut ». Applied Microbiology and Biotechnology 97 (18): 7943-62.

Kontur, Wayne S., Charles N. Olmsted, Larissa M. Yusko, Alyssa V. Niles, Kevin A. Walters, Emily T. Beebe, Kirk A. Vander Meulen, et al. 2019. « A Heterodimeric Glutathione S-Transferase That Stereospecifically Breaks Lignin's β(R)-Aryl Ether Bond Reveals the Diversity of Bacterial β-Etherases ». Journal of Biological Chemistry 294 (6): 1877-90.

Korostin, Dmitriy, Nikolay Kulemin, Vladimir Naumov, Vera Belova, Dmitriy Kwon, et Alexey Gorbachev. 2020. « Comparative Analysis of Novel MGISEQ-2000 Sequencing Platform vs Illumina HiSeq 2500 for Whole-Genome Sequencing ». PLOS ONE 15 (3): e0230301.

Kozlov, Alexey M., Diego Darriba, Tomáš Flouri, Benoit Morel, et Alexandros Stamatakis. 2019. « RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference ». Bioinformatics 35 (21): 4453-55.

« Kraken: ultrafast metagenomic sequence classification using exact alignments | Genome Biology | Full Text ». s. d. Consulté le 9 mai 2019. https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r46.

Kuijk, Sandra J. A. van, Anton S. M. Sonnenberg, Johan J. P. Baars, Wouter H. Hendriks, José C. del Río, Jorge Rencoret, Ana Gutiérrez, Norbert C. A. de Ruijter, et John W. Cone. 2017. « Chemical Changes and Increased Degradability of Wheat Straw and Oak Wood Chips Treated with the White Rot Fungi Ceriporiopsis Subvermispora and Lentinula Edodes ». Biomass and Bioenergy 105 (octobre): 381-91.

Kurtz, Zachary D., Christian L. Müller, Emily R. Miraldi, Dan R. Littman, Martin J. Blaser, et Richard A. Bonneau. 2015. « Sparse and Compositionally Robust Inference of Microbial Ecological Networks ». PLOS Computational Biology 11 (5): e1004226.

Lahens, Nicholas F., Emanuela Ricciotti, Olga Smirnova, Erik Toorens, Eun Ji Kim, Giacomo Baruzzo, Katharina E. Hayer, Tapan Ganguly, Jonathan Schug, et Gregory R. Grant. 2017. « A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression ». BMC Genomics 18 (1): 602.

Lam, Felix H., Adel Ghaderi, Gerald R. Fink, et Gregory Stephanopoulos. 2014. « Engineering alcohol tolerance in yeast ». Science (New York, N.Y.) 346 (6205): 71-75.

Lam, Kathy N., Jiujun Cheng, Katja Engel, Josh D. Neufeld, et Trevor C. Charles. 2015. « Current and Future Resources for Functional Metagenomics ». Frontiers in Microbiology 6.

Lam, Pui Ying, Yuki Tobimatsu, Naoyuki Matsumoto, Shiro Suzuki, Wu Lan, Yuri Takeda, Masaomi Yamamura, et al. 2019. « OsCAldOMT1 Is a Bifunctional O -Methyltransferase Involved in the Biosynthesis of Tricin-Lignins in Rice Cell Walls ». Scientific Reports 9 (1): 1-13.

Lam, Pui Ying, Yuki Tobimatsu, Yuri Takeda, Shiro Suzuki, Masaomi Yamamura, Toshiaki Umezawa, et Clive Lo. 2017. « Disrupting Flavone Synthase II Alters Lignin and Improves Biomass Digestibility1[OPEN] ». Plant Physiology 174 (2): 972-85.

Lambertz, Camilla, Megan Garvey, Johannes Klinger, Dirk Heesel, Holger Klose, Rainer Fischer, et Ulrich Commandeur. 2014. « Challenges and advances in the heterologous expression of cellulolytic enzymes: a review ». Biotechnology for Biofuels 7 (octobre).

Lan, Jing, Xirong Huang, Ming Hu, Yuezhong Li, Yinbo Qu, Peiji Gao, et Daocheng Wu. 2006. « High Efficient Degradation of Dyes with Lignin Peroxidase Coupled with Glucose Oxidase ». Journal of Biotechnology 123 (4): 483-90.

Lan, Wu, Fachuang Lu, Matthew Regner, Yimin Zhu, Jorge Rencoret, Sally A. Ralph, Uzma I. Zakai, Kris Morreel, Wout Boerjan, et John Ralph. 2015. « Tricin, a Flavonoid Monomer in Monocot Lignification1[OPEN] ». Plant Physiology 167 (4): 1284-95.

Lan, Yemin, Gail Rosen, et Ruth Hershberg. 2016. « Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains ». Microbiome 4 (mai).

Lane, D J, B Pace, G J Olsen, D A Stahl, M L Sogin, et N R Pace. 1985. « Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. » Proceedings of the National Academy of Sciences of the United States of America 82 (20): 6955-59.

Lapébie, Pascal, Vincent Lombard, Elodie Drula, Nicolas Terrapon, et Bernard Henrissat. 2019. « Bacteroidetes Use Thousands of Enzyme Combinations to Break down Glycans ». Nature Communications 10 (1): 2043.

Lapidus, Alla L., et Anton I. Korobeynikov. 2021. « Metagenomic Data Assembly – The Way of Decoding Unknown Microorganisms ». Frontiers in Microbiology 12: 653.

Lapierre, Catherine, Bernard Monties, Christian Rolando, et Laboratorie de Chirale. 1985. « Thioacidolysis of Lignin: Comparison with Acidolysis ». Journal of Wood Chemistry and Technology 5 (2): 277-92.

Lapierre, Catherine, Brigitte Pollet, et Christian Rolando. 1995. « New Insights into the Molecular Architecture of Hardwood Lignins by Chemical Degradative Methods ». Research on Chemical Intermediates 21 (3-5): 397.

Laserson, Jonathan, Vladimir Jojic, et Daphne Koller. 2011. « Genovo: De Novo Assembly for Metagenomes ». Journal of Computational Biology: A Journal of Computational Molecular Cell Biology 18 (3): 429-43.

Latorre-Pérez, Adriel, Pascual Villalba-Bermell, Javier Pascual, et Cristina Vilanova. 2020. « Assembly Methods for Nanopore-Based Metagenomic Sequencing: A Comparative Study ». Scientific Reports 10 (1): 13588.

Lazuka, Adèle, Lucas Auer, Sophie Bozonnet, Diego P. Morgavi, Michael O'Donohue, et Guillermina Hernandez-Raquet. 2015. « Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium ». Bioresource Technology 196 (novembre): 241-49.

Lazuka, Adèle, Lucas Auer, Michael O'Donohue, et Guillermina Hernandez-Raquet. 2018. « Anaerobic lignocellulolytic microbial consortium derived from termite gut: enrichment, lignocellulose degradation and community dynamics | Biotechnology for Biofuels | Full Text ». Biotechnology for Biofuels 11 (284).

Lazuka, Adèle, Cécile Roland, Abdellatif Barakat, Fabienne Guillon, Michael O'Donohue, et Guillermina Hernandez-Raquet. 2017. « Ecofriendly Lignocellulose Pretreatment to Enhance the Carboxylate Production of a Rumen-Derived Microbial Consortium ». Bioresource Technology 236 (juillet): 225-33.

Le Roes-Hill, Marilize, Jeffrey Rohland, et Stephanie Burton. 2011. « Actinobacteria Isolated from Termite Guts as a Source of Novel Oxidative Enzymes ». Antonie Van Leeuwenhoek 100 (4): 589-605.

Lee, Hojun, Ki-Wook Lee, Taeseob Lee, Donghyun Park, Jongsuk Chung, Chung Lee, Woong-Yang Park, et Dae-Soon Son. 2018. « Performance Evaluation Method for Read Mapping Tool in Clinical Panel Sequencing ». Genes & Genomics 40 (2): 189-97.

Lee, Siseon, Minsik Kang, Jung-Hoon Bae, Jung-Hoon Sohn, et Bong Hyun Sung. 2019. « Bacterial Valorization of Lignin: Strains, Enzymes, Conversion Pathways, Biosensors, and Perspectives ». Frontiers in Bioengineering and Biotechnology 7.

Letunic, Ivica, et Peer Bork. 2019. « Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments ». Nucleic Acids Research 47 (W1): W256-59.

Levasseur, Anthony, Elodie Drula, Vincent Lombard, Pedro M. Coutinho, et Bernard Henrissat. 2013. « Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes ». Biotechnology for Biofuels 6 (1): 41.

Li, Dinghua. (2014) 2021. MEGAHIT. C++. https://github.com/voutcn/megahit/blob/7dde1cae4dfa8ced0a9bd524894df36e9cd4185b/CHANGELOG.md.

Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, et Tak-Wah Lam. 2015. « MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph ». Bioinformatics (Oxford, England) 31 (10): 1674-76.

Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko Sadakane, Hiroshi Yamashita, et Tak-Wah Lam. 2016. « MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices ». Methods (San Diego, Calif.) 102: 3-11.

Li, Fei, Fuying Ma, Honglu Zhao, Shu Zhang, Lei Wang, Xiaoyu Zhang, et Hongbo Yu. 2019. « A Lytic Polysaccharide Monooxygenase from a White-Rot Fungus Drives the Degradation of Lignin by a Versatile Peroxidase ». Applied and Environmental Microbiology 85 (9): e02803-18.

Li, Heng. 2013. « Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM ». arXiv:1303.3997 [q-bio], mai. http://arxiv.org/abs/1303.3997.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, et Richard Durbin. 2009. « The Sequence Alignment/Map format and SAMtools ». Bioinformatics 25 (16): 2078-79.

Li, Hongjie, Daniel J. Yelle, Chang Li, Mengyi Yang, Jing Ke, Ruijuan Zhang, Yu Liu, et al. 2017. « Lignocellulose Pretreatment in a Fungus-Cultivating Termite ». Proceedings of the National Academy of Sciences 114 (18): 4709-14.

Lima Brossi, Maria Julia de, Diego Javier Jiménez, Larisa Cortes-Tolalpa, et Jan Dirk van Elsas. 2016. « Soil-Derived Microbial Consortia Enriched with Different Plant Biomass Reveal Distinct Players Acting in Lignocellulose Degradation ». Microbial Ecology 71: 616-27.

Lin, Hsin-Hung, et Yu-Chieh Liao. 2016. « Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes ». Scientific Reports 6 (avril).

Lin, You-Yu, Chia-Hung Hsieh, Jiun-Hong Chen, Xuemei Lu, Jia-Horng Kao, Pei-Jer Chen, Ding-Shinn Chen, et Hurng-Yi Wang. 2017. « De novo assembly of highly polymorphic metagenomic data using in situ generated reference sequences and a novel BLAST-based assembly pipeline ». BMC Bioinformatics 18 (1): 223.

Liu, Qingquan, Le Luo, et Luqing Zheng. 2018. « Lignins: Biosynthesis and Biological Functions in Plants ». International Journal of Molecular Sciences 19 (2).

Liu, Xiaoying, Steven M. Zicari, Guangqing Liu, Yeqing Li, et Ruihong Zhang. 2015. « Pretreatment of Wheat Straw with Potassium Hydroxide for Increasing Enzymatic and Microbial Degradability ». Bioresource Technology 185 (juin): 150-57.

Lončar, Nikola, Dana I. Colpa, et Marco W. Fraaije. 2016. « Exploring the Biocatalytic Potential of a DyP-Type Peroxidase by Profiling the Substrate Acceptance of Thermobifida Fusca DyP Peroxidase ». Tetrahedron, Modern Developments in Biotransformations, 72 (46): 7276-81

Lozupone, Catherine, et Rob Knight. 2005. « UniFrac: A New Phylogenetic Method for Comparing Microbial Communities ». Applied and Environmental Microbiology 71 (12): 8228-35

Lui, Lauren M., Torben N. Nielsen, et Adam P. Arkin. 2021. « A Method for Achieving Complete Microbial Genomes and Improving Bins from Metagenomics Data ». PLOS Computational Biology 17 (5): e1008972.

Macrelli, Stefano, Johan Mogensen, et Guido Zacchi. 2012. « Techno-economic evaluation of 2nd generation bioethanol production from sugar cane bagasse and leaves integrated with the sugar-based ethanol process ». Biotechnology for Biofuels 5 (avril): 22.

Maeda, Allyn H., Shinro Nishi, Shun'ichi Ishii, Yasuhiro Shimane, Hideki Kobayashi, Junko Ichikawa, Kanako Kurosawa, Wataru Arai, Hideto Takami, et Yukari Ohta. 2018. « Complete Genome Sequence of Altererythrobacter sp. Strain B11, an Aromatic Monomer-Degrading Bacterium, Isolated from Deep-Sea Sediment under the Seabed off Kashima, Japan ». Genome Announcements 6 (12): e00200-18.

Magoč, Tanja, et Steven L. Salzberg. 2011. « FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies ». Bioinformatics (Oxford, England) 27 (21): 2957-63.

Mahé, Frédéric, Torbjørn Rognes, Christopher Quince, Colomban de Vargas, et Micah Dunthorn. 2014. « Swarm: Robust and Fast Clustering Method for Amplicon-Based Studies ». PeerJ 2 (septembre): e593.

Mahé, Frédéric, Torbjørn Rognes, Christopher Quince, Colomban de Vargas, et Micah Dunthorn. 2015. « Swarm v2: highly-scalable and high-resolution amplicon clustering ». PeerJ 3 (décembre).

Majira, Amel, Blandine Godon, Laurence Foulon, Jacinta C. van der Putten, Laurent Cézard, Marina Thierry, Florian Pion, et al. 2019. « Enhancing the Antioxidant Activity of Technical Lignins by Combining Solvent Fractionation and Ionic-Liquid Treatment ». ChemSusChem 12 (21): 4799-4809.

Majumdar, Sudipta, Tiit Lukk, Jose O. Solbiati, Stefan Bauer, Satish K. Nair, John E. Cronan, et John A. Gerlt. 2014. « Roles of Small Laccases from Streptomyces in Lignin Degradation ». Biochemistry 53 (24): 4047-58

Malherbe, S., et T. E. Cloete. 2002. « Lignocellulose Biodegradation: Fundamentals and Applications ». Reviews in Environmental Science and Biotechnology 1 (2): 105-14.

Mallawaarachchi, Vijini G., Anuradha S. Wickramarachchi, et Yu Lin. 2021. « Improving metagenomic binning results with overlapped bins using assembly graphs ». Algorithms for Molecular Biology 16 (1): 3.

Mansfield, Shawn D., Hoon Kim, Fachuang Lu, et John Ralph. 2012. « Whole Plant Cell Wall Characterization Using Solution-State 2D NMR ». Nature Protocols 7 (9): 1579-89.

Martin, Andri Fadillah, Yuki Tobimatsu, Ryosuke Kusumi, Naoyuki Matsumoto, Takuji Miyamoto, Pui Ying Lam, Masaomi Yamamura, Taichi Koshiba, Masahiro Sakamoto, et Toshiaki Umezawa. 2019. « Altered Lignocellulose Chemical Structure and Molecular Assembly in CINNAMYL ALCOHOL DEHYDROGENASE -Deficient Rice ». Scientific Reports 9 (1): 17153.

Martin, Marcel. 2011. « CUTADAPT removes adapter sequences from high-throughput sequencing reads ». EMBnet journal 17 (1).

Martínez-Porchas, Marcel, Enrique Villalpando-Canchola, et Francisco Vargas-Albores. 2016. « Significant loss of sensitivity and specificity in the taxonomic classification occurs when short 16S rRNA gene sequences are used ». Heliyon 2 (9): e00170.

Maryana, Roni, Dian Ma'rifatun, A. I. Wheni, K. W. Satriyo, et W. Angga Rizal. 2014. « Alkaline Pretreatment on Sugarcane Bagasse for Bioethanol Production ». Energy Procedia, Conference and Exhibition Indonesia Renewable Energy & Energy Conservation [Indonesia EBTKE-CONEX 2013], 47 (janvier): 250-54.

Marynowska, Martyna, Xavier Goux, David Sillam-Dussès, Corinne Rouland-Lefèvre, Rashi Halder, Paul Wilmes, Piotr Gawron, Yves Roisin, Philippe Delfosse, et Magdalena Calusinska. 2020. « Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites ». Microbiome 8 (1): 96.

Marynowska, Martyna, Xavier Goux, David Sillam-Dussès, Corinne Rouland-Lefèvre, Yves Roisin, Philippe Delfosse, et Magdalena Calusinska. 2017. « Optimization of a metatranscriptomic approach to

study the lignocellulolytic potential of the higher termite gut microbiome ». BMC Genomics 18 (septembre).

Marzullo, L., R. Cannio, P. Giardina, M. T. Santini, et G. Sannia. 1995. « Veratryl Alcohol Oxidase from Pleurotus Ostreatus Participates in Lignin Biodegradation and Prevents Polymerization of Laccase-Oxidized Substrates ». The Journal of Biological Chemistry 270 (8): 3823-27.

Masai, E., Y. Katayama, S. Nishikawa, M. Yamasaki, N. Morohoshi, et T. Haraguchi. 1989. « Detection and Localization of a New Enzyme Catalyzing the Beta-Aryl Ether Cleavage in the Soil Bacterium (Pseudomonas Paucimobilis SYK-6) ». FEBS Letters 249 (2): 348-52.

Masai, Eiji, Yoshihiro Katayama, et Masao Fukuda. 2007. « Genetic and Biochemical Investigations on Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds ». Bioscience, Biotechnology, and Biochemistry 71 (1): 1-15.

Matsushika, Akinori, Hiroyuki Inoue, Tsutomu Kodaki, et Shigeki Sawayama. 2009. « Ethanol Production from Xylose in Engineered Saccharomyces Cerevisiae Strains: Current State and Perspectives ». Applied Microbiology and Biotechnology 84 (1): 37-53.

McMurdie, Paul J., et Susan Holmes. 2013a. « Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data ». PLOS ONE 8 (4): e61217.

———. 2013b. « Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data ». PLOS ONE 8 (4): e61217.

Méchin, Valérie, Aurélia Laluc, Frédéric Legée, Laurent Cézard, Dominique Denoue, Yves Barrière, et Catherine Lapierre. 2014. « Impact of the Brown-Midrib Bm5 Mutation on Maize Lignins ». Journal of Agricultural and Food Chemistry 62 (22): 5102-7.

Meijenfeldt, F. A. Bastiaan von, Ksenia Arkhipova, Diego D. Cambuy, Felipe H. Coutinho, et Bas E. Dutilh. 2019. « Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT ». Genome Biology 20 (1): 217. Menzel, Peter, Kim Lee Ng, et Anders Krogh. 2016. « Fast and Sensitive Taxonomic Classification for Metagenomics with Kaiju ». Nature Communications 7 (avril): 11257.

Meyer, Fernando, Peter Hofmann, Peter Belmann, Ruben Garrido-Oter, Adrian Fritz, Alexander Sczyrba, et Alice C. McHardy. 2018. « AMBER: Assessment of Metagenome BinnERs ». GigaScience 7 (6).

Meyer, Fernando, Till-Robin Lesker, David Koslicki, Adrian Fritz, Alexey Gurevich, Aaron E. Darling, Alexander Sczyrba, Andreas Bremges, et Alice C. McHardy. 2021. « Tutorial: Assessing Metagenomics Software with the CAMI Benchmarking Toolkit ». Nature Protocols 16 (4): 1785-1801.

Meziti, Alexandra, Luis M. Rodriguez-R, Janet K. Hatt, Angela Peña-Gonzalez, Karen Levy, et Konstantinos T. Konstantinidis. s. d. « The Reliability of Metagenome-Assembled Genomes (MAGs)

in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample ». Applied and Environmental Microbiology 87 (6): e02593-20.

Mikaelyan, Aram, Carsten Dietrich, Tim Köhler, Michael Poulsen, David Sillam-Dussès, et Andreas Brune. 2015. « Diet Is the Primary Determinant of Bacterial Community Structure in the Guts of Higher Termites ». Molecular Ecology 24 (20): 5284-95.

Mikaelyan, Aram, Tim Köhler, Niclas Lampert, Jeffrey Rohland, Hamadi Boga, Katja Meuser, et Andreas Brune. 2015. « Classifying the Bacterial Gut Microbiota of Termites and Cockroaches: A Curated Phylogenetic Reference Database (DictDb) ». Systematic and Applied Microbiology 38 (7): 472-82.

Min, D. Y., H. Jameel, H. M. Chang, L. Lucia, Z. G. Wang, et Y. C. Jin. 2014. « The Structural Changes of Lignin and Lignin–Carbohydrate Complexes in Corn Stover Induced by Mild Sodium Hydroxide Treatment ». RSC Advances 4 (21): 10845-50.

Mitchell, R. B., M. R. Schmer, W. F. Anderson, V. Jin, K. S. Balkcom, J. Kiniry, A. Coffin, et P. White. 2016. « Dedicated Energy Crops and Crop Residues for Bioenergy Feedstocks in the Central and Eastern USA ». BioEnergy Research 9 (2): 384-98.

Miyamoto, Takuji, Asako Mihashi, Masaomi Yamamura, Yuki Tobimatsu, Shiro Suzuki, Rie Takada, Yoshinori Kobayashi, et Toshiaki Umezawa. 2018. « Comparative Analysis of Lignin Chemical Structures of Sugarcane Bagasse Pretreated by Alkaline, Hydrothermal, and Dilute Sulfuric Acid Methods ». Industrial Crops and Products 121 (octobre): 124-31.

Monlau, F., A. Barakat, J. P. Steyer, et H. Carrere. 2012. « Comparison of Seven Types of Thermo-Chemical Pretreatments on the Structural Features and Anaerobic Digestion of Sunflower Stalks ». Bioresource Technology 120 (septembre): 241-47.

Montella, Salvatore, Valeria Ventorino, Vincent Lombard, Bernard Henrissat, Olimpia Pepe, et Vincenza Faraco. 2017. « Discovery of Genes Coding for Carbohydrate-Active Enzyme by Metagenomic Analysis of Lignocellulosic Biomasses ». Scientific Reports 7 (1): 42623.

Moraes, Eduardo C., Thabata M. Alvarez, Gabriela F. Persinoti, Geizecler Tomazetto, Livia B. Brenelli, Douglas A. A. Paixão, Gabriela C. Ematsu, et al. 2018. « Lignolytic-consortium omics analyses reveal novel genomes and pathways involved in lignin modification and valorization ». Biotechnology for Biofuels 11 (1): 75

Morozova, O. V., G. P. Shumakovich, S. V. Shleev, et Ya I. Yaropolov. 2007. « Laccase-Mediator Systems and Their Applications: A Review ». Applied Biochemistry and Microbiology 43 (5): 523-35.

Moshkelani, Maryam, Mariya Marinova, Michel Perrier, et Jean Paris. 2013. « The forest biorefinery and its implementation in the pulp and paper industry: Energy overview ». Applied Thermal Engineering, Combined Special Issues: ECP 2011 and IMPRES 2010, 50 (2): 1427-36

Mysara, Mohamed, Yvan Saeys, Natalie Leys, Jeroen Raes, et Pieter Monsieurs. 2015. « CATCh, an Ensemble Classifier for Chimera Detection in 16S rRNA Sequencing Studies ». Applied and Environmental Microbiology 81 (5): 1573-84.

Nakashima, K., H. Watanabe, H. Saitoh, G. Tokuda, et J.-I. Azuma. 2002. « Dual Cellulose-Digesting System of the Wood-Feeding Termite, Coptotermes Formosanus Shiraki ». Insect Biochemistry and Molecular Biology 32 (7): 777-84.

Namiki, Toshiaki, Tsuyoshi Hachiya, Hideaki Tanaka, et Yasubumi Sakakibara. 2012. « MetaVelvet: An Extension of Velvet Assembler to de Novo Metagenome Assembly from Short Sequence Reads ». Nucleic Acids Research 40 (20): e155.

Naron, D. R., F. -X. Collard, L. Tyhoda, et J. F. Görgens. 2017. « Characterisation of lignins from different sources by appropriate analytical methods: Introducing thermogravimetric analysis-thermal desorption-gas chromatography–mass spectroscopy ». Industrial Crops and Products 101 (juillet): 61-74.

Nishiyama, Yoshiharu, Paul Langan, et Henri Chanzy. 2002. « Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction ». Journal of the American Chemical Society 124 (31): 9074-82.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, et Pavel A. Pevzner. 2017. « MetaSPAdes: A New Versatile Metagenomic Assembler ». Genome Research 27 (5): 824-34.

O'Connor, Roberta M., Jennifer M. Fung, Koty H. Sharp, Jack S. Benner, Colleen McClung, Shelley Cushing, Elizabeth R. Lamkin, et al. 2014. « Gill Bacteria Enable a Novel Digestive Strategy in a Wood-Feeding Mollusk ». Proceedings of the National Academy of Sciences 111 (47): E5096-5104.

Okello, Collins, Stefania Pindozzi, Salvatore Faugno, et Lorenzo Boccia. 2013. « Bioenergy potential of agricultural and forest residues in Uganda ». Biomass and Bioenergy 56 (septembre): 515-25.

Patel, Ravi K., et Mukesh Jain. 2012. « NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data ». PLoS ONE 7 (2): e30619.

Pawar, Sandip V., Steven J. Hallam, et Vikramaditya G. Yadav. 2018. « Metagenomic Discovery of a Novel Transaminase for Valorization of Monoaromatic Compounds ». RSC Advances 8 (40): 22490-97.

Pereira-Marques, Joana, Anne Hout, Rui M. Ferreira, Michiel Weber, Ines Pinto-Ribeiro, Leen-Jan van Doorn, Cornelis Willem Knetsch, et Ceu Figueiredo. 2019. « Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis ». Frontiers in Microbiology 10: 1277.

Perez, J., et T. W. Jeffries. 1993. « Role of Organic Acid Chelators in Manganese Regulation of Lignin Degradation By<Emphasis Type="Italic">Phanerochaete Chrysosporium</Emphasis> ». Applied Biochemistry and Biotechnology 39-40 (1): 227.

Peschel, Stefanie, Christian L Müller, Erika von Mutius, Anne-Laure Boulesteix, et Martin Depner. 2020. « NetCoMi: network construction and

Peterson, Brittany F., et Michael E. Scharf. 2016. « Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions ». BMC Genomics 17 (1): 772.

Ployet, Raphaël. 2017. « Régulation de La Formation Du Bois Chez l'eucalyptus Lors Du Développement et En Réponse à Des Contraintes Environnementales ». Phd, Université de Toulouse, Université Toulouse III - Paul Sabatier.

Poirier, Simon, Olivier Rué, Raphaëlle Peguilhan, Gwendoline Coeuret, Monique Zagorec, Marie-Christine Champomier-Vergès, Valentin Loux, et Stéphane Chaillou. 2018. « Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing ». PLoS ONE 13 (9).

Polansky, Ondrej, Zuzana Sekelova, Marcela Faldynova, Alena Sebkova, Frantisek Sisak, et Ivan Rychlik. 2016. « Important Metabolic Pathways and Biological Processes Expressed by Chicken Cecal Microbiota ». Applied and Environmental Microbiology 82 (5): 1569-76.

Polat, Yusuf, Elena Stojanovska, Tolera A. Negawo, Elmas Doner, et Ali Kilic. 2017. « Lignin as an Additive for Advanced Composites ». In Green Biocomposites, 71-89. Green Energy and Technology. Springer, Cham

Pollard, Martin O., Deepti Gurdasani, Alexander J. Mentzer, Tarryn Porter, et Manjinder S. Sandhu. 2018. « Long Reads: Their Purpose and Place ». Human Molecular Genetics 27 (R2): R234-41.

Pollegioni Loredano, Tonin Fabio, et Rosini Elena. 2015. « Lignin-degrading enzymes ». The FEBS Journal 282 (7): 1190-1213.

Prabhakaran, Madhu, Matthew B. Couger, Colin A. Jackson, Tyler Weirick, et Babu Z. Fathepure. 2015. « Genome Sequences of the Lignin-Degrading Pseudomonas Sp. Strain YS-1p and Rhizobium Sp. Strain YS-1r Isolated from Decaying Wood ». Genome Announcements 3 (2).

Puentes-Téllez, Pilar Eliana, et Joana Falcao Salles. 2018. « Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation ». Microbial Ecology 76 (2): 419-29

Qian, Yong, Xueqing Qiu, et Shiping Zhu. 2015. « Lignin: A Nature-Inspired Sun Blocker for Broad-Spectrum Sunscreens ». Green Chemistry 17 (1): 320-24.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, et Frank Oliver Glöckner. 2013. « The SILVA ribosomal RNA gene database project: improved data processing and web-based tools ». Nucleic Acids Research 41 (D1): D590-96

Quince, Christopher, Alan W. Walker, Jared T. Simpson, Nicholas J. Loman, et Nicola Segata. 2017. « Shotgun Metagenomics, from Sampling to Analysis ». Nature Biotechnology 35 (9): 833-44.

Qureshi, N. 2017. « Solvent (Acetone–Butanol: AB) Production☆ ». In Reference Module in Life Sciences. Elsevier.

Ragauskas, Arthur J., Gregg T. Beckham, Mary J. Biddy, Richard Chandra, Fang Chen, Mark F. Davis, Brian H. Davison, et al. 2014. « Lignin Valorization: Improving Lignin Processing in the Biorefinery ». Science 344 (6185): 1246843. https://doi.org/10.1126/science.1246843.

Rahmanpour, Rahman, et Timothy D. H. Bugg. 2015. « Characterisation of Dyp-Type Peroxidases from Pseudomonas Fluorescens Pf-5: Oxidation of Mn(II) and Polymeric Lignin by Dyp1B ». Archives of Biochemistry and Biophysics, A protein fold with multiple functions: chlorite dismutase, HemQ and DyP-type peroxidase, 574 (mai): 93-98.

Rahmanpour, Rahman, Dean Rea, Shirin Jamshidi, Vilmos Fülöp, et Timothy D. H. Bugg. 2016. « Structure of Thermobifida Fusca DyP-Type Peroxidase and Activity towards Kraft Lignin and Lignin Model Compounds ». Archives of Biochemistry and Biophysics 594 (mars): 54-60.

Rai, Amrita, Johann P. Klare, Patrick Y. A. Reinke, Felix Englmaier, Jörg Fohrer, Roman Fedorov, Manuel H. Taft, et al. 2021. « Structural and Biochemical Characterization of a Dye-Decolorizing Peroxidase from Dictyostelium Discoideum ». International Journal of Molecular Sciences 22 (12): 6265

Raja, HuzefaA., Andrew N. Miller, Cedric J. Pearce, et Nicholas H. Oberlies. 2017. « Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community ». Journal of Natural Products 80 (3): 756-70.

Ralph, John, Catherine Lapierre, et Wout Boerjan. 2019. « Lignin Structure and Its Engineering ». Current Opinion in Biotechnology, Food Biotechnology • Plant Biotechnology, 56 (avril): 240-49.

Ralph, John, Knut Lundquist, Gösta Brunow, Fachuang Lu, Hoon Kim, Paul F. Schatz, Jane M. Marita, et al. 2004. « Lignins: Natural Polymers from Oxidative Coupling of 4-Hydroxyphenyl- Propanoids ». Phytochemistry Reviews 3 (1-2): 29-60.

Ramachandra, M, D L Crawford, et G Hertel. 1988. « Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete Streptomyces viridosporus. » Applied and Environmental Microbiology 54 (12): 3057-63.

Ravi, Krithika, Javier García-Hidalgo, Marie F Gorwa-Grauslund, et Gunnar Lidén. 2017. « Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost ». Applied Microbiology and Biotechnology 101 (12): 5059-70.

Rezende, Camila Alves, Marisa Aparecida de Lima, Priscila Maziero, Eduardo Ribeiro deAzevedo, Wanius Garcia, et Igor Polikarpov. 2011. « Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility ». Biotechnology for Biofuels 4 (novembre): 54.

Riley, Robert, Asaf A. Salamov, Daren W. Brown, Laszlo G. Nagy, Dimitrios Floudas, Benjamin W. Held, Anthony Levasseur, et al. 2014. « Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi ». Proceedings of the National Academy of Sciences of the United States of America 111 (27): 9923-28.

Río, José C. del, Jorge Rencoret, Pepijn Prinsen, Ángel T. Martínez, John Ralph, et Ana Gutiérrez. 2012. « Structural Characterization of Wheat Straw Lignin as Revealed by Analytical Pyrolysis, 2D-NMR, and Reductive Cleavage Methods ». Journal of Agricultural and Food Chemistry 60 (23): 5922-35.

Riyadi, Fatimah Azizah, Analhuda Abdullah Tahir, Nurtasbiyah Yusof, Nurul Syazwani Ahmad Sabri, Megat Johari Megat Mohd Noor, Fazrena Nadia M. D. Akhir, Nor'azizi Othman, Zuriati Zakaria, et Hirofumi Hara. 2020. « Enzymatic and Genetic Characterization of Lignin Depolymerization by Streptomyces Sp. S6 Isolated from a Tropical Environment ». Scientific Reports 10 (1): 7813.

Roell, Garrett W., Jian Zha, Rhiannon R. Carr, Mattheos A. Koffas, Stephen S. Fong, et Yinjie J. Tang. 2019. « Engineering microbial consortia by division of labor ». Microbial Cell Factories 18 (1): 35.

Rognes, Torbjørn, Tomáš Flouri, Ben Nichols, Christopher Quince, et Frédéric Mahé. 2016. « VSEARCH: a versatile open source tool for metagenomics ». PeerJ 4 (octobre).

Rohart, Florian, Benoît Gautier, Amrit Singh, et Kim-Anh Lê Cao. 2017. « MixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration ». PLOS Computational Biology 13 (11): e1005752.

Rolando, C., B. Monties, et C. Lapierre. 1992. « Thioacidolysis ». In Methods in Lignin Chemistry, 334-49. Springer Series in Wood Science. Springer, Berlin, Heidelberg.

Rooney, William L., Jürg Blumenthal, Brent Bean, et John E. Mullet. 2007. « Designing Sorghum as a Dedicated Bioenergy Feedstock ». Biofuels, Bioproducts and Biorefining 1 (2): 147-57.

Rosenberg, Eugene. 2014. « The Family Chitinophagaceae ». In The Prokaryotes: Other Major Lineages of Bacteria and The Archaea, édité par Eugene Rosenberg, Edward F. DeLong, Stephen Lory, Erko Stackebrandt, et Fabiano Thompson, 493-95. Berlin, Heidelberg: Springer

Rossmassler, Karen, Carsten Dietrich, Claire Thompson, Aram Mikaelyan, James O. Nonoh, Rudolf H. Scheffrahn, David Sillam-Dussès, et Andreas Brune. 2015. « Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites ». Microbiome 3 (novembre).

Sadhukhan, Jhuma, et Elias Martinez-Hernandez. 2017. « Material Flow and Sustainability Analyses of Biorefining of Municipal Solid Waste ». Bioresource Technology 243 (novembre): 135-46.

Sadhukhan, Jhuma, Kok Siew Ng, et Elias Martinez Hernandez. 2014. Biorefineries and Chemical Processes: Design, Integration and Sustainability Analysis. John Wiley & Sons.

Salvachúa, Davinia, Alicia Prieto, Ángel T. Martínez, et María Jesús Martínez. 2013. « Characterization of a Novel Dye-Decolorizing Peroxidase (DyP)-Type Enzyme from Irpex Lacteus and Its Application in Enzymatic Hydrolysis of Wheat Straw ». Applied and Environmental Microbiology 79 (14): 4316-24.

Salvachúa, Davinia, Allison Z. Werner, Isabel Pardo, Martyna Michalska, Brenna A. Black, Bryon S. Donohoe, Stefan J. Haugen, et al. 2020. « Outer Membrane Vesicles Catabolize Lignin-Derived Aromatic Compounds in Pseudomonas Putida KT2440 ». Proceedings of the National Academy of Sciences 117 (17): 9302-10

Sannigrahi, Poulomi, Arthur J. Ragauskas, et Gerald A. Tuskan. 2010. « Poplar as a Feedstock for Biofuels: A Review of Compositional Characteristics ». Biofuels, Bioproducts and Biorefining 4 (2): 209-26

Savelli, Bruno, Qiang Li, Mark Webber, Achraf Mohamed Jemmat, Alexis Robitaille, Marcel Zamocky, Catherine Mathé, et Christophe Dunand. 2019. « RedoxiBase: A Database for ROS Homeostasis Regulated Proteins ». Redox Biology 26 (septembre): 101247.

Scheller, Henrik Vibe, et Peter Ulvskov. 2010. « Hemicelluloses ». Annual Review of Plant Biology 61 (1): 263-89

Schloss, Patrick D. 2020. « Reintroducing mothur: 10 Years Later ». Applied and Environmental Microbiology 86 (2): e02343-19.

Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. « Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities ». Applied and Environmental Microbiology 75 (23): 7537-41.

Schubert, Mikkel, Stinus Lindgreen, et Ludovic Orlando. 2016. « AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging ». BMC Research Notes 9 (février): 88.

Sczyrba, Alexander, Peter Hofmann, Peter Belmann, David Koslicki, Stefan Janssen, Johannes Dröge, Ivan Gregor, et al. 2017. « Critical Assessment of Metagenome Interpretation – a benchmark of computational metagenomics software ». Nature methods 14 (11): 1063-71.

Seemann, Torsten. 2014. « Prokka: Rapid Prokaryotic Genome Annotation ». Bioinformatics 30 (14): 2068-69.

Segata, Nicola, Daniela Börnigen, Xochitl C. Morgan, et Curtis Huttenhower. 2013. « PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes ». Nature communications 4: 2304.

Sethupathy, Sivasamy, Gabriel Murillo Morales, Yixuan Li, Yongli Wang, Jianxiong Jiang, Jianzhong Sun, et Daochen Zhu. 2021. « Harnessing microbial wealth for lignocellulose biomass valorization through secretomics: a review ». Biotechnology for Biofuels 14 (1): 154.

Shang, Jing, Fei Zhu, Wanwipa Vongsangnak, Yifei Tang, Wenyu Zhang, et Bairong Shen. 2014. « Evaluation and Comparison of Multiple Aligners for Next-Generation Sequencing Data Analysis ». Research article. BioMed Research International. 2014.

Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, et Jillian F. Banfield. 2018. « Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy ». Nature Microbiology 3 (7): 836.

Sigoillot, Jean-Claude, Jean-Guy Berrin, Mathieu Bey, Laurence Lesage-Meessen, Anthony Levasseur, Anne Lomascolo, Eric Record, et Eva Uzan-Boukhris. 2012. « Chapter 8 - Fungal Strategies for Lignin Degradation ». In Advances in Botanical Research, édité par Lise Jouanin et Catherine Lapierre, 61:263-308. Lignins. Academic Press.

Silva, Gabriela Ghizzi D., Marie Couturier, Jean-Guy Berrin, Alain Buléon, et Xavier Rouau. 2012. « Effects of Grinding Processes on Enzymatic Degradation of Wheat Straw ». Bioresource Technology 103 (1): 192-200.

Silva, Jéssica P., Alonso R. P. Ticona, Pedro R. V. Hamann, Betania F. Quirino, et Eliane F. Noronha. 2021. « Deconstruction of Lignin: From Enzymes to Microorganisms ». Molecules 26 (8): 2299.

Smibert, Robert M. 1981. « The Genus Treponema ». In The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria, édité par Mortimer P. Starr, Heinz Stolp, Hans G. Trüper, Albert Balows, et Hans G. Schlegel, 564-77. Berlin, Heidelberg: Springer.

Sondhi, Sonica, Prince Sharma, Nancy George, Prakram Singh Chauhan, Neena Puri, et Naveen Gupta. 2015. « An extracellular thermo-alkali-stable laccase from Bacillus tequilensis SN4, with a potential to biobleach softwood pulp ». 3 Biotech 5 (2): 175-85.

Song, Wei-Zhi, et Torsten Thomas. 2017. « Binning_refiner: Improving Genome Bins through the Combination of Different Binning Programs ». Bioinformatics (Oxford, England) 33 (12): 1873-75

Souza, Adriana Carvalho de, Tim Rietkerk, Coralie G. M. Selin, et Peter P. Lankhorst. 2013. « A robust and universal NMR method for the compositional analysis of polysaccharides ». Carbohydrate Polymers 95 (2): 657-63.

Standard Methods Committee, 2010. s. d. « 5550 Tannin and Lignin (2010) ». Consulté le 20 août 2018. https://www.standardmethods.org/store/ProductView.cfm?ProductID=415.

Strachan, Cameron, Dave VanInsberghe, et Dominique Williams. 2012. « Ligninase Activity Is Not Consistently Predicted by the Presence of Manganese Coordinating Residues in Dyp-Like Proteins » 16: 7.

Strous, Marc, Beate Kraft, Regina Bisdorf, et Halina E. Tegetmeyer. 2012. « The Binning of Metagenomic Contigs for Microbial Physiology of Mixed Cultures ». Frontiers in Microbiology 3 (décembre)..

Su, Lijuan, Lele Yang, Shi Huang, Yan Li, Xiaoquan Su, Fengqin Wang, Cunpei Bo, En Tao Wang, et Andong Song. 2017. « Variation in the Gut Microbiota of Termites (Tsaitermes Ampliceps) Against Different Diets ». Applied Biochemistry and Biotechnology 181 (1): 32-47..

Sugano, Y. 2009. « DyP-Type Peroxidases Comprise a Novel Heme Peroxidase Family ». Cellular and Molecular Life Sciences: CMLS 66 (8): 1387-1403.

Sugano, Yasushi, et Toru Yoshida. 2021. « DyP-Type Peroxidases: Recent Advances and Perspectives ». International Journal of Molecular Sciences 22 (11): 5556.

Suzuki, Ryota, et Hidetoshi Shimodaira. 2006. « Pvclust: an R package for assessing the uncertainty in hierarchical clustering ». Bioinformatics 22 (12): 1540-42.

Svartström, Olov, Johannes Alneberg, Nicolas Terrapon, Vincent Lombard, Ino de Bruijn, Jonas Malmsten, Ann-Marie Dalin, et al. 2017. « Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation ». The ISME Journal 11 (11): 2538-51.

Szklarz, Grażyna, et Andrzej Leonowicz. 1986. « Cooperation between fungal laccase and glucose oxidase in the degradation of lignin derivatives ». Phytochemistry 25 (11): 2537-39.

Talebnia, Farid, Dimitar Karakashev, et Irini Angelidaki. 2010. « Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation ». Bioresource Technology, Special Issue on Lignocellulosic Bioethanol: Current Status and Perspectives, 101 (13): 4744-53.

Tamames, Javier, Marta Cobo-Simón, et Fernando Puente-Sánchez. 2019. « Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes ». BMC Genomics 20 (1): 960.

Tanizawa, Yasuhiro, Takatomo Fujisawa, et Yasukazu Nakamura. 2018. « DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication ». Bioinformatics 34 (6): 1037-39.

Tarmadi, Didi, Yuki Tobimatsu, Masaomi Yamamura, Takuji Miyamoto, Yasuyuki Miyagawa, Toshiaki Umezawa, et Tsuyoshi Yoshimura. 2018. « NMR studies on lignocellulose deconstructions in the digestive system of the lower termite Coptotermes formosanus Shiraki ». Scientific Reports 8 (janvier).

Tarmadi, Didi, Tsuyoshi Yoshimura, Yuki Tobimatsu, Masaomi Yamamura, Takuji Miyamoto, Yasuyuki Miyagawa, et Toshiaki Umezawa. 2017. « The Effects of Various Lignocelluloses and Lignins on Physiological Responses of a Lower Termite, *Coptotermes Formosanus*». Journal of Wood Science 63 (5):

Tatusov, Roman L., Michael Y. Galperin, Darren A. Natale, et Eugene V. Koonin. 2000. « The COG database: a tool for genome-scale analysis of protein functions and evolution ». Nucleic Acids Research 28 (1): 33-36.

Taupp, Marcus, Keith Mewis, et Steven J. Hallam. 2011. « The Art and Design of Functional Metagenomic Screens ». Current Opinion in Biotechnology 22 (3): 465-72.

Taylor, C. R., E. M. Hardiman, M. Ahmad, P. D. Sainsbury, P. R. Norris, et T. D. H. Bugg. 2012. « Isolation of Bacterial Strains Able to Metabolize Lignin from Screening of Environmental Samples ». Journal of Applied Microbiology 113 (3): 521-30.

ten Have Rimko, Hartmans Sybe, Teunissen Pauline J.M, et Field Jim A. 1998. « Purification and characterization of two lignin peroxidase isozymes produced by Bjerkandera sp. strain BOS55 ». FEBS Letters 422 (3): 391-94.

Teng, Fei, Sree Sankar Darveekaran Nair, Pengfei Zhu, Shanshan Li, Shi Huang, Xiaolan Li, Jian Xu, et Fang Yang. 2018. « Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling ». Scientific Reports 8 (novembre).

Teramoto, Maki, Masahito Suzuki, Ariani Hatmanti, et Shigeaki Harayama. 2010. « The Potential of Cycloclasticus and Altererythrobacter Strains for Use in Bioremediation of Petroleum-Aromatic-Contaminated Tropical Marine Environments ». Journal of Bioscience and Bioengineering 110 (1): 48-52.

Terrapon, Nicolas, Vincent Lombard, Élodie Drula, Pascal Lapébie, Saad Al-Masaudi, Harry J. Gilbert, et Bernard Henrissat. 2018. « PULDB: The Expanded Database of Polysaccharide Utilization Loci ». Nucleic Acids Research 46 (D1): D677-83.

Thomas, François, Jan-Hendrik Hehemann, Etienne Rebuffet, Mirjam Czjzek, et Gurvan Michel. 2011. « Environmental and Gut Bacteroidetes: The Food Connection ». Frontiers in Microbiology 2 (mai).

Thurston, Christopher F. 1994. « The structure and function of fungal laccases ». Microbiology 140 (1): 19-26

Tien, Ming, et T. Kent Kirk. 1984. « Lignin-degrading enzyme from Phanerochaete chrysosporium: Purification, characterization, and catalytic properties of a unique H2O2-requiring oxygenase ». Proceedings of the National Academy of Sciences of the United States of America 81 (8): 2280-84.

Tilocca, Bruno, Katharina Burbach, Charlotte M. E. Heyer, Ludwig E. Hoelzle, Rainer Mosenthin, Volker Stefanski, Amélia Camarinha-Silva, et Jana Seifert. 2017. « Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome—from days to weeks ». Microbiome 5 (octobre).

Tobimatsu, Yuki, Fang Chen, Jin Nakashima, Luis L. Escamilla-Treviño, Lisa Jackson, Richard A. Dixon, et John Ralph. 2013. « Coexistence but Independent Biosynthesis of Catechyl and Guaiacyl/Syringyl Lignin Polymers in Seed Coats[W][OPEN] ». The Plant Cell 25 (7): 2587-2600.

Tokuda, Gaku, Nathan Lo, Hirofumi Watanabe, Michael Slaytor, Tadao Matsumoto, et Hiroaki Noda. 1999. « Metazoan Cellulase Genes from Termites: Intron/Exon Structures and Sites of Expression ». Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1447 (2): 146-59.

Tokuda, Gaku, Aram Mikaelyan, Chiho Fukui, Yu Matsuura, Hirofumi Watanabe, Masahiro Fujishima, et Andreas Brune. 2018. « Fiber-Associated Spirochetes Are Major Agents of Hemicellulose Degradation in the Hindgut of Wood-Feeding Higher Termites ». Proceedings of the National Academy of Sciences 115 (51): E11996-4.

Tran, Quang, et Vinhthuy Phan. 2020. « Assembling Reads Improves Taxonomic Classification of Species ». Genes 11 (8): 946.

Ufarté, Lisa, Gabrielle Potocki-Veronese, Davide Cecchini, Alexandra S. Tauzin, Angeline Rizzo, Diego P. Morgavi, Bernard Cathala, et al. 2018. « Highly Promiscuous Oxidases Discovered in the Bovine Rumen Microbiome ». Frontiers in Microbiology 9.

Ufarté, Lisa, Gabrielle Potocki-Veronese, et Élisabeth Laville. 2015. « Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology ». Frontiers in Microbiology 6 (juin).

UNFCCC. 2006. « ANNEX 18 DEFINITION OF RENEWABLE BIOMASS ». 2006. http://cdm.unfccc.int/EB/023/eb23_repan18.pdf.

Vanholme, Ruben, Brecht Demedts, Kris Morreel, John Ralph, et Wout Boerjan. 2010. « Lignin Biosynthesis and Structure ». Plant Physiology 153 (3): 895-905.

Vaser, Robert, Dario Pavlović, et Mile Šikić. 2016. « SWORD—a Highly Efficient Protein Database Search ». Bioinformatics 32 (17): i680-84.

Vasudevan, N., et A. Mahadevan. 1991. « Degradation of Lignin and Lignin Derivatives by Acinetobacter Sp. » Journal of Applied Bacteriology 70 (2): 169-76.

Vikram, Surendra, Joel D. Arneodo, Javier Calcagno, Maximiliano Ortiz, Maria Laura Mon, Clara Etcheverry, Don A. Cowan, et Paola Talia. 2021. « Diversity Structure of the Microbial Communities in the Guts of Four Neotropical Termite Species ». PeerJ 9 (avril): e10959.

Vogel, John. 2008. « Unique Aspects of the Grass Cell Wall ». Current Opinion in Plant Biology, Physiology and Metabolism - Edited by Markus Pauly and Kenneth Keegstra, 11 (3): 301-7.

Vollmers, John, Sandra Wiegand, et Anne-Kristin Kaster. 2017. « Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! » PLOS ONE 12 (1): e0169662.

Vos, Michiel, Christopher Quince, Agata S. Pijl, Mattias de Hollander, et George A. Kowalchuk. 2012. « A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity ». PLoS ONE 7 (2).

Voß, Hauke, Carina Amata Heck, Marcus Schallmey, et Anett Schallmey. 2020. « Database Mining for Novel Bacterial β-Etherases, Glutathione-Dependent Lignin-Degrading Enzymes ». Applied and Environmental Microbiology 86 (2).

Voxeur, Aline, Ludivine Soubigou-Taconnat, Frédéric Legée, Kaori Sakai, Sébastien Antelme, Mylène Durand-Tardif, Catherine Lapierre, et Richard Sibout. 2017. « Altered Lignification in Mur1-1 a Mutant Deficient in GDP-L-Fucose Synthesis with Reduced RG-II Cross Linking ». PLOS ONE 12 (9): e0184820.

Wade, William. 2002. « Unculturable bacteria—the uncharacterized organisms that cause oral infections ». Journal of the Royal Society of Medicine 95 (2): 81-83.

Waidele, Lena, Judith Korb, Christian R. Voolstra, Franck Dedeine, et Fabian Staubach. 2019. « Ecological specificity of the metagenome in a set of lower termite species supports contribution of the microbiome to adaptation of the host ». Animal Microbiome 1 (1): 13.

Wang, Cheng, Da Dong, Haoshu Wang, Karin Müller, Yong Qin, Hailong Wang, et Weixiang Wu. 2016. « Metagenomic Analysis of Microbial Consortia Enriched from Compost: New Insights into the Role of Actinobacteria in Lignocellulose Decomposition ». Biotechnology for Biofuels 9 (1): 22

Warnecke, Falk, Peter Luginbühl, Natalia Ivanova, Majid Ghassemian, Toby H. Richardson, Justin T. Stege, Michelle Cayouette, et al. 2007. « Metagenomic and Functional Analysis of Hindgut Microbiota of a Wood-Feeding Higher Termite ». Nature 450 (7169): 560-65.

Wei, Ze-Gang, Xiao-Dan Zhang, Ming Cao, Fei Liu, Yu Qian, et Shao-Wu Zhang. 2021. « Comparison of Methods for Picking the Operational Taxonomic Units From Amplicon Sequences ». Frontiers in Microbiology 12: 474.

Weijde, Tim van der, Andreas Kiesel, Yasir Iqbal, Hilde Muylle, Oene Dolstra, Richard G. F. Visser, Iris Lewandowski, et Luisa M. Trindade. 2017. « Evaluation of Miscanthus Sinensis Biomass Quality as Feedstock for Conversion into Different Bioenergy Products ». GCB Bioenergy 9 (1): 176-90

« Welcome to the Tidyverse ». s. d. Consulté le 26 juillet 2020. https://tidyverse.tidyverse.org/articles/paper.html.

Whittaker, Robert H. 1960. « Vegetation of the Siskiyou Mountains, Oregon and California Ecological Monographs », Ecological Monographs, .

Wick, Ryan R., et Kathryn E. Holt. 2021. « Benchmarking of long-read assemblers for prokaryote whole genome sequencing ». F1000Research 8 (février): 2138.

Wilhelm, Roland C., Rahul Singh, Lindsay D. Eltis, et William W. Mohn. 2019. « Bacterial Contributions to Delignification and Lignocellulose Degradation in Forest Soils with Metagenomic and Quantitative Stable Isotope Probing ». The ISME Journal 13 (2): 413.

Wong, Mabel T., Weijun Wang, Michael Lacourt, Marie Couturier, Elizabeth A. Edwards, et Emma R. Master. 2016. « Substrate-Driven Convergence of the Microbial Community in Lignocellulose-Amended Enrichments of Gut Microflora from the Canadian Beaver (Castor Canadensis) and North American Moose (Alces Americanus) ». Frontiers in Microbiology 7.

Woo, Hannah L, Nicholas R Ballor, Terry C Hazen, Julian L Fortney, Blake Simmons, Karen Walston Davenport, Lynne Goodwin, et al. 2014. « Complete genome sequence of the lignin-degrading bacterium Klebsiella sp. strain BRL6-2 ». Standards in Genomic Sciences 9

Wu, Dongying, Guillaume Jospin, et Jonathan A. Eisen. 2013. « Systematic Identification of Gene Families for Use as “Markers” for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and Archaea and Their Major Subgroups ». PLoS ONE 8 (10).

Wu, Yu-Wei, Blake A. Simmons, et Steven W. Singer. 2016. « MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets ». Bioinformatics (Oxford, England) 32 (4): 605-7.

Xie, Fei, Wei Jin, Huazhe Si, Yuan Yuan, Ye Tao, Junhua Liu, Xiaoxu Wang, et al. 2021. « An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants ». Microbiome 9 (1): 137.

Xie, L., L. Zhang, Y. Zhong, N. Liu, Y. Long, S. Wang, X. Zhou, Z. Zhou, Y. Huang, et Q. Wang. 2012. « Profiling the Metatranscriptome of the Protistan Community in Coptotermes Formosanus with Emphasis on the Lignocellulolytic System. » Genomics 99 (4): 246-55.

Xu, Jiele, Ziyu Wang, et Jay J. Cheng. 2011. « Bermuda grass as feedstock for biofuel production: A review ». Bioresource Technology 102 (17): 7613-20.

Yang, Chenxian, Fangfang Yue, Yanlong Cui, Yuanmei Xu, Yuanyuan Shan, Bianfang Liu, Yuan Zhou, et Xin Lü. 2018. « Biodegradation of Lignin by Pseudomonas Sp. Q18 and the Characterization of a Novel Bacterial DyP-Type Peroxidase ». Journal of Industrial Microbiology and Biotechnology 45 (10): 913-27.

Yarza, Pablo, Pelin Yilmaz, Elmar Pruesse, Frank Oliver Glöckner, Wolfgang Ludwig, Karl-Heinz Schleifer, William B. Whitman, Jean Euzéby, Rudolf Amann, et Ramon Rosselló-Móra. 2014. « Uniting the Classification of Cultured and Uncultured Bacteria and Archaea Using 16S RRNA Gene Sequences ». Nature Reviews Microbiology 12 (9): 635-45.

Yelle, Daniel J., John Ralph, Fachuang Lu, et Kenneth E. Hammel. 2008. « Evidence for Cleavage of Lignin by a Brown Rot Basidiomycete ». Environmental Microbiology 10 (7): 1844-49.

Yilmaz, Pelin, Laura Wegener Parfrey, Pablo Yarza, Jan Gerken, Elmar Pruesse, Christian Quast, Timmy Schweer, Jörg Peplies, Wolfgang Ludwig, et Frank Oliver Glöckner. 2014. « The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks ». Nucleic Acids Research 42 (D1): D643-48.

Yoo, Chang Geun, Yunqiao Pu, Mi Li, et Arthur J. Ragauskas. 2016. « Elucidating Structural Characteristics of Biomass Using Solution-State 2 D NMR with a Mixture of Deuterated Dimethylsulfoxide and Hexamethylphosphoramide ». ChemSusChem 9 (10): 1090-95.

Yuki, Masahiro, Hirokazu Kuwahara, Masaki Shintani, Kazuki Izawa, Tomoyuki Sato, David Starns, Yuichi Hongoh, et Moriya Ohkuma. 2015. « Dominant Ectosymbiotic Bacteria of Cellulolytic Protists in the Termite Gut Also Have the Potential to Digest Lignocellulose ». Environmental Microbiology 17 (12): 4942-53.

Yuki, Masahiro, Kenshiro Oshima, Wataru Suda, Mitsuo Sakamoto, Toshiya Iida, Masahira Hattori, et Moriya Ohkuma. 2014. « Draft Genome Sequence of Bacteroides reticulotermitis Strain JCM 10512T, Isolated from the Gut of a Termite ». Genome Announcements 2 (1): e00072-14. Zeng, J., D. Singh, D. D. Laskar, et S. Chen. 2013. « Degradation of Native Wheat Straw Lignin by Streptomyces Viridosporus T7A ». International Journal of Environmental Science and Technology 10 (1): 165-74.

Zhang, Han, Tanner Yohe, Le Huang, Sarah Entwistle, Peizhi Wu, Zhenglu Yang, Peter K Busk, Ying Xu, et Yanbin Yin. 2018. « dbCAN2: a meta server for automated carbohydrate-active enzyme annotation ». Nucleic Acids Research 46 (W1): W95-101.

Zhang, Wen, Xueyan Ren, Qiong Lei, et Lei Wang. 2021. « Screening and Comparison of Lignin Degradation Microbial Consortia from Wooden Antiques ». Molecules 26 (10): 2862.

Zhang, Xinning, et Jared R. Leadbetter. s. d. « Evidence for Cascades of Perturbation and Adaptation in the Metabolic Genes of Higher Termite Gut Symbionts ». mBio 3 (4): e00223-

Zheng, Qi, Tiantian Zhou, Yibin Wang, Xiaohua Cao, Songqing Wu, Meili Zhao, Haoyuan Wang, et al. 2018. « Pretreatment of Wheat Straw Leads to Structural Changes and Improved Enzymatic Hydrolysis ». Scientific Reports 8 (1): 1321.

Zhou, Haizhu, Wei Guo, Bo Xu, Zhanwei Teng, Dapeng Tao, Yujie Lou, et Yunhang Gao. 2017. « Screening and Identification of Lignin-Degrading Bacteria in Termite Gut and the Construction of LiP-Expressing Recombinant Lactococcus Lactis ». Microbial Pathogenesis 112 (septembre): 63-69.

Zhou, Man, Peng Guo, Tao Wang, Lina Gao, Huijun Yin, Cheng Cai, Jie Gu, et Xin Lü. 2017. « Metagenomic Mining Pectinolytic Microbes and Enzymes from an Apple Pomace-Adapted Compost Microbial Community ». Biotechnology for Biofuels 10 (1): 198.

Zielińska, Sylwia, Piotr Radkowski, Aleksandra Blendowska, Agnieszka Ludwig-Gałęzowska, Joanna M. Łoś, et Marcin Łoś. 2017. « The Choice of the DNA Extraction Method May Influence the Outcome of the Soil Microbial Community Structure Analysis ». MicrobiologyOpen 6 (4): e00453.

Zilber-Rosenberg, Ilana, et Eugene Rosenberg. 2008. « Role of Microorganisms in the Evolution of Animals and Plants: The Hologenome Theory of Evolution ». FEMS Microbiology Reviews 32 (5): 723-35.

Zoghlami, Aya, et Gabriel Paës. 2019. « Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis ». Frontiers in Chemistry 7.

Zuo, Tao, Fen Zhang, Grace C.Y. Lui, Yun Kit Yeoh, Amy Y.L. Li, Hui Zhan, Yating Wan, et al. 2020. « Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization ». Gastroenterology 159 (3): 944-955.e8.

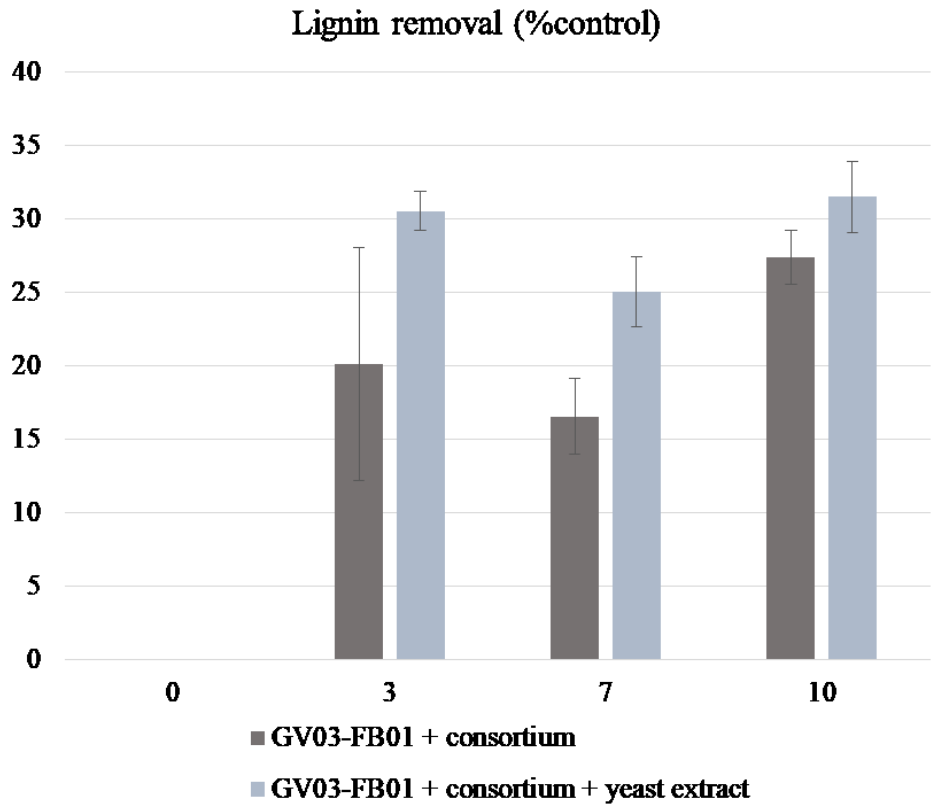
IX.

ANNEX

Annex: Degradation of ethyl-acetate extracted lignin by the consortium

As one of the main goal of the project was to degrade technical lignins, so that we could have more definitive answer on the lignolytic potential of NE15 consortium, we cultivated it on a ethyl acetate extracted technical lignin. We reached a total of 30% removal with or without yeast extract proving once again that NE15 can not only degrade lignin in a natural matrix but also pure lignin unlike the initial gut microbiota. (Figure 1)

Figure 1 : Lignin removal in ethyl acetate-extracted GV03 (GV03-FB01) lignin over time



After 3 days of incubation alsmost all monomers were degraded even in absence of yeast extract. Oligomers took more time but around 30% of them were degraded in ten days without yeast extract and 70% (Figure 2)with it which means another source of energy may be usefull to extract the consortium full lignolytic potential.

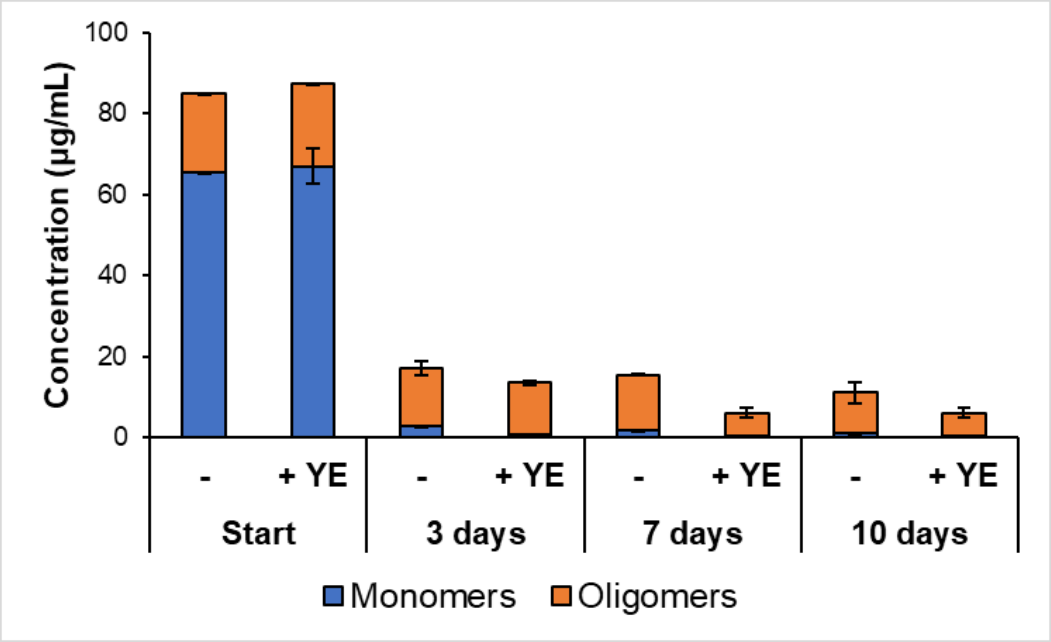


Figure 2: Evolution of the concentration of soluble lignin monomers and oligomers in GV03-FB01 in presence or absence of yeast extract after various incubation time with NE15 consortium

La clef de toutes les sciences est sans contredit le point d'interrogation ; nous devons la plupart des grandes découvertes au comment ? Et la sagesse dans la vie consiste peut-être à se demander, à tout propos, pourquoi ?

Honoré de Balzac