



**HAL**  
open science

# Vers la compréhension du fonctionnement des consortia microbiens lignocellulolytiques : contribution des approches méta-omiques et biostatistiques

Emeline Flajollet

## ► To cite this version:

Emeline Flajollet. Vers la compréhension du fonctionnement des consortia microbiens lignocellulolytiques : contribution des approches méta-omiques et biostatistiques. Biotechnologies. INSA de Toulouse, 2021. Français. NNT : 2021ISAT0037 . tel-03723828

**HAL Id: tel-03723828**

**<https://theses.hal.science/tel-03723828>**

Submitted on 15 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# 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  
Emeline FLAJOLLET**

Le 10 décembre 2021

**Vers la compréhension du fonctionnement des consortia  
microbiens lignocellulolytiques : contribution des approches  
méta-omiques et biostatistiques**

---

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

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :

**TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering**

**Directrice de Thèse**

**Mme GUILLERMINA HERNANDEZ-RAQUET**, Directrice de recherche, INRAE

**Jury**

<b>M. JEAN ARMENGAUD</b> , Directeur de recherche, CEA	Rapporteur
<b>M. Olivier CHAPLEUR</b> , Ingénieur de recherche, INRAE	Rapporteur
<b>Mme EVELYNE FORANO</b> , Directrice de recherche, INRAE	Rapporteuse
<b>M. FADHEL BEN CHAABANE</b> , Ingénieur de recherche, IFPEN	Examineur
<b>Mme Claire DUMAS</b> , Chargée de recherche, INRAE	Examinatrice
<b>Mme Claire DUMON</b> , Chargée de recherche, INRAE	Présidente



# Résumé

La lignocellulose, principal composant des parois végétales et sous-produit agricole très abondant, est une source de carbone renouvelable les plus prometteuses pour la production d'énergie, de produits chimiques ou de biocarburants. L'utilisation de la lignocellulose pour la production de carboxylates repose sur l'utilisation de consortia microbiens. Pour améliorer les rendements de production de carboxylates, ces travaux de thèse se sont intéressés à i) caractériser le fonctionnement des consortia sélectionnés pour leur potentiel lignocellulolytique à partir de différents écosystèmes naturels et ii) à mieux comprendre leurs réponses fonctionnelles face à des changements dans la composition et la structure physicochimique du substrat. Pour cela, des approches d'écologie microbienne et des outils multi-omiques ont été appliqués. Ainsi, des consortia dérivés de rumen bovin ou d'intestin de termites ont été étudiés par des approches de métabarcoding 16S et de métaprotéomique comparative dite *label-free*. Leurs réponses fonctionnelles, enzymatiques et taxonomiques ont été évaluées en utilisant la paille de blé prétraitée par des méthodes mécaniques et/ou chimiques. Ces recherches ont permis de mettre en lumière le rôle et la complémentarité des différentes populations composant les consortia microbiens, ainsi que l'effet des modifications induites par les prétraitements du substrat sur la régulation de l'expression des enzymes impliquées dans la bioconversion des carbohydrates (CAZymes) en carboxylates. De plus, un consortium issu de rumen bovin et enrichi sur des résidus de maïs a également été étudié. L'étude de la dynamique des populations bactériennes au cours de l'enrichissement et de la bioconversion de la lignocellulose a permis de mettre en lumière le rôle des microorganismes libres et attachés au substrat, et de leurs enzymes, fournissant un aperçu inédit du fonctionnement des communautés lignocellulolytiques en bioréacteur.

**Mots clés :** lignocellulose ; métaprotéomique ; digestion anaérobie ; consortia microbiens ; CAZymes ; Métabarcoding ARNr 16S

# Summary

Lignocellulose, the main component of plant walls and very abundant agricultural by-product, is one of the most promising renewable carbon sources for the production of energy, chemicals or biofuels. The use of lignocellulose for the production of carboxylates is based on the use of microbial consortia. To improve carboxylate production yields, this thesis work focused on i) characterizing the functioning of consortia selected for their lignocellulolytic potential from different natural ecosystems and ii) understanding their functional changes in the face of modifications in the physicochemical structure and composition of the substrate. For this, microbial ecology approaches and multi-omics tools were applied. Thus, consortia derived from bovine rumen or termite gut have been studied by 16S metabarcoding and label-free metaproteomics approaches. Their enzyme activities and taxonomic structure were evaluated using wheat straw pretreated by mechanical and/or chemical methods. This research has shed light on the role and complementarity of the different populations making up microbial consortia, as well as the effect of the substrate changes induced by the pretreatment on the regulation of the enzymes expression during the bioconversion of carbohydrates (CAZymes) and carboxylates. In addition, a consortium derived from bovine rumen and enriched on corn stover residues was also performed. The study of the dynamics of bacterial populations during the enrichment and bioconversion of lignocellulose has made it possible to shed light on the role of free microorganisms and those attached to the substrate, and of their enzymes. This thesis work provides an unprecedented insight of the functioning of lignocellulolytic communities in a bioreactor.

**Mots clés :** lignocellulose ; metaproteomic ; anaerobic digestion ; microbial consortia ; CAZymes ; 16S rRNA metabarcoding





## Remerciements

*La soutenance d'une thèse marque la fin d'une étape dans la jeune vie d'un chercheur. Je profite de cette occasion pour exprimer mes plus sincères remerciements à tous ceux et celles qui m'ont encouragé et soutenu à poursuivre ce travail de thèse et à le terminer. Je ne peux tous vous citer donc Merci.*

*Mes remerciements vont en premier lieu à Guillermina Hernandez-Raquet et Claire Dumas pour m'avoir offert la possibilité de réaliser cette thèse. Votre accompagnement quotidien et votre implication m'ont aidé à explorer toutes les possibilités qu'offraient ce travail de thèse : du réacteur à la bioinformatique en passant par la microscopie, le 16S, la métaprotéomique et toutes les autres analyses qu'on a faites pour essayer de comprendre chaque facette de la dégradation de la lignocellulose par nos petites bactéries. Merci Guillermina pour ta présence presque quotidienne et ta volonté de toujours aller plus loin. Tu as toujours été très disponible pour discuter dans ton bureau, souvent tard les vendredis soirs devant R... je te remercie également pour tous les bons moments de vie et discussions que nous avons partagés à l'extérieur du labo lors de concerts ou autour d'une bonne table. Merci Claire pour ton encadrement, ton suivi et tous les conseils sur la fermentation et particulièrement la forte teneur en solides. Ta bienveillance m'a aussi permis de passer des caps et de reprendre confiance en moi quand il le fallait.*

*Je voudrais remercier chaleureusement Madame Evelyne Forano, et Messieurs Olivier Chapleur et Jean Armengaud, qui ont accepté d'être les rapporteurs de cette thèse. Merci également à Madame Claire Dumon d'avoir accepté la présidence de mon jury de thèse, et Monsieur Fadhel Ben Chaabane pour avoir examiné ce travail. Vous avez tous contribué, par vos précieuses remarques, à améliorer la qualité de ce travail et je vous en suis très reconnaissante.*

*Ensuite, j'aimerais remercier l'équipe des Techs : Mansour, Evrard et Simon pour leur aide dans les réacteurs et les analyses. Désolée de vous avoir trop de fois embêté, aussi bien avec mes réacteurs à forte TS qui n'agissent pas ou qui font sauter les plombs du labo, qu'avec mes Sartorius, leurs sondes capricieuses et puis les fuites... ah ces fuites... vive l'anaérobiose ☺*

*Merci à tous les membres de l'équipe Symbiose pour les bons moments passés, leur sympathie et leur encouragement tout au long de ces 3 ans au labo. Ca y est, c'est la fin !*

*Merci à Adèle de m'avoir transmis et confié le bébé de métaprotéomique. Je pense que tu es la seule à pouvoir comprendre la passion qu'on peut mettre dans la thèse sur une aussi longue durée ;) ...*

*Merci à mes stagiaires qui m'ont accompagné durant cette aventure : Guillaume, Emilie, Lévy, Coralie, Léa, Maria, Luc, Alexandre. Cette thèse est aussi une partie de vous et votre travail. J'espère vous avoir en retour apporté un petit quelque chose pour vos futures et belles carrières respectives...*

*En dehors de l'équipe, il y a d'autres personnes qui ont suivi de près ou de loin ce travail, et chacun a eu un rôle à jouer. Vous voir, vous parler, cela m'a toujours fait du bien et vous avez toujours eu un petit geste pour égayer mes journées... Merci Marie-Line pour le suivi de mon périple après le Master et les pauses-café à la MRV. Merci Mr Cam, Florence, Brice et Gilles pour l'aventure iGEMesque qui m'a fait mettre un pied au LISBP. Merci à toute l'équipe d'Adisséo pour votre soutien lorsque j'ai candidaté à cette thèse (pensées à notre Sandrine...).*

*Merci à Mathieu mon binôme d'aventures depuis l'iGEM, puis durant le Master, puis le stage et puis la thèse...(quelle est la prochaine étape ?) et son fidèle acolyte Arthur. Merci à vous deux de m'avoir fait autant rire durant cette thèse. Merci à Louison de m'avoir supporté dans le bureau lorsque je refaisais le monde des CAZymes ☺ Et puis merci à Emilie, ma petite stagiaire revenue en thèse avec nous et qui est devenue une amie*

## Remerciements

---

*en or ♥. Et puis petit clin d'oeil à mes amis de la vraie vie, Stéphan, Aurélie, Jérôme, Nils qui me remontaient et me remontent toujours le moral ! ♥*

*Je vais terminer par le plus important : la famille ! Merci à mes frères et sœurs et leurs petites familles d'avoir été là. Merci à mes parents de m'avoir laissé partir faire mes études loin d'eux à Toulouse, je sais que vous êtes fiers de ce travail et vous pouvez l'être.*

*Merci aussi à ma belle famille et particulièrement à mes beaux parents qui m'ont soutenu cette dernière année à jongler entre manuscrit de thèse, conseil municipal, casquette de com et biberons.*

*Je termine par celui que je n'aurais pas rencontré sans cette thèse : Lucas. Je ne pense pas que cette thèse aurait pu être aussi aboutie sans toi et toutes nos discussions sur le sujet, je crois que tu sais déjà tout : tu as été mon pilier durant cette thèse, et maintenant, dans la vie de tous les jours et on y est enfin arrivé.*

*Enfin, cette thèse je la dédie à mes M&MS, à Maika ma plus fidèle compagne du haut de ses 4 pattes durant ces journées et ces nuits passées sur R, mais surtout à ma petite Marine, pour qui clôturer cette thèse fut le plus beau premier cadeau d'anniversaire que je pouvais lui offrir.*

# Sommaire

## Table des matières

<i>Résumé</i>	<b>1</b>
<i>Remerciements</i>	<b>3</b>
<i>Sommaire</i>	<b>5</b>
<i>Table des illustrations et tableaux</i>	<b>9</b>
<i>Publications, communications, collaborations scientifiques et encadrement</i>	<b>15</b>
<i>Introduction générale</i>	<b>21</b>
<b>I. Chapitre 1 – Etat de l’art</b>	<b>27</b>
<b>A. La lignocellulose</b>	<b>27</b>
1. Composition de la lignocellulose	27
2. Structure de la lignocellulose	33
3. Hydrolyse de la lignocellulose	34
<b>B. La bioraffinerie lignocellulosique</b>	<b>44</b>
1. De la lignocellulose aux carboxylates	45
2. Stratégies d’optimisation de la production de carboxylates	55
<b>C. Approches omiques pour la compréhension des consortia microbiens lignocellulolytiques</b>	<b>73</b>
1. Le métabarcoding 16S	74
2. L’analyse métagénomique	82
3. Limite des techniques omiques	88
<b>II. Chapitre 2 – Matériel et méthodes</b>	<b>91</b>
<b>A. Substrats utilisés</b>	<b>91</b>
1. Les résidus de paille de blé bruts	91
2. Les résidus de paille de blé prétraités	91
3. Les résidus de maïs bruts	92
<b>B. Origine des consortia microbiens</b>	<b>93</b>
1. Inocula dérivés de rumen bovin	93
2. Inoculum issu d’intestin de termite supérieur	93
<b>C. Mise en œuvre des bioréacteurs</b>	<b>94</b>

1.	Enrichissement d'un consortium microbien	98
2.	Essais de stabilité face à la cryo-conservation et mise à l'échelle	98
3.	Cinétiques de caractérisation du consortium microbien	99
<b>D.</b>	<b>Analyses macrocinétiques</b>	<b>99</b>
1.	Dosage des acides gras volatils par chromatographie à phase gazeuse	99
2.	Suivi des gaz produits	100
3.	Dosage de l'activité xylanase	100
<b>E.</b>	<b>Analyses omiques</b>	<b>101</b>
1.	Caractérisation de la diversité microbienne par métabarcoding 16S	101
2.	Caractérisation de la diversité fonctionnelle et taxonomique des protéines par analyse métaprotéomique	105
<b>III.</b>	<b><i>Chapitre 3 – Analyse métaprotéomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV</i></b>	<b>111</b>
<b>A.</b>	<b>Introduction</b>	<b>111</b>
<b>B.</b>	<b>Horizontal metaproteomics and CAZymes analysis of lignocellulose decomposition by microbial consortia derived from cow rumen and termite gut</b>	<b>113</b>
1.	Abstract	114
2.	Introduction	116
3.	Materials and methods	120
4.	Results	128
5.	Discussion	149
6.	Conclusions	158
7.	Additional files	160
8.	Acknowledgements	181
9.	Funding	181
10.	Author contributions	181
<b>C.</b>	<b>Conclusion</b>	<b>183</b>
<b>IV.</b>	<b><i>Chapitre 4 – Analyse métaprotéomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements</i></b>	<b>187</b>
<b>A.</b>	<b>Introduction</b>	<b>187</b>
<b>B.</b>	<b>Impact of substrate changes induced by dry-alkali pretreatment on the lignocellulolytic bioconversion of rumen-derived microbial consortium</b>	<b>189</b>
1.	Abstract	190

2.	Introduction	192
3.	Materials and methods	193
4.	Results and discussion	197
5.	Conclusions	211
6.	Additional files	213
7.	Funding	220
8.	Author contributions	220
<b>C.</b>	<b>Conclusion</b>	<b>221</b>
<b>V.</b>	<b><i>Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs</i></b>	<b>227</b>
<b>A.</b>	<b>Introduction</b>	<b>227</b>
<b>B.</b>	<b>Enrichment, stability and scale-up of a cow-rumen derived microbial consortium producing carboxylates from corn stover</b>	<b>229</b>
1.	Abstract	229
2.	Introduction	229
3.	Materials and methods	231
4.	Results and discussion	235
5.	Conclusion	252
6.	Additional files	254
7.	Acknowledgements	262
8.	Funding	262
<b>C.</b>	<b>Conclusion</b>	<b>263</b>
<b>VI.</b>	<b><i>A propos de la transformation CLR des données compositionnelles</i></b>	<b>267</b>
	<b><i>Conclusions générales et perspectives</i></b>	<b>273</b>
	<b><i>Références bibliographiques</i></b>	<b>281</b>



## Table des illustrations et tableaux

### Figures

<i>Figure I-1 : Structure moléculaire de la cellulose.</i>	28
<i>Figure I-2 : Représentation schématique des liaisons glycosidiques <math>\beta</math>-(1,4) (violet) et des liaisons hydrogène intra-chaînes (rouges) et inter-chaînes (bleues) de la cellulose.</i>	28
<i>Figure I-3 : Organisation des fibres de cellulose (A) et états structuraux d'une microfibrille de cellulose (B) avec des régions amorphes et cristallines.</i>	29
<i>Figure I-4 : Exemples de structures de trois hémicelluloses (xyloglucane, <math>\beta</math>-glucane, glucuronoarabinoxylane) retrouvées dans les parois végétales.</i>	30
<i>Figure I-5 : Structure moléculaire de la lignine résultant de la polymérisation tridimensionnelle de trois monolignols (l'alcool <i>p</i>-coumarylique, l'alcool coniférylique et l'alcool sinapylique).</i>	31
<i>Figure I-6 : Représentation schématique des pectines.</i>	32
<i>Figure I-7 : Représentation tridimensionnelle de la structure de la lignocellulose.</i>	33
<i>Figure I-8 : Représentation des activités enzymatiques responsables de la dégradation de la cellulose.</i>	35
<i>Figure I-9 : Représentation des activités enzymatiques responsables de la dégradation de l'arabinoxylane.</i>	37
<i>Figure I-10 : Représentation des différents modules du cellulosome de la bactérie anaérobie <i>Clostridium cellulovorans</i>.</i>	42
<i>Figure I-11 : Représentation du système Sus (Starch utilization system) chez <i>Bacteroides thetaioatomicron</i> pour l'hydrolyse du glucane.</i>	43
<i>Figure I-12 : Schéma des composés chimiques et leurs dérivés obtenus à partir des carbohydrates (C<sub>5</sub>-C<sub>6</sub>) présents dans la lignocellulose.</i>	46
<i>Figure I-13 : Représentation simplifiée des étapes de la digestion anaérobie : de l'hydrolyse de la matière organique complexe à la méthanogénèse.</i>	48
<i>Figure I-14 : Représentation schématique du concept de la lignocellulose comme plateforme de carboxylates pour la production de biocarburants et de composés chimiques.</i>	52
<i>Figure I-15 : Schéma simplifié des voies métaboliques aboutissant à la production d'acides gras volatils (acétate, propionate, butyrate, valérate, caproate) et rendements en NADH, ATP et pyruvate lors de la dégradation du xylose et du glucose.</i>	53
<i>Figure I-16 : Exemple de procédés de post-traitements chimiques permettant de convertir les carboxylates en biocarburants ou en solvants.</i>	55
<i>Figure I-17 : Résumé des optimisations possibles des paramètres de la digestion anaérobie pour la production d'acides gras volatils.</i>	57



<i>Figure I-18 : Représentation schématique de l'effet destructurant d'un prétraitement sur la biomasse lignocellulosique.</i>	59
<i>Figure I-19 : Représentation des principales voies métaboliques de dégradation de la lignocellulose dans le rumen bovin et implication des principaux groupes microbiens.</i>	69
<i>Figure I-20 : Principe d'enrichissement d'un consortium microbien sur substrat lignocellulosique par repiquage successif en bioréacteurs discontinus séquentiels (SBR, Sequential Batch Reactor).</i>	71
<i>Figure I-21 : Structure de l'ARNr 16S et position des neuf régions hypervariables (V1-V9).</i>	76
<i>Figure I-22 : Schéma résumant les principales étapes de préparation des librairies Illumina et la construction des amplicons séquencés.</i>	78
<i>Figure I-23 : Principes généraux du séquençage Illumina : amplification des clusters par ponts (à gauche) et séquençage par synthèse (à droite).</i>	79
<i>Figure I-24 : Schéma général du pipeline d'analyse FROGS.</i>	80
<i>Figure I-25 : Processus d'identification des protéines et de la création des groupes de protéines à partir des peptides identifiées en spectrométrie de masse en tandem (MS/MS).</i>	84
<i>Figure I-26 : Suite des étapes de bioinformatique pour l'analyse métaprotéomique shotgun label-free.</i>	86
<i>Figure II-1 : Présentation du système de prélèvement utilisé lors de la conduite des bioréacteurs en anaérobiose.</i>	96
<i>Figure III-1 : A. Taxonomic community composition at genus level deduced from the identified proteins (metaproteomics data) for RWS (top graphs) and TWS (bottom graphs). B. Principal components analysis (PCA) ordination plots with 95% confidence ellipses at the major genus level based on protein abundance (CLR-transformed data) at different sampling times of RWS (blue) and TWS (green) microbial consortia.</i>	132
<i>Figure III-2 : Clusters of orthologous groups (COGs) in the metaproteomes of RWS and TWS replicates (1 and 2 indicate the biological duplicates).</i>	134
<i>Figure III-3 : Relative abundance of CAZyme classes (glycoside hydrolases (GH), carbohydrate esterases (CE), carbohydrate binding modules (CBM), glycosyl transferases (GT), polysaccharide lyases (PL), S-layer homology (SLH), dockerins (DOC), cohesins (COH)) across sampling time of RWS and TWS consortia replicates.</i>	136
<i>Figure III-4 : Comparison of abundance of CAZyme families classified by their function and taxonomy in the metaproteomes of RWS and TWS consortia.</i>	139
<i>Figure III-5 : Clustered Image Maps (CIM) represented the most discriminant CAZymes for RWS and TWS consortia at the different sampling times.</i>	145

<i>Figure III-6 : Relative abundance and phylogenetic origin of bacterial enzymes involved in butyrate, propionate and acetate production in the metaproteomes of RWS and TWS (1 and 2 indicate the biological duplicates).</i>	148
<i>Figure III-7 : Main CAZymes and enzyme activities discriminating RWS and TWS consortia and the wheat straw degradation phases.</i>	153
<i>Figure III-8 : A. Comparison of the protein abundance (CLR-transformed data) affiliated to the major genus present in RWS and TWS metaproteomes. B. Comparison of the proteins abundance (CLR-transformed data) affiliated to the major genus in phase A (T1 point) and phase B (T2-T3-T4 points).</i>	167
<i>Figure IV-1 : Wheat straw degradation percentage (%), VFA production (mCmol/L) and xylanase levels (UA/L) over time for each pretreatment (mean of two replicates per pretreatment, coloured as detailed in legend).</i>	194
<i>Figure IV-2 : Three-dimensional Principal Component Analysis (PCA) of Centered Log Ratio (CLR)-transformed metaproteomic data (PC1 to PC3).</i>	199
<i>Figure IV-3 : Taxonomic community composition of RWS deduced from identified proteins for each time point and substrate condition in the biological replicate bioreactors.</i>	200
<i>Figure IV-4 : Comparison of the protein abundance affiliated to major genera found in RWS during fermentation of non-alkali (A and B) and dry-alkali pretreated (C and D) substrates (CLR-transformed data).</i>	202
<i>Figure IV-5 : Comparison of the abundances of the nine CAZy families (CLR-transformed data).</i>	205
<i>Figure IV-6 : Comparison of the allocation index (AI) of Bacteroides and Clostridium between alkali-treated and non-treated substrates for the nine CAZy families with significant abundance differences.</i>	207
<i>Figure IV-7 : Comparison of Allocation index (AI) between treatments abundances of the proteins related to VFA production.</i>	210
<i>Figure IV-8 : Représentation du PUL20 de Bacteroides graminisolvens identifié et référencé dans la PULDB</i>	223
<i>Figure IV-9 : Représentation du PUL23 de Bacteroides graminisolvens identifié et référencé dans la PULDB</i>	223
<i>Figure V-1 : Volatile fatty acids (VFA) production (gC/L) at the end of each cycle of the two corn-stover enriched communities throughout the enrichment process.</i>	236
<i>Figure V-2 : Relative distribution of the bacterial taxa (Family level) of the two corn-stover enriched communities throughout the enrichment process.</i>	238
<i>Figure V-3 : Temporal dynamic of RCS corn stover bioconversion into VFAs and xylanase activity in duplicate inoculated bioreactors (Reactors 1 and 2) and one uninoculated control (Corn stover).</i>	244

## Table des illustrations et tableaux

---

*Figure V-4 : Dynamics of the bacterial taxa (Family level) throughout the corn stover bioconversion by RCS consortium.* 247

*Figure V-5 : Longitudinal analysis integrating dynamics of RCS main bacterial families (CLR transformed data) and individual fatty acid produced in each reactor, using timeOmics R package following the authors guidelines (Bodein et al., 2019).* 249

*Figure V-6 : Xylanase activity and community composition in each sample fraction (free, weakly attached and attached to the substrate).* 251

*Figure VI-1 : Limites de la transformation CLR. Exemple de données issues du Chapitre 4 lors de la comparaison entre substans prétraités (A-B vs C-D) de l'expression de quelques familles CAZy.* 269

*Figure VI-2 : Limites de la transformation CLR. Exemple théorique illustrant l'effet de la compositionnalité et des différences de moyenne géométrique sur les interprétations de données CLR.* 270

---

## Tableaux

---

<i>Tableau I-1 : Descriptif des activités enzymatiques pour l'hydrolyse de la cellulose (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).</i>	36
<i>Tableau I-2 : Descriptif des principales activités enzymatiques existantes pour l'hydrolyse des hémicelluloses (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).</i>	39
<i>Tableau I-3 : Descriptif des principales activités enzymatiques existantes pour l'hydrolyse des pectines (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).</i>	41
<i>Tableau I-4 : Compositions biochimiques des différentes fractions majeures (% cellulose, hémicelluloses, lignine exprimés en matière sèche) de biomasses lignocellulosiques issus des résidus agricoles, référencées dans la littérature</i>	58
<i>Tableau I-5 : Effets sur la lignocellulose, avantages et inconvénients des grands types de prétraitements de la biomasse lignocellulosique</i>	60
<i>Tableau II-1 : Résumé des conditions des prétraitements mécaniques et chimiques pour chacun des substrats de paille de blé A-B-C-D utilisés dans ces travaux de thèse (MS : matière sèche).</i>	92
<i>Table III-1 : RWS and TWS sample characteristics</i>	121
<i>Table III-2 : Diversity of species (OTUs) and proteins detected in RWS and TWS samples</i>	129
<i>Table III-3 : Diversity of CAZymes found in the metaproteomics dataset</i>	137
<i>Table III-4 : PERMANOVA analysis based on Euclidean distance, using inoculum source (“inoculum”: RWS or TWS), degradation phase (“phase” : phase A (T1 points) or phase B (T2-T3-T4 points)) and interaction between them (Inoculum:Phase”) as factors.</i>	144
<i>Tableau III-5 : Comparative analysis of main bacterial GH families targeting plant cell wall degradation found in this study and in other omics studies concerning bovine rumen and termite hindgut microbiomes.</i>	156



## Publications, communications, collaborations scientifiques et encadrement

### Publications dans journaux à comité de lecture

**Horizontal metaproteomics and CAZymes analysis of lignocellulose decomposition by microbial consortia derived from cow rumen and termite gut.** [Emeline Auer](#), Adèle Lazuka, Nico Jehmlich, Sébastien Déjean, Vincent Lombard, Bernard Henrissat, Michael O'Donohue, Guillermina Hernandez-Raquet. **En préparation.**

**Impact of substrate changes induced by dry-alkali pretreatment on the lignocellulolytic bioconversion of rumen-derived microbial consortium.** [Emeline Auer](#), Nico Jehmlich, Vincent Lombard, Bernard Henrissat, Guillermina Hernandez-Raquet. **En préparation.**

**Enrichment, stability and scale-up of a cow-rumen derived microbial consortium producing carboxylates from corn stover.** [Emeline Auer](#), Emilie Alaux, Guillaume Caillou, Diego P. Morgavi, Claire Dumas, Guillermina Hernandez-Raquet. **En préparation.**

### Communications

[Flajollet E.](#), Barakat A., Dumas C., Dumon C., Guillon F., Hernandez-Raquet G. *Impact du prétraitement par le procédé d'extrusion-réactive sur la diversité fonctionnelle d'un consortium microbien responsable de la transformation de la lignocellulose à haute teneur en solides.* 13<sup>ème</sup> Colloque de Génomique Environnementale, 26-28 octobre 2015, Montpellier (France). **Présentation poster.**

Lazuka A., Flajollet E., Abot A., Auer L., Dumon C., Le Berre V., O'Donohue M., von Bergen M., Jehmlich N., [Hernandez-Raquet G.](#) *Diversity of microorganisms and enzymes involved on wheat straw transformation by microbiomes derived from cow rumen and termite gut.* 38<sup>th</sup> Symposium on Biotechnology for Fuels and Chemicals, 25-28 avril 2016, Baltimore (Etats-Unis). **Présentation orale.**

Lazuka A., Auer L., Flajollet E., von Bergen M., O'Donohue M., Jehmlich N., Hernandez-Raquet G. *Dynamics of proteins and microorganisms during wheat straw transformation by microbial consortia*. Lignobiotech 4<sup>th</sup> Symposium on biotechnology applied to lignocelluloses, 19-22 juin 2016, Madrid (Espagne). **Présentation orale.**

Flajollet E., Lazuka A., Auer L., O'Donohue M., Jehmlich N., von Bergen M., Hernandez-Raquet G. *Dynamic of carbohydrate active enzymes (CAZymes) throughout lignocellulose degradation by a microbial consortium : Impact of dry chemo-mechanical pretreatment*. 39<sup>th</sup> Symposium on Biotechnology for Fuels and Chemicals, 1<sup>er</sup>-4 mai 2017, San Francisco (Etats-Unis). **Présentation poster.**

Flajollet E., Lazuka A., Auer L., Jehmlich N., Henrissat B., von Bergen M., O'Donohue M., Hernandez-Raquet G. *Metagenomics and metaproteomics provide insights into the roles of Bacteroidetes and Firmicutes in lignocellulose deconstruction by wheat straw-enriched microbial consortia*. 2<sup>nd</sup> International Metaproteomics Symposium, 14-16 juin 2017, Alghero (Sardaigne). **Présentation orale.**

Flajollet E., Auer L., Lazuka A., O'Donohue M., Jehmlich N., von Bergen M., Henrissat B., Hernandez-Raquet G. *Functional relationship between Bacteroidetes and Firmicutes in lignocellulolytic microbial consortia : an integrative multi-omics approach*. 17<sup>th</sup> International Symposium on Microbial Ecology, 12-17 août 2018, Leipzig (Allemagne). **Présentation poster. (Travel award).**

Flajollet E., Lazuka A., Lombard V., Henrissat B., O'Donohue M., Dumas C., Jehmlich N., Hernandez-Raquet G. *Deciphering the temporal dynamics of lignocellulolytic enzymes in microbial consortia derived from cow rumen and termite gut throughout wheat straw degradation*. 3<sup>rd</sup> International Metaproteomics Symposium, 3-5 décembre 2018, Leipzig (Allemagne). **Présentation orale.**

Flajollet E., Lazuka A., Auer L., Jehmlich N., Henrissat B., Hernandez-Raquet G. *Impact of biomass pretreatment on CAZymes expression of a lignocellulolytic consortium : a metaproteomic assessment*. 41<sup>st</sup> Symposium on Biotechnology for Fuels and Chemicals, 28 avril-1<sup>er</sup> mai 2019, Seattle (Etats-Unis). **Présentation orale.**

## Collaboration scientifique

**SystOmics (Systems approaches to study microbial consortia : integrating meta-omics to elucidate the functioning of lignocellulolytic microbial consortia).** Post-doctorant : Myriam Badawi. Coordinateur : G. Hernandez-Raquet. Financement Metaprogramme MEM (Métaomiques et écosystèmes microbiens) INRA 2016.

**Metaproteomic analysis of hydrolytic samples derived from natural ecosystems.** UFZ Leipzig-Germany. (2014-2015). Financement INRA DARESE et Campus France

## Encadrement

**Guillaume Caillou**, 6 mois – 5<sup>ème</sup> année UTC Compiègne (Caractérisation de l'enrichissement à faible teneur en solides sur résidus de maïs non prétraités)

**Emilie Alaux**, 6 mois – 5<sup>ème</sup> année SupAgro Montpellier (Caractérisation de l'enrichissement et des cinétiques à faible teneur en solides sur résidus de maïs non prétraités)

**Léa Ledaine**, 6 mois – 4<sup>ème</sup> année UTC Compiègne (*Mise au point du protocole FISH sur des bioréacteurs lignocellulosiques pour la visualisation de la colonisation microbienne*) Encadrée par Claire Dumas

**Coralie Roy-Camille**, 5 mois – Césure Master 2 Université Paul Sabatier Toulouse (Essais enzymatiques à faible et forte teneurs en solides et préparation d'un run de séquençage)

**Maria Tovilla Coutiño**, 6 mois – 5<sup>ème</sup> année SupAgro Montpellier (Screening/Cinétiques des prétraitements du projet Hi-SOLIDS à faible teneur en solides)

**Luc Durand**, 6 mois – 5<sup>ème</sup> année Ecole d'Ingénieurs Toulouse (Cinétiques à forte teneur en solides sur paille de maïs prétraitée)

**Alexandre Delarouzée**, 6 mois – 5<sup>ème</sup> année UTC Compiègne (Caractérisation des cinétiques à faible et forte teneurs en solides sur résidus de maïs prétraités)





*Le succès n'est pas final. L'échec n'est pas fatal.  
C'est le courage de continuer qui compte.*  
Winston Churchill

A notre fille, Marine



## Introduction générale

Les micro-organismes, longtemps restés invisibles et inconnus, sont pourtant partout dans notre environnement et l'ont même parfois façonné depuis plus de trois milliards d'années. Extrêmement variés, ils présentent une diversité métabolique inégalée dans le vivant, qui leur permet d'être présents dans tous les milieux, même les plus extrêmes. Leur apparente simplicité comparée aux organismes pluricellulaires macroscopiques cache leur importance pour les écosystèmes. Les micro-organismes jouent en effet des rôles primordiaux dans tous les cycles biogéochimiques : carbone, azote, phosphore... Ils occupent également des positions clés à la base de la plupart des chaînes alimentaires, en tant que recycleurs ou producteurs primaires. Même quand on les pense absents, ils sont très souvent présents à l'intérieur d'autres organismes qu'ils colonisent et avec lesquels ils établissent des interactions, aussi nombreuses que variées dans le monde vivant. Ces interactions peuvent être bénéfiques pour l'organisme hôte qui profite alors des capacités métaboliques des micro-organismes, on parle alors de symbiose ou de mutualisme. Ainsi, les plantes légumineuses, associées à une bactérie du genre *Rhizobium*, deviennent capables de fixer l'azote. Les algues *Xanthelles* permettent aux coraux de développer leur exosquelette calcaire quand elles s'y associent. Enfin, la plupart des animaux présentent des microbiotes intestinaux : leur système digestif est colonisé par des micro-organismes qui participent à son fonctionnement, et permettent d'améliorer les capacités de digestion de l'organisme hôte. Les herbivores ou les insectes xylophages, dont le régime alimentaire est intégralement constitué de végétaux, sont incapables d'en digérer les principaux composants sans interaction avec leur flore intestinale. Ainsi, l'écologie microbienne est la science qui étudie les interactions faisant intervenir un ou plusieurs micro-organismes. Etant donné la grande diversité de ces interactions, c'est un domaine très vaste, mais aussi très récent puisque les outils moléculaires permettant leur étude datent de quelques dizaines d'années seulement.

Néanmoins, même si les processus à l'œuvre n'étaient alors pas compris, les capacités des micro-organismes ont été exploitées depuis les débuts de la civilisation humaine, et elles ont même probablement contribué à façonner notre histoire. La fermentation alcoolique de céréales en bière est attestée depuis plus de dix millénaires, et est avancée par certains archéologues comme un des principaux moteurs de la sédentarisation de l'Homme et du développement de

l'agriculture. Ces premières « biotechnologies » préhistoriques exploitent des processus naturels pré-existants afin de transformer différents substrats en produits d'intérêt. Outre la fermentation alcoolique de céréales et de fruits, d'autres processus permettant de transformer et/ou mieux conserver les aliments sont empiriquement maîtrisés depuis plusieurs siècles, et parfois même depuis plusieurs millénaires : panification des céréales, fermentation lactique de lait (crème, beurre, fromage, Kéfir...), de légumes (choucroute, Kimchi coréen, Takuan japonais...) ou de poisson (Surströmming norvégien, Nuoc-mâm chinois...), ou encore ensilage de fourrage à destination des animaux d'élevage. Les processus impliqués ne se limitent pas à l'alimentation, puisque le compostage des déchets végétaux est une adaptation humaine du processus naturel d'humification et est largement répandu dès l'Antiquité. Les premiers procédés d'épuration de l'eau datent de l'époque moderne et reposent aussi sur des micro-organismes. Cependant, ils remplissaient leur fonction sans qu'on comprenne les mécanismes sous-jacents ou qu'on soit capable de bien les identifier. C'est seulement avec la génétique et les progrès de la microbiologie qu'on a pu commencer à étudier les populations microbiennes et leurs fonctions dans ces procédés. La révolution des omiques permise par les progrès du séquençage de l'ADN (génomique et transcriptomique) et de la spectrométrie de masse (protéomique) puis leur utilisation dans le contexte de l'étude des communautés microbiennes (méta-omiques), ont ouvert une nouvelle ère dans l'étude de ces écosystèmes microbiens, qu'ils soient naturels ou façonnés par l'Homme.

Les biotechnologies, au sens moderne du terme, ont connu un essor très rapide depuis le début du XXI<sup>ème</sup> siècle. Avec l'utilisation des micro-organismes génétiquement modifiés et le développement de procédés pour la production industrielle de protéines ou d'enzymes, elles se sont plutôt orientées vers le génie génétique et enzymatique. Les stratégies mises en place consistent à copier et transférer (et éventuellement optimiser) une propriété unique d'un micro-organisme vivant, portée par un ou plusieurs gènes. C'est donc une approche différente de la plupart des procédés plus anciens et empiriquement maîtrisés faisant intervenir beaucoup plus d'acteurs (fermentation, compostage, traitement des eaux), et des processus microbiologiques naturels tels que les processus naturels de dégradation de la matière végétale. La bioconversion de cette dernière peut prendre de nombreuses formes comme la digestion par les herbivores et les xylophages, l'humification dans les sols forestiers ou la méthanisation dans les tourbières. Elle peut concerner des substrats très différents, mais tous ont en commun un composant majoritaire :

la lignocellulose, principal constituant des parois cellulaires végétales. Ce biopolymère est le plus abondant sur Terre et possède une structure physico-chimique complexe. Les trois fractions qui composent la lignocellulose (cellulose, hémicelluloses et lignine en quantité variables selon le type de plante et son origine) sont entremêlées et chimiquement liées les unes aux autres, et lui confèrent des propriétés de rigidité et de résistances mécanique et biotique. En effet, la grande diversité de molécules et de liaisons chimiques constituant la lignocellulose requière une diversité équivalente d'enzymes pour la dégrader. Malgré sa grande résistance et dans le contexte actuel de recherche d'alternatives à l'utilisation des ressources fossiles, la lignocellulose est vue comme une source potentielle majeure de molécules carbonées renouvelables puisqu'elle est le sous-produit de nombreuses activités agricoles. La valorisation des résidus agricoles lignocellulosiques sous forme de biocarburants, de biogaz ou en produits à haute valeur ajoutée, est donc un enjeu majeur. Cependant, les premières approches de biotechnologies « classiques » par hydrolyse enzymatique se sont heurtées à la faiblesse des rendements obtenus malgré la découverte de nombreuses enzymes produites par des micro-organismes lignocellulolytiques, leur optimisation et leur emploi combiné en « cocktails » enzymatiques. Les procédés de prétraitements du substrat ont permis d'améliorer l'accessibilité de la biomasse aux enzymes ainsi que les rendements, mais ils restent insuffisants à ce jour pour qu'une bioraffinerie lignocellulosique puisse être économiquement rentable.

Et pourtant, la bioconversion de la lignocellulose est un processus naturellement réalisé, avec des rendements et des vitesses de dégradation largement supérieurs dans le rumen de chaque ruminant, le cæcum de chaque cæcotrophe ou plus généralement dans le système digestif de tous les herbivores, qu'ils soient des Mammifères, Insectes ou Invertébrés. Dans le cas du rumen des ruminants, les produits de cette bioconversion par la flore microbienne ruminale sont des gaz mais aussi, en partie, des acides gras volatils (aussi dénommés carboxylates ou AGV). Les AGV sont des petites molécules carbonées d'un grand intérêt en chimie verte comme pour la production de bioplastiques ou de biocarburants. L'étude de ces flores digestives permet d'obtenir des connaissances clés pour améliorer la bioconversion de la biomasse lignocellulosique, et l'utilisation directe en culture de ces communautés microbiennes, productrices d'enzymes lignocellulolytiques et d'AGV, a émergé comme une alternative potentiellement viable par rapport à l'emploi d'enzymes seules. Cependant, cette approche présente certaines difficultés : les sources

d'inoculum sont nombreuses, mais dans une optique d'utilisation industrielle, les populations microbiennes (ou *a minima* leurs fonctions lignocellulolytiques) doivent pouvoir être maintenues en bioréacteur, et être stables dans le temps face à des perturbations. Pour ce faire, on utilise des conditions de culture contrôlées en bioréacteur les plus proches possibles des conditions naturelles pour un inoculum donné (température, pH, anaérobiose). De nombreuses études se sont appliquées à transférer en bioréacteur des communautés microbiennes issues de sols et de sédiments puisque ces environnements présentent une diversité microbienne très élevée et des conditions environnementales assez simples à reproduire. Les procédés d'enrichissement en micro-organismes actifs par cultures successives ont permis d'obtenir de nombreux consortia microbiens stables, avec de bonnes capacités de bioconversion de la biomasse. Au cours de ces enrichissements, bien que la fonction lignocellulolytique soit maintenue dans la plupart des cas, l'enrichissement induit de profonds changements au sein de la structure de la communauté microbienne. En effet, même si leurs capacités de bioconversion sont plus élevées (notamment leur vitesse) et leurs produits plus intéressants, la complexité de leur environnement naturel, régulé par la physiologie de l'hôte, rend plus difficile leur transfert en bioréacteur. Cependant, plusieurs études ont montré qu'il était possible de stabiliser des consortia microbiens actifs à partir d'inocula issus du rumen de différents Mammifères. Les travaux de deux précédentes thèses réalisées au sein du laboratoire ont notamment permis d'obtenir un consortium issu de rumen bovin et un consortium issu d'intestins de termites du genre *Nasutitermes*. Les deux consortia enrichis présentent des rendements élevés de bioconversion de la paille de blé. La description des communautés bactériennes obtenues en fin d'enrichissement et l'analyse des dynamiques de population au cours de la bioconversion ont permis d'identifier de probables complémentarités entre taxons composant chaque consortium, sur la base des connaissances de leurs génomes. Cependant, seules des analyses fonctionnelles de métatranscriptomique ou de métagénomique pourraient confirmer ces hypothèses.

Ainsi, en combinant les approches de macro-cinétiques conventionnelles, d'écologie microbienne et de métagénomique et en s'appuyant sur trois consortia lignocellulolytiques stabilisés (deux obtenus précédemment sur paille de blé et un nouvellement enrichi sur résidus de maïs), ces travaux de thèse s'attachent à améliorer notre compréhension du fonctionnement de la bioconversion de la lignocellulose par des communautés bactériennes en bioréacteur. Dans

un second temps, ils s'intègrent dans un projet plus large d'optimisation des procédés de bioconversion, s'intéressant au prétraitement de résidus de maïs par extrusion-réactive et à la fermentation en voie sèche, permettant de limiter les besoins en eau. Les données et résultats obtenus sont organisés dans ce manuscrit en cinq chapitres. Le 1<sup>er</sup> chapitre a pour but de présenter au lecteur les bases et les principaux points clés permettant de comprendre le contexte de ces travaux et la méthodologie employée. Le chapitre 2 présente l'ensemble des procédures expérimentales utilisées dans les chapitres de résultats suivants, écrits sous la forme d'articles scientifiques en préparation. Le chapitre 3 présente les résultats de l'analyse métagénomique comparative de deux consortia bactériens, RWS et TWS, respectivement issus de rumen bovin et d'intestin de termite, précédemment enrichis au laboratoire sur paille de blé. Ces résultats ont permis de valider notre approche de métagénomique *shotgun* sans marquage appliquée aux bioréacteurs, et de confirmer *in situ* au niveau protéique, les hypothèses de complémentarité fonctionnelle entre les Bacteroidetes et les Firmicutes, précédemment formulées sur la base des résultats de diversité 16S. Le chapitre 4 présente une étude, à ce jour inédite, de l'impact de prétraitements modifiant la structure physico-chimique de la paille de blé sur le métagénomique du consortium RWS au cours de la bioconversion du substrat. Ces résultats ont amélioré nos connaissances sur l'impact des prétraitements et du substrat sur les communautés microbiennes, et nous ont renseigné sur les fonctions enzymatiques soumises à régulation chez les principaux taxons impliqués. Ces résultats ouvrent de nouvelles pistes pour piloter les consortia microbiens et pour améliorer les vitesses de bioconversion de la lignocellulose. Enfin, le chapitre 5 propose une approche d'écologie microbienne pour étudier l'enrichissement d'un nouveau consortium à partir de rumen bovin pour la bioconversion de résidus de maïs en acides gras volatils à faible teneur en solides. Sa stabilité et sa robustesse ont été éprouvées lors d'essais de cryoconservation et de mise à l'échelle. Une analyse longitudinale reliant la dynamique des populations et la production des AGV lors d'une cinétique de bioconversion ainsi que le fractionnement des populations bactériennes et de l'activité xylanase en fractions libres, faiblement liées ou liées au substrat tendent à montrer l'importance d'enrichir des taxons tardifs et peu abondants dans un consortium et à supposer leur rôle majeur dans la bioconversion de la lignocellulose en bioréacteur.



Ce travail s'inscrit dans le projet Hi-Solids (ANR-14-CE19-0013-01) et a été financé par l'Agence Nationale de la Recherche. Il a été réalisé au sein du Toulouse Biotechnology Institute (anciennement Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, LISBP), dans l'équipe Symbiose (EAD9). Il a fait l'objet d'une collaboration avec les équipes françaises de l'EAD1 du même laboratoire, de l'Unité Mixte de Recherche Ingénierie des Agropolymères et Technologies Emergentes (UMR IATE) à Montpellier, du Laboratoire de Biotechnologie de l'Environnement (LBE) à Narbonne, et de l'Unité Biopolymères Interactions Assemblages (BIA) à Nantes. Les analyses métagénomiques ont été conduites en collaboration avec le Helmholtz Centre of Environment à Leipzig (Allemagne) et avec l'équipe du Dr Bernard Henrissat qui gère la base de données CAZy au Laboratoire Architecture et Fonction des Macromolécules Biologiques (AFMB) à Marseille (France).

## I. Chapitre 1 – Etat de l’art

### A. La lignocellulose

La lignocellulose est le composant principal de la paroi cellulaire végétale et le biopolymère le plus abondant sur la Terre. La principale fonction de la lignocellulose est d’assurer les résistances mécanique et physique des cellules végétales. L’agriculture produit plus de  $2.10^{14}$  tonnes de biomasse lignocellulosique par an (Sawatdeenarunat et al., 2015), dont une grande partie est brûlée ou considérée comme déchets. Dans le contexte de réchauffement climatique et de réduction de l’utilisation des ressources fossiles, l’utilisation de cette biomasse pour la production d’énergie est devenue un enjeu majeur. Du fait d’une teneur en eau assez élevée, sa combustion n’est pas intéressante, d’autant qu’elle est riche en polysaccharides et composés aromatiques pouvant être valorisés sous forme de biocarburants, bioplastiques et synthons d’intérêt industriel. Néanmoins, les propriétés de résistance de la lignocellulose, bien qu’intéressantes pour les plantes en termes de défenses et de maintien de leur intégrité, la rendent difficilement hydrolysable pour en récupérer les sucres fermentescibles et est un facteur limitant pour son utilisation comme source de carbone renouvelable (Yang et al., 2015; Marriott et al., 2016).

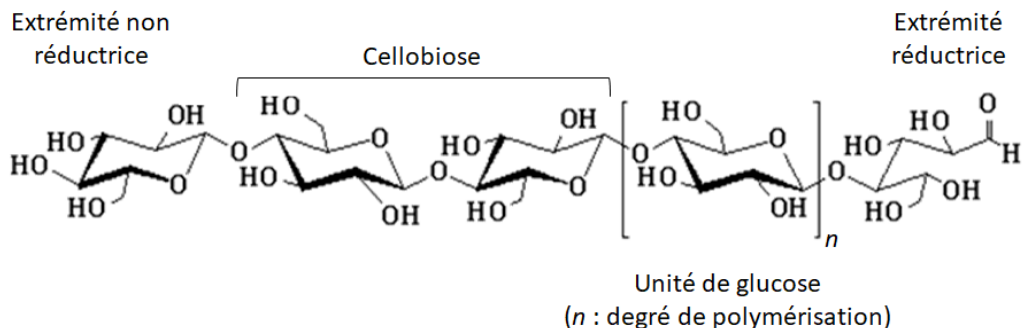
#### 1. Composition de la lignocellulose

La lignocellulose est une matrice hétérogène composée de trois fractions principales : la cellulose (35-50 % du poids sec), les hémicelluloses (20-35 %) et les lignines (5-30 %) (Lynd et al., 2002). Elle présente également des fractions minoritaires : des pectines, des cendres, des sucres solubles, des protéines et des minéraux (Vassilev et al., 2012). Le pourcentage de chaque fraction dépend de l’espèce, de la maturation, des tissus et des organes de la plante (racine, tige, feuille) et les conditions dans lesquelles la plante croît (Wilson, 1993).

##### a) La cellulose

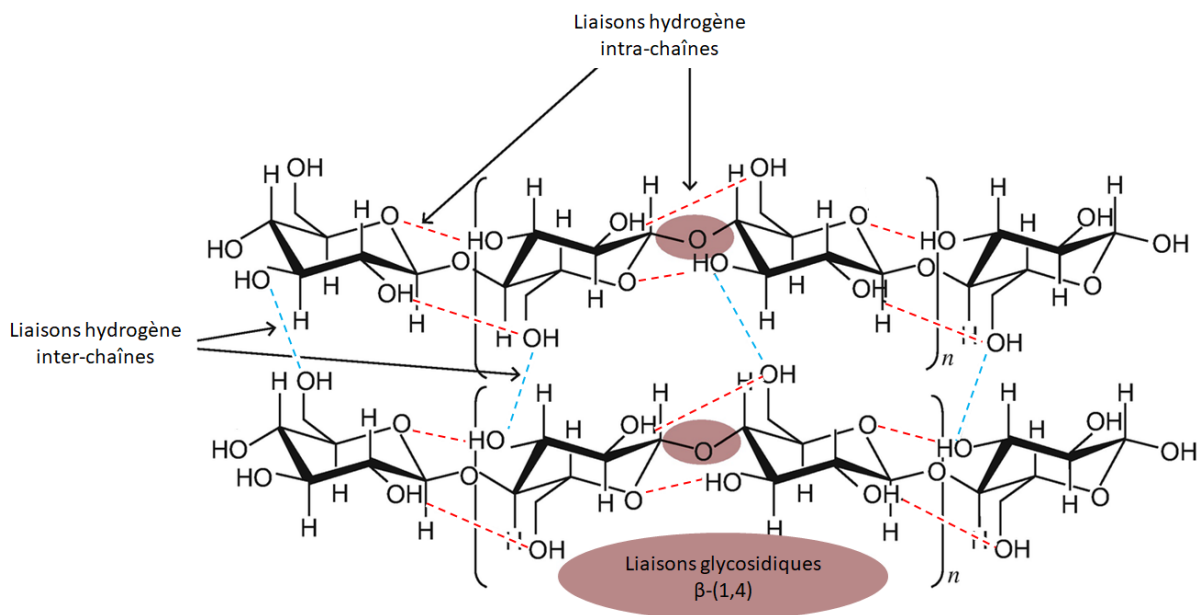
La cellulose est le composant majeur des parois végétales (35-50% du poids sec en moyenne) et est le biopolymère le plus abondant sur la planète avec une production totale estimée à  $10^{11}$ - $10^{12}$  tonnes par an (Lynd et al., 2002; Foroughi et al., 2021). La cellulose est un homopolymère insoluble, relativement stable, et linéaire dont la formule chimique est  $(C_5H_{10}O_5)_n$ .

La cellulose est constituée de molécules de  $\beta$ -D-glucose reliées par des liaisons  $\beta$ -(1,4) avec un degré de polymérisation  $n$  variant de 300 à 14 000 unités (Figure I-1).



**Figure I-1 : Structure moléculaire de la cellulose.**

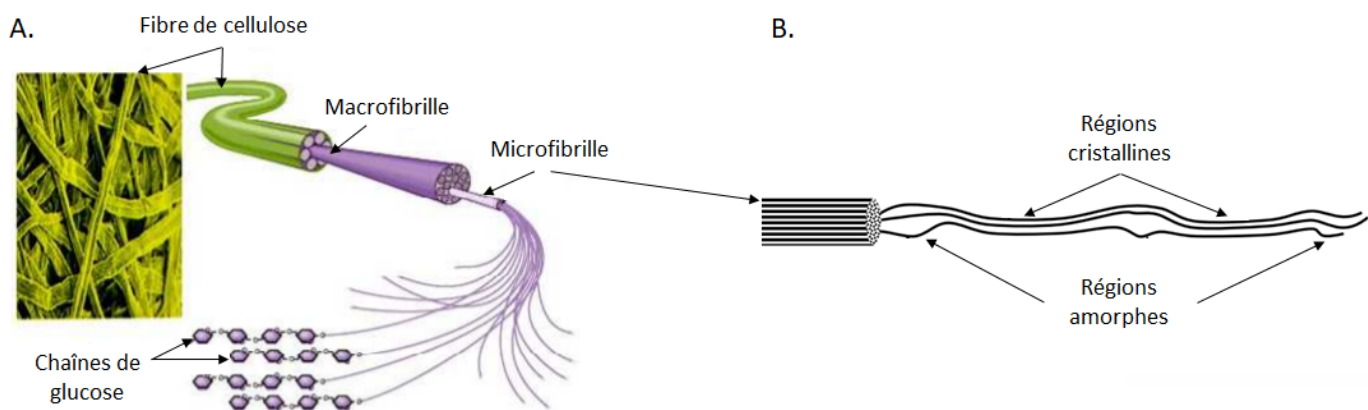
Les chaînes de cellulose sont associées entre elles de façon parallèle (Figure I-2) grâce aux liaisons hydrogènes intra- et inter- chaînes et aux forces de Van der Waals (O’Sullivan, 1997).



**Figure I-2 : Représentation schématique des liaisons glycosidiques  $\beta$ -(1,4) (violet) et des liaisons hydrogène intra-chaînes (rouges) et inter-chaînes (bleues) de la cellulose.**

Les chaînes de cellulose se regroupent ensemble en microfibrilles en formant des structures cristallines ordonnées, dont la taille peut varier de 2 nm à 20 nm (Figure I-3A). Les microfibrilles

de cellulose s’associent entre elles pour donner des structures ordonnées de taille plus importante appelées microfibrilles. Ces dernières peuvent également se regrouper en fibres de cellulose, constituant ainsi une structure stable et organisée (Bayer et al., 1998).



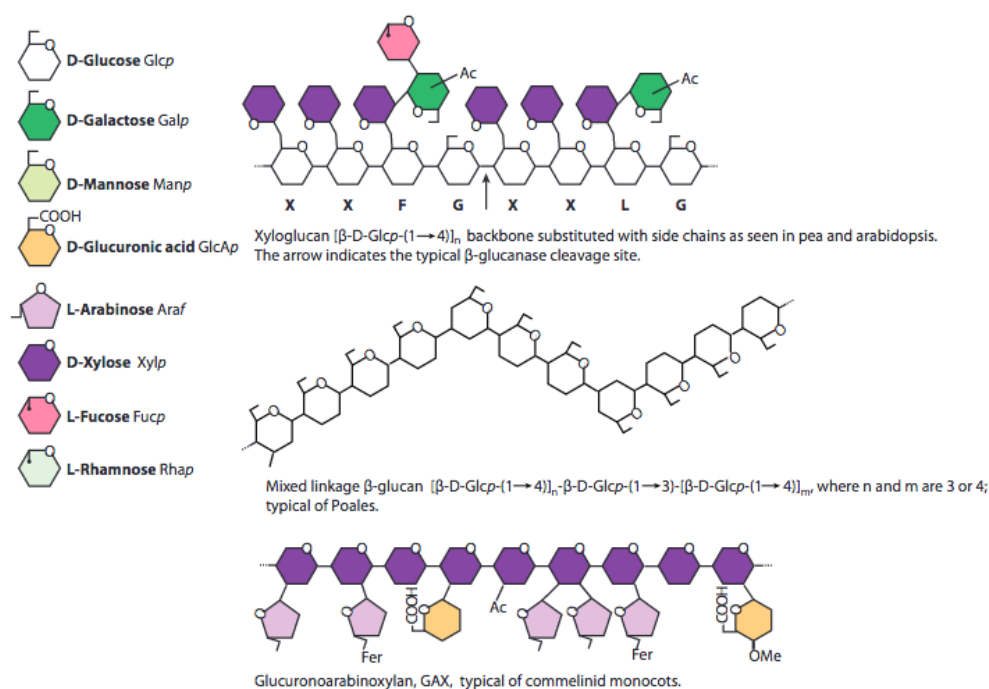
**Figure I-3 : Organisation des fibres de cellulose (A) et états structuraux d'une microfibrille de cellulose (B) avec des régions amorphes et cristallines.**

L’assemblage des chaînes de cellulose n’est pas homogène et présente deux états structuraux différents: la cellulose amorphe et la cellulose cristalline (Figure I-3B) (Larsson et al., 1997). Les régions amorphes présentent un assemblage non ordonné et aléatoire des chaînes de cellulose tandis que les régions cristallines sont caractérisées par des chaînes de cellulose liées de manière ordonnée par des liaisons hydrogène, formant alors un motif cristallin. Les régions cristallines confèrent à la lignocellulose un caractère difficilement biodégradable, peu soluble en solution aqueuse et gonflant en présence d’eau. La lignocellulose peut être caractérisée par son degré de cristallinité (ratio entre la quantité de cellulose cristalline et cellulose amorphe) à l’aide de techniques analytiques comme la cristallographie par rayons X ou la spectroscopie à infrarouge (O’Sullivan, 1997).

### ***b) Les hémicelluloses***

La composition des hémicelluloses est très variable et dépend essentiellement de l’espèce et du stade de développement des plantes (20-35% du poids sec). Les hémicelluloses sont des hétéropolysaccharides linéaires ou branchés, nommées en fonction de l’ose (pentose ou hexose) le plus présent dans le squelette principal (Figure I-4). On y retrouve les xylanes (enchaînement de

$\beta$ -D-xyloses reliés par des liaisons  $\beta$ -(1,4)), les mannanes (enchaînement de  $\beta$ -D-mannoses reliés par des liaisons  $\beta$ -(1,4)), les  $\beta$ -glucanes (enchaînement de  $\beta$ -D-glucoses reliés par des liaisons mixtes  $\beta$ -(1,4) et  $\beta$ -(1,3)), les arabinogalactanes (enchaînement de  $\alpha$ -D-galactoses et  $\alpha$ -L-arabinoses), et les xyloglucanes (enchaînement de  $\beta$ -D-xyloses et  $\beta$ -D-glucoses) (Gírio et al., 2010).



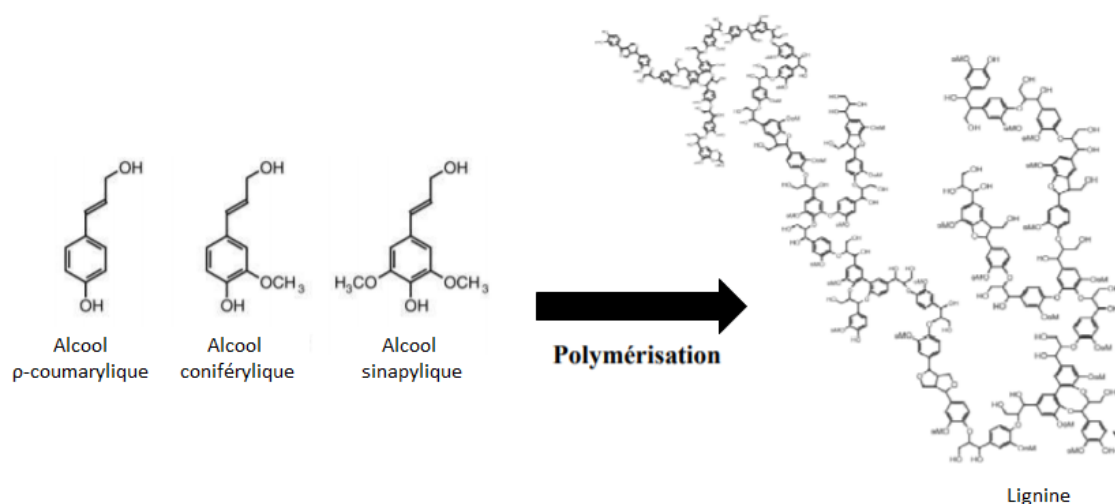
**Figure I-4 : Exemples de structures de trois hémicelluloses (xyloglucane,  $\beta$ -glucane, glucuronoarabinoxylane) retrouvées dans les parois végétales. D’après Scheller and Ulvskov, (2010).**

Les chaînes principales peuvent être ramifiées par des chaînes latérales constituées de divers motifs ( $\alpha$ -L-arabinose,  $\beta$ -D-xylose,  $\alpha$ -D-galactose,  $\alpha$ -L-fucose,  $\alpha$ -L-rhamnose, acide  $\alpha$ -D-glucuronique,  $\alpha$ -D-acide galacturonique,  $\alpha$ -D-4-O acide méthylglucuronique, acide férulique, acide p-coumarique ou groupements acétyles) (Pérez et al., 2002). Le degré de polymérisation dépend du tissu et de l’espèce végétale mais est moins important que celui de la cellulose, avec une moyenne de 160 unités. Par contraste avec la cellulose, les hémicelluloses sont plus facilement hydrolysables par voie enzymatique ou par voie chimique puisqu’une partie est soluble en milieux aqueux alcalin (Gírio et al., 2010).

Les xylanes sont un groupe important des hémicelluloses constituant les cellules végétales des céréales et des graminées (jusqu’à 30% de la composition, Vassilev et al., (2012)). Ils sont constitués de chaînes principales de  $\beta$ -(1,4)-xyloses pouvant être substitués par des chaînes latérales constituées de manière variable d’ $\alpha$ -L-arabinofuranoses, d’ $\alpha$ -D-galactoses, d’ $\alpha$ -D-acide glucuronique et d’acides férulique (Raven et al., 2014).

### c) La lignine

La lignine est une fraction représentant 5 à 20% du poids sec des tiges herbacées, 15 à 35% des tiges ligneuses, et jusqu’à 25 à 40% du bois, Vassilev et al., (2012)). La lignine est un polymère phénolique et ramifié de haut poids moléculaire. Elle résulte de la polymérisation tridimensionnelle de trois monolignols (Figure I-5) qui sont : l’alcool p-coumarylique ou sous unité H (p-hydroxyphényle), l’alcool coniférylique ou sous unité G (guaïacyle ou 4-hydroxyl-3-methoxyphenyl) et l’alcool sinapylique ou sous unité S (syringyle ou 4-hydroxy-3,5dimethoxyphenyl) (Figure I-5). La proportion de chaque monolignol varie en fonction de l’origine de la plante (Pandey and Kim, 2011). De par sa structure, la lignine assure un rôle de barrière physico-chimique contre les agents pathogènes.



**Figure I-5 : Structure moléculaire de la lignine résultant de la polymérisation tridimensionnelle de trois monolignols (l’alcool p-coumarylique, l’alcool coniférylique et l’alcool sinapylique).**

La lignine possède une structure amorphe, complexe et apparaît comme aléatoire et désorganisée (Vassilev et al., 2012). En raison de son caractère hydrophobe (Yang and Pan, 2016), la lignine représente un obstacle majeur pour l’hydrolyse des fibres de lignocellulose, point qui sera abordé dans la section I.A.3. Hydrolyse de la lignocellulose.

**d) Les fractions minoritaires**

La lignocellulose est composée d’autres fractions minoritaires (Vassilev et al., 2012): des pectines, des minéraux, des protéines, et des extractibles.

Les pectines possèdent un squelette principal formé d’acides galacturoniques hautement méthylés ou acétylés et reliés par des liaisons  $\alpha$ -(1,4). L’homogalacturonane (HGA) est la forme la plus simple des pectines avec une seule chaîne linéaire d’acides galacturoniques. Les pectines plus complexes sont les rhamnogalacturonanes de type I (RG-I) constitués d’un squelette d’acides galacturoniques et de rhamnoses, ramifiés par des résidus galactoses et arabinoses et les rhamnogalacturonanes de type II (RG-II) constitués d’un squelette d’acides galacturoniques ramifiés par des résidus rhamnoses, galactoses, arabinoses et d’autres résidus saccharides (Figure I-6, Harholt et al., 2010). La teneur en pectine varie en fonction du tissu végétal, du type et de l’âge de la plante. Les pectines sont assez peu mentionnées dans les compositions de différentes biomasses retrouvées dans la littérature.

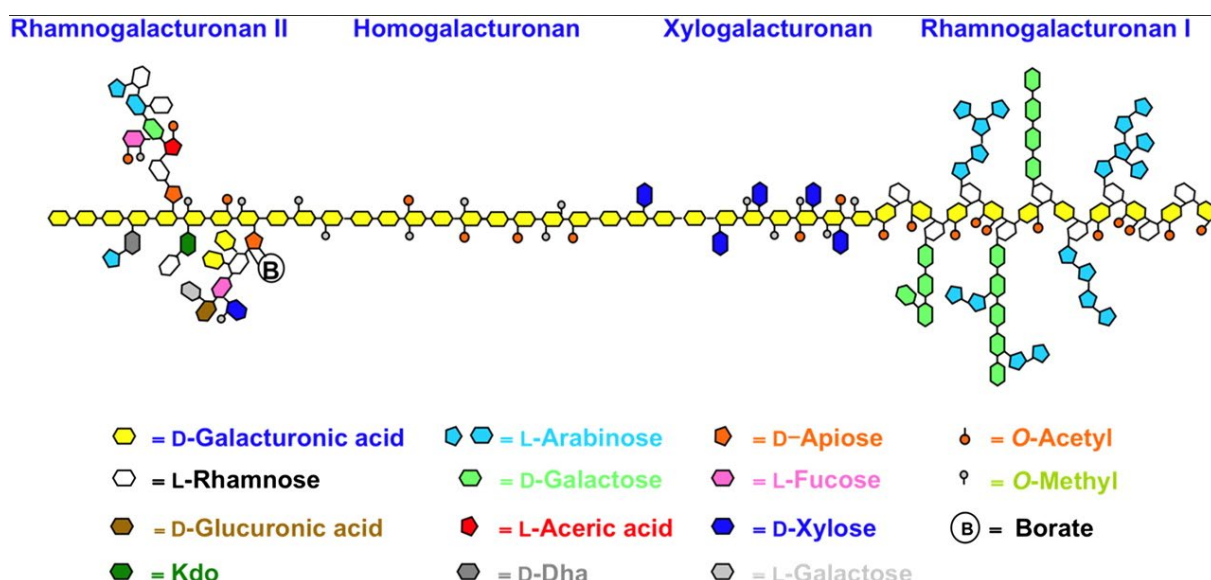


Figure I-6 : Représentation schématique des pectines. D’après Harholt et al., (2010).

Des éléments minéraux sont récupérables dans les cendres de la biomasse lignocellulosique et correspondent en général à 5 à 10% du poids sec de la biomasse, mais peut atteindre jusqu’à 15% du poids sec dans le cas de la paille de riz. On y retrouve une large quantité de silicium sous forme de silice, du calcium, du potassium, du soufre, du magnésium, du chlore, du phosphore, et du sodium. Une faible fraction de la biomasse lignocellulosique est également constituée de protéines (1 à 11% du poids sec en fonction de l’origine de la plante et du stade de développement). Les protéines peuvent être des enzymes mais aussi des protéines structurales. D’autres composés, appelés les extractibles, représentent aussi quelques pourcents du poids sec de la biomasse. Il s’agit de composés carbonés non structuraux tels que des acides gras, des acides résiniques, des stérols, des cires, des stérides, et des acides phénoliques. Ils sont extraits lors de la solubilisation de la biomasse par un solvant.

## 2. Structure de la lignocellulose

L’assemblage des différentes fractions de la lignocellulose est complexe et se fait par l’intermédiaire de liaisons covalentes. Comme nous l’avons vu précédemment, les fibres de cellulose sont stabilisées par des liaisons hydrogène intra- et inter- chaînes, résistantes à l’hydrolyse. Elles sont reliées aux hémicelluloses via des liaisons covalentes ou hydrogènes, mais aussi à la lignine par des liaisons fortes de type esters, leur conférant une résistance à l’hydrolyse enzymatique (Chen, 2014). Ainsi, la résistance des parois végétales s’explique par cette structure renforcée de la cellulose par les hémicelluloses et la lignine (Figure I-7).

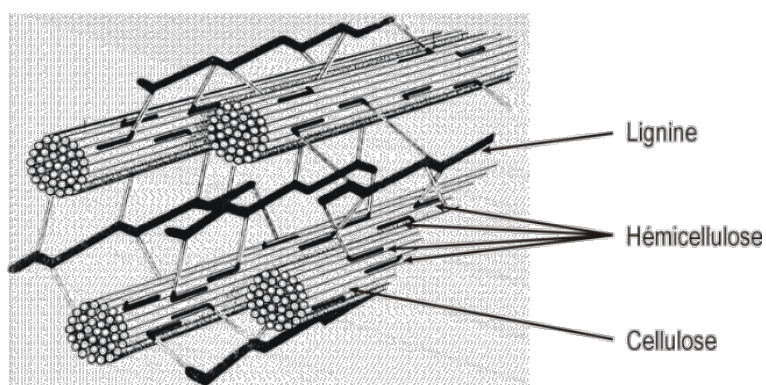


Figure I-7 : Représentation tridimensionnelle de la structure de la lignocellulose.



### 3. Hydrolyse de la lignocellulose

*Dans ce chapitre, bien que la dégradation de la lignocellulose puisse être également réalisée en aérobiose par des champignons et/ou des bactéries, seules les activités bactériennes anaérobies, correspondant aux populations observées dans les conditions de ces travaux, seront abordées.*

En raison du caractère insoluble et récalcitrant de la lignocellulose, les micro-organismes ont adapté des stratégies singulières pour hydrolyser la lignocellulose en produisant une large variété d’enzymes qui agissent en complémentarité et en synergie (Van Dyk and Pletschke, 2012; Cortes-Totalpa et al., 2017; Malgas et al., 2017). L’hydrolyse bactérienne de la lignocellulose se produit le plus souvent de manière exocellulaire, soit par des enzymes extracellulaires libres et solubles (Bayer et al., 2013), soit par des cellulosomes en association avec la paroi cellulaire (Bayer et al., 1998; Schwarz, 2001; Doi and Kosugi, 2004). Ces enzymes sont appelées les *Carbohydrate Active enZymes* (CAZymes) et sont référencées dans la base de données CAZy (<http://www.cazy.org>, Lombard et al., (2014)). Cette base de données classe les CAZymes en 5 catégories :

- les glycosides hydrolases ou glycosidases (GH) ;
- les carbohydrates estérases (CE) ;
- les glycosyltransférases (GT) ;
- les polysaccharides lyases (PL) ;
- et les enzymes dites à activités auxiliaires (AA).

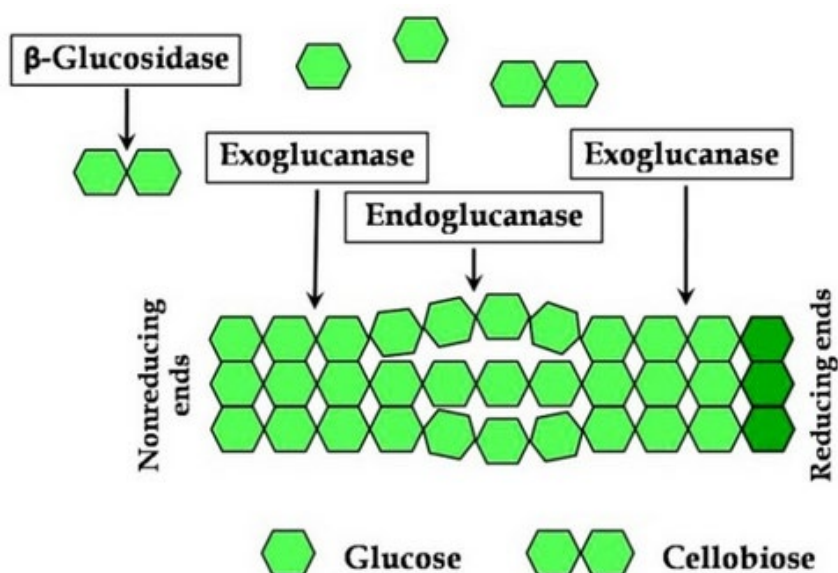
Chaque catégorie est divisée en familles au sein desquelles les enzymes possèdent des structures tertiaires et des mécanismes d’actions catalytiques similaires. A l’heure actuelle (mise à jour du 6 août 2021), la base de données CAZy regroupe 171 familles GH, 114 familles GT, 42 familles PL, 19 familles CE, et 17 familles AA. De plus, cinq familles GH sont également divisées en sous-familles en raison de leur polyspécificité : les GH5 (Aspeborg et al., 2012), GH13 (Stam et al., 2006), GH16 (Viborg et al., 2019), GH30 (St John et al., 2010) et GH43 (Mewis et al., 2016). A cette base de données, il faut aussi ajouter les domaines de liaison au substrat (appelés en anglais *Carbohydrate-Binding Modules*, CBM) qui permettent à certaines familles CAZymes de cibler et de se lier à leur substrat et de faciliter l’activité du site catalytique (Linder and Teeri, 1997). Il existe actuellement 88 familles de CBM dans la base de données CAZy (mise à jour du 6 août 2021).

Trois grandes catégories d’enzymes dégradent la lignocellulose et sont regroupées selon le substrat qu’elles catalysent : les cellulases, les hémicellulases, les ligninases.

**a) Les cellulases**

Quatre groupes d’enzymes sont capables de dégrader la cellulose et sont classés dans les familles des Glycoside hydrolases (GH) dans la base de données CAZy, parmi entre autres les familles GH1, 3, 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 74, 124 (Lynd et al., 2002; Bayer et al., 2013; López-Mondéjar et al., 2016a).

Les exo- $\beta$ -(1,4)-glucanases (EC 3.2.1.91) et les cellobiohydrolases (EC 3.2.1.176 et EC 3.2.1.74) hydrolysent les extrémités réductrices libres des chaînes de cellulose en libérant des monomères de glucose ou du cellobiose (dimère de glucose). Les endo- $\beta$ -(1,4)-glucanases (EC 3.2.1.4) hydrolysent aléatoirement les liaisons osidiques à l’intérieur des chaînes de cellulose en générant des oligosaccharides de différentes tailles. Les  $\beta$ -glucosidases (EC 3.2.1.21) hydrolysent les oligomères de glucose et les dimères de cellobiose en monomères de glucose solubles (Figure I-8 et Tableau I-1).



**Figure I-8 : Représentation des activités enzymatiques responsables de la dégradation de la cellulose. D’après Ratanakhanokchai et al., (2013).**

**Tableau I-1 : Descriptif des activités enzymatiques pour l'hydrolyse de la cellulose (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).**

Fraction	Substrat	Activité enzymatique	Liaisons hydrolysées	Produits issus de l'hydrolyse enzymatique
Cellulose	Cellulose (extrémités)	Exo- $\beta$ -(1,4)-glucanase	$\beta$ -(1,4)-glucose	Glucose
		Cellobiohydrolases	$\beta$ -(1,4)-glucose	Cellobiose
	Cellulose (intra-chaîne)	Endo- $\beta$ -(1,4)-glucanase	$\beta$ -(1,4)-glucose	Oligosaccharides (taille variable)
	Cellobiose	$\beta$ -glucosidase	$\beta$ -(1,4)-glucose	Glucose

La plupart des cellulases agissent préférentiellement sur la cellulose amorphe, mais les cellobiohydrolases et les exo- $\beta$ -(1,4)-glucanases sont capables de cibler préférentiellement et d'hydrolyser efficacement la cellulose cristalline (Carrard et al., 2000).

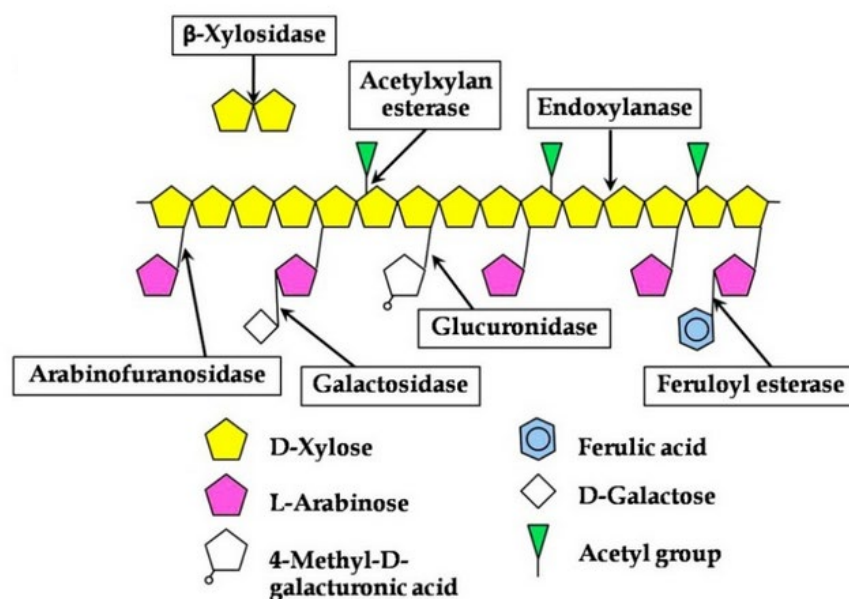
Une nouvelle classe d'enzymes, appelées les LPMO (en anglais, *Lytic Polysaccharide Monooxygenases*) existe depuis peu (Hemsworth et al., 2015) et a permis de reclassifier certaines GH en AA. Cette classe regroupe des enzymes oxydatives des familles AA9 (EC 1.14.99.54 et EC 1.14.99.56), AA10 (EC 1.14.99.54, EC 1.14.99.53 et EC 1.14.99.56), AA11 et AA13 (EC 1.14.99.55), AA14 (EC 1.14.99.-), AA15 (EC 1.14.99.54 et EC 1.14.99.53) et AA16 (EC 1.14.99.54) dans la base de données CAZy. Ces enzymes créent des brèches dans la structure cristalline des fibres de cellulose (Villares Garicochea et al., 2017).

Les bactéries cellulolytiques anaérobies les plus connues et étudiées appartiennent aux genres *Acetivibrio*, *Bacillus*, *Bacteroides*, *Butyrivibrio*, *Caldicellulosiruptor*, *Cellulomonas*, *Clostridium*, *Erwinia*, *Eubacterium*, *Fibrobacter*, *Ruminococcus*, *Spirochaeta*, *Streptomyces*, *Thermomonospora* (Lynd et al., 2002).

### **b) Les hémicelluloses**

La composition chimique plus hétérogène des hémicelluloses et l'organisation structurale en chaînes latérales courtes, en font un polymère plus facilement hydrolysable que la cellulose.

Cependant, l’hydrolyse complète des hémicelluloses nécessite une grande diversité de CAZymes (Figure I-9 et Tableau I-2), entre autres les glycosides hydrolases (telles que les familles GH2, 5, 7, 8, 10, 11, 16, 26, 30, 31, 36, 39, 42, 43, 51, 53, 67, 74, 92, 95, 115) et les carbohydrates esterases (CE) (López-Mondéjar et al., 2016a; Thapa et al., 2020).



**Figure I-9 : Représentation des activités enzymatiques responsables de la dégradation de l'arabinoxylane. D'après Ratanakhanokchai et al., (2013).**

Les hémicellulases sont classées en 3 catégories et se différencient en fonction du substrat hydrolysé :

- les endohémicellulases clivent de façon aléatoire les chaînes principales des hémicelluloses comme le xylane. Ainsi, la chaîne principale de xylane peut être attaquée à l’intérieur de la chaîne par des endo- $\beta$ -(1,4)-xylanases (EC 3.2.1.8) et les extrémités de chaînes peuvent libérer du xylobiose par l’action des exo- $\beta$ -(1,4)-xylanases (EC 3.2.1.37). Les endomannanases (EC 3.2.1.78) et les xyloglucanases (EC 3.2.1.151) dégradent préférentiellement le mannane et le xyloglucane respectivement.
- Les enzymes de débranchement (*debranching enzymes*) clivent les motifs accessoires pour permettre l’accès au squelette de la chaîne principale. Les plus fréquentes sont les  $\alpha$ -L-

arabinofuranosidases (EC 3.2.1.55), les acétyl xylan estérases (EC 3.1.1.72) ou les  $\alpha$ -D-glucuronidases (EC 3.2.1.1).

- Les exo-hémicellulases produisent des monomères de sucre à partir des courtes chaînes issues de l’hydrolyse des endohémicellulases, comme les  $\beta$ -xyloxydases (EC 3.2.1.37) qui engendrent deux unités de xylose à partir d’une unité de xylobiose.

**Tableau I-2 : Descriptif des principales activités enzymatiques existantes pour l'hydrolyse des hémicelluloses (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).**

<b>Fraction</b>	<b>Substrat</b>	<b>Activité enzymatique</b>	<b>Liaisons hydrolysées</b>	<b>Produits issus de l'hydrolyse enzymatique</b>
Hémicelluloses	Xylane	Endo- $\beta$ -(1,4)-xylanase	$\beta$ -(1,4)-xylose	Oligomères de xylose
	Xylane	Exo- $\beta$ -(1,4)-xylanase	$\beta$ -(1,4)-xylose	Xylobiose
	Xylobiose	$\beta$ -(1,4)-xylosidase	$\beta$ -(1,4)-xylose	Monomères de xylose
	Arabinoxylane	$\alpha$ -L-arabinofuranosidase	$\alpha$ -1,3-xylose	Arabinoxylane et arabinose
	Glucuronoxylane	$\alpha$ -glucuronidase	$\alpha$ -1,3 ou 1,2 acide glucuronique	Glucuronoxylane et acide glucuronique
	Acétyl xylane	O-acétyl xylan estérase	Acétyl ester	Acétyl xylane et acétate
	$\beta$ -glucane	$\beta$ -1,4-glucanase	$\beta$ -1,4-glucose	Oligomères de $\beta$ -glucane

### **c) Les enzymes lignolytiques**

De par sa structure aléatoire et l’absence de motif répété, la lignine constitue un polymère très résistant à la dégradation enzymatique. La complexité structurale des lignines, leurs hauts poids moléculaires et leur insolubilité dans l’eau rendent leur hydrolyse très difficile. Elle est néanmoins réalisée par des enzymes extracellulaires oxydatives, telles que les laccases, les peroxydases ou les enzymes accessoires, produites majoritairement par des champignons

lignolytiques (Bugg et al., 2011; López-Mondéjar et al., 2019). Ces enzymes sont toutes regroupées dans certaines familles des AA au sein de la base de données CAZy (familles AA1, AA2, AA4, AA6).

Les laccases (EC 1.10.3.2) sont des oxydoréductases dont l’action catalytique repose sur le couplage entre la réduction du dioxygène en eau et l’oxydation de substrats comme les phénols ou les monolignols. Les peroxydases à hème sont capables d’oxyder des substrats aromatiques phénoliques ou non phénoliques en réagissant avec un équivalent de peroxyde d’hydrogène. Il s’agit d’une réaction d’oxydoréduction à deux électrons au cours de laquelle le peroxyde d’hydrogène est réduit en eau pendant que l’enzyme est oxydée. Il existe quatre types de peroxydases : les lignines peroxydases (LiP, EC 1.11.1.14), les manganèse peroxydases (MnP, EC 1.11.1.13), les peroxydases versatiles (VP, EC 1.11.1.16) et les peroxydases décolorantes de pigments (DyP, EC 1.11.1.19). Les enzymes accessoires ne sont pas directement impliquées dans la dégradation de la lignine mais permettent de réduire les composés dérivés de la lignine en aval de l’action des ligninases. Ce sont principalement des oxydases permettant la production de peroxyde d’hydrogène dont sont dépendantes les peroxydases à hème. Ainsi, l’oxygène semble essentiel pour le catabolisme des composés aromatiques de par son implication dans les réactions d’oxydoréduction. Dans les environnements anaérobies, les bactéries ne sont pas capables de dégrader la lignine sous sa forme native mais certaines molécules dérivées de la lignine peuvent être converties en méthane et en dioxyde de carbone (Grbić-Galić, 1983).

La dégradation de la lignine est principalement réalisée par des champignons lignolytiques (pourriture blanche et pourriture brune) et des bactéries aérobies issus des Phyla Actinobacteria et Proteobacteria mais quelques familles comme les *Enterobacteriaceae*, *Aeromonadaceae*, *Veillonellaceae*, *Rhodospirillaceae*, *Clostridiaceae*, *Bacillaceae*, *Eubacteriaceae*, *Nocardiaceae* ont été identifiées pour dépolymériser la lignine en anaérobiose (Xu et al., 2018; Khan and Ahring, 2019).

### **d) Autres enzymes**

En raison de leurs structures complexes, l’hydrolyse des pectines implique plusieurs catégories d’enzymes de la base de données CAZy. On y retrouve principalement les GH (GH28, 78, 88, 105), les CE (CE8, 12) et les PL (PL1, 2, 3, 4, 9, 10, 11) (Schols et al., 2009; López-Mondéjar et al., 2019). Les pectines sont hydrolysées par les polygalacturonases (EC 3.2.1.15) qui hydrolysent le squelette principal d’acide galacturonique, les pectines méthyl estérases (EC

3.1.1.11) qui clivent les liaisons esters entre le squelette principal et les ramifications, et les pectines lyases (EC 4.2.2.10) qui hydrolysent le squelette principal estérifié (Tableau I-3).

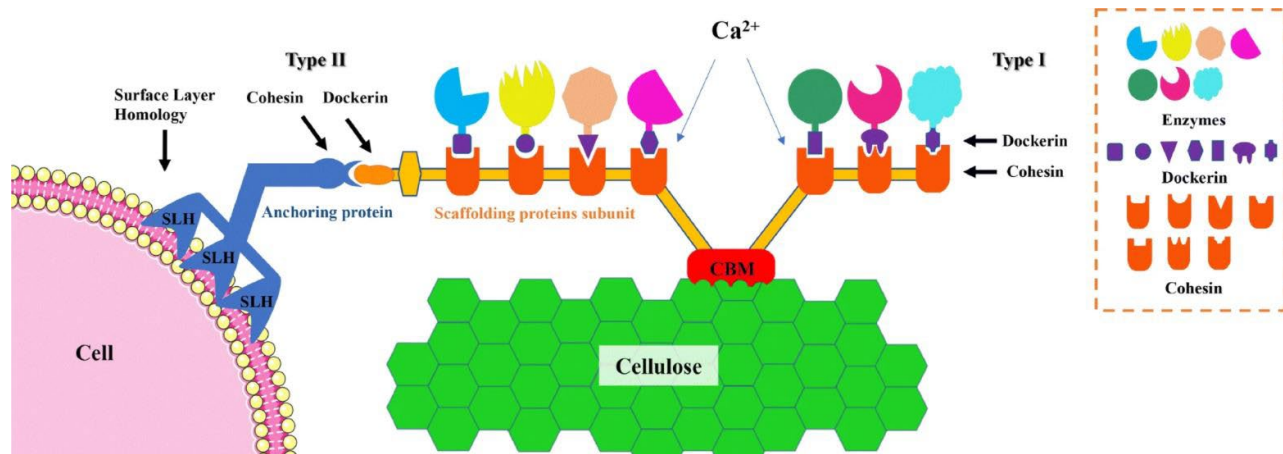
**Tableau I-3 : Descriptif des principales activités enzymatiques existantes pour l'hydrolyse des pectines (substrat hydrolysé, liaisons hydrolysées et produits issus de l'hydrolyse enzymatique).**

Fraction	Substrat	Activité enzymatique	Liaisons hydrolysées	Produits issus de l'hydrolyse enzymatique
Pectines	Polygalacturonane	Polygalacturonase	$\alpha$ -1,4 galacturonide	Acide polygalacturonique
	Pectines	Pectine méthyl estérase	Méthyl ester	Méthanol et polygalacturonane
	Pectines	Polygalacturonate lyase	$\alpha$ -1,4 galacturonide	Acide polygalacturonique

### e) Les cellulosomes

La plupart des bactéries hydrolytiques en anaérobie présentent des systèmes cellulolytiques multi-enzymatiques complexes appelés cellulosomes (Figure I-10). Les cellulosomes sont des protubérances à la surface des cellules bactériennes qui permettent une dégradation efficace du substrat. Ils sont constitués d’une sous-unité structurale appelée “*Scaffoldin*” sur laquelle s’assemblent, grâce aux domaines cohésine et dockerine de type I, des modules enzymatiques et des modules de liaison aux carbohydrates (*Carbohydrate Binding Module*, CBM), qui sont positionnés de façon à créer une synergie entre les unités catalytiques. En effet, les cellulosomes présentent une activité plus forte que la somme des activités des enzymes isolées en raison de cette synergie (Munir and Levin, 2016). Les domaines SLH (*Surface-layer homology*) permettent d’ancrer les cellulosomes sur les surfaces bactériennes grâce aux domaines cohésine et dockerine de type II.





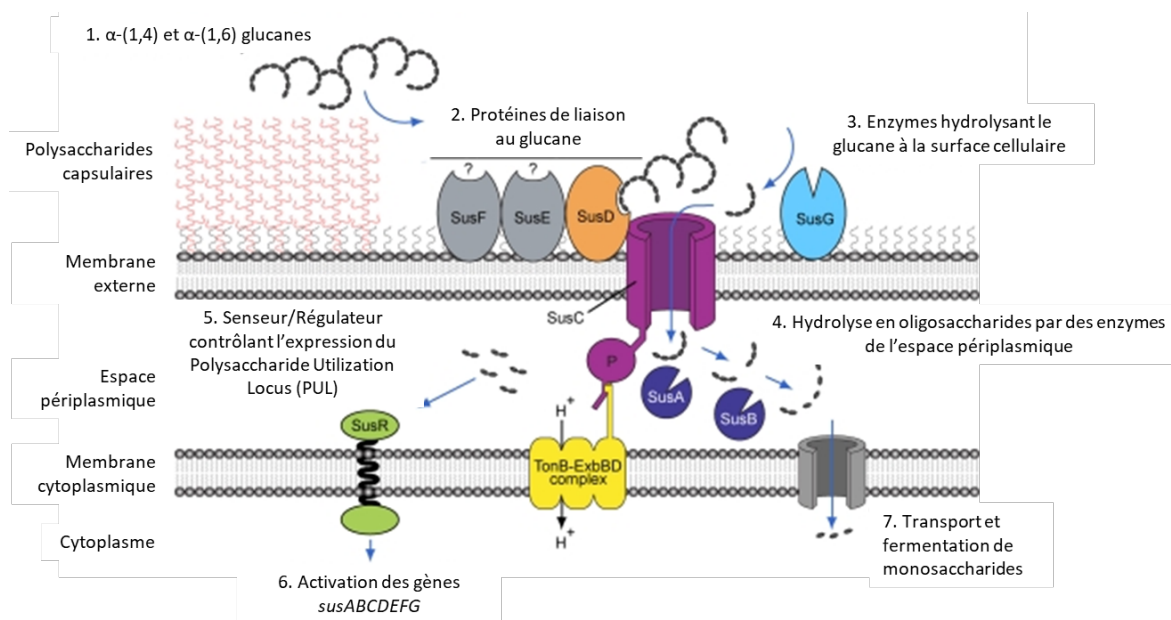
**Figure I-10 : Représentation des différents modules du cellulosome de la bactérie anaérobie *Clostridium cellulovorans*.** Les modules enzymatiques, colorés différemment pour indiquer la diversité enzymatique, sont assemblés sur la sous unité structurale appelée *scaffoldin* grâce aux domaines cohésines et dockerines de type I. Le cellulosome contient également des domaines de liaison au substrat (appelés *Carbohydrate Binding Module*, CBM) permettant au cellulosome d’adhérer au substrat. Le cellulosome se fixe à la surface de la cellule bactérienne grâce aux domaines cohésines et dockerines de type II. D’après Jiang et al., (2021).

Les cellulosomes ont été rencontrés chez plusieurs bactéries anaérobies parmi les genres *Acetivibrio*, *Clostridium* et *Ruminococcus* (Phylum Firmicutes), et *Bacteroides* (Phylum Bacteroidetes) (Ding et al., 2001; Xu et al., 2003; Pinheiro et al., 2009; Zhivin et al., 2017).

### f) Les Polysaccharide Utilization Loci (PULs)

Les PULs (*Polysaccharide Utilization Loci*) sont des loci génétiques, c’est-à-dire des suites de gènes co-régulés qui codent des systèmes multi-protéiques (entre autres des glycosides hydrolases, des modules de liaisons aux carbohydrates et des transporteurs de polysaccharides, ainsi qu’un système de régulation spécifique) qui sont associés à la membrane et sont dédiés spécifiquement à un type de polysaccharide. Ils servent à coder des glycosides hydrolases pour hydrolyser efficacement les fractions de la lignocellulose (cellulose, hémicelluloses, pectines) et à l’internalisation du produit d’hydrolyse grâce à un transporteur. Les PULs ont jusqu’ici été décrits exclusivement dans le génome des bactéries anaérobies du phylum *Bacteroidetes*. L’analyse génomique de l’ordre des *Bacteroidales* montre une grande diversité du nombre et des fonctions des PULs entre espèces et même entre souches, avec une forte prévalence dans le genre *Bacteroides* (Naas et al., 2014; Mackenzie et al., 2015). Les PULs sont nommés en fonction du polysaccharide à hydrolyser. Ainsi, chez *Bacteroides thetaiotaomicron*, bactérie amylolytique

présente dans le système digestif humain, le système multiprotéique associé à la membrane et impliqué dans la dégradation de l’amidon a été nommé Sus pour *Starch utilization system*. Son organisation a servi de base pour la définition des PULs (Figure I-11 ; Bjursell et al., 2006).



**Figure I-11 : Représentation du système Sus (*Starch utilization system*) chez *Bacteroides thetaioatomicron* pour l'hydrolyse du glucane. D'après Martens et al., (2009).**

Des études ultérieures ont permis de découvrir des dérivés du système Sus, chez d'autres représentants du phylum Bacteroidetes provenant de divers écosystèmes. Un système analogue pour le xylan a été appelé « Xus » (*Xylan Utilization System*) chez les *Bacteroidetes* xylanolytiques (Zhang et al., 2014) tandis qu'un PUL dédié à la dégradation des galactomannanes a été détecté chez *Bacteroides ovatus* (Bågenholm et al., 2017).

### **g) Les facteurs limitant l'hydrolyse**

Dans la nature, en condition anaérobie, entre 5 et 10% de la cellulose seulement est dégradée (Pérez et al., 2002). Plusieurs propriétés de la lignocellulose limitent son hydrolyse. Tout d'abord, la lignine est reconnue pour être le 1<sup>er</sup> facteur limitant en raison de son hydrophobicité et de ses liaisons à la cellulose et aux hémicelluloses, qui sont reconnues pour être résistantes à l'hydrolyse chimique et enzymatique (Yang and Pan, 2016). Les effets négatifs de la lignine ou ses dérivés sur la bioconversion de la lignocellulose incluent également une interférence avec les enzymes

cellulolytiques qui s’associent de manière non productive aux complexes lignine-carbohydrates et la libération de composés toxiques pour les micro-organismes (Kumar et al., 2009b). Néanmoins, la récalcitrance de la lignocellulose n’est pas uniquement due à sa teneur en lignine, mais est liée également à ses propriétés structurelles telles que la composition en lignine et en hémicelluloses, le degré de polymérisation, la porosité du substrat, la surface accessible aux enzymes hydrolytiques, et le pourcentage de cristallinité de la fraction cellulosique (McMillan, 1994). En effet, les régions amorphes de la cellulose présentent des niveaux d’hydrolyse 70% plus élevés que la fraction cristalline (Pérez et al., 2002), traduisant le fait que plus le degré de cristallinité de la biomasse lignocellulosique est faible, plus son hydrolyse est élevée (Laureano-Perez et al., 2005; Xu et al., 2012; Barakat et al., 2014a).

### **B. La bioraffinerie lignocellulosique**

La bioraffinerie est définie comme le processus de transformation durable de la biomasse en bioénergie (biocarburant, électricité, biogaz) et en bioproduits à haute valeur ajoutée destinés à la conception de matériaux biosourcés ou de produits pour l’industrie agroalimentaire (Kim and Dale, 2004; Kobayashi et al., 2012; Anwar et al., 2014). La lignocellulose, polymère le plus abondant sur terre, constitue une source de carbone renouvelable très attractive pour la bioraffinerie.

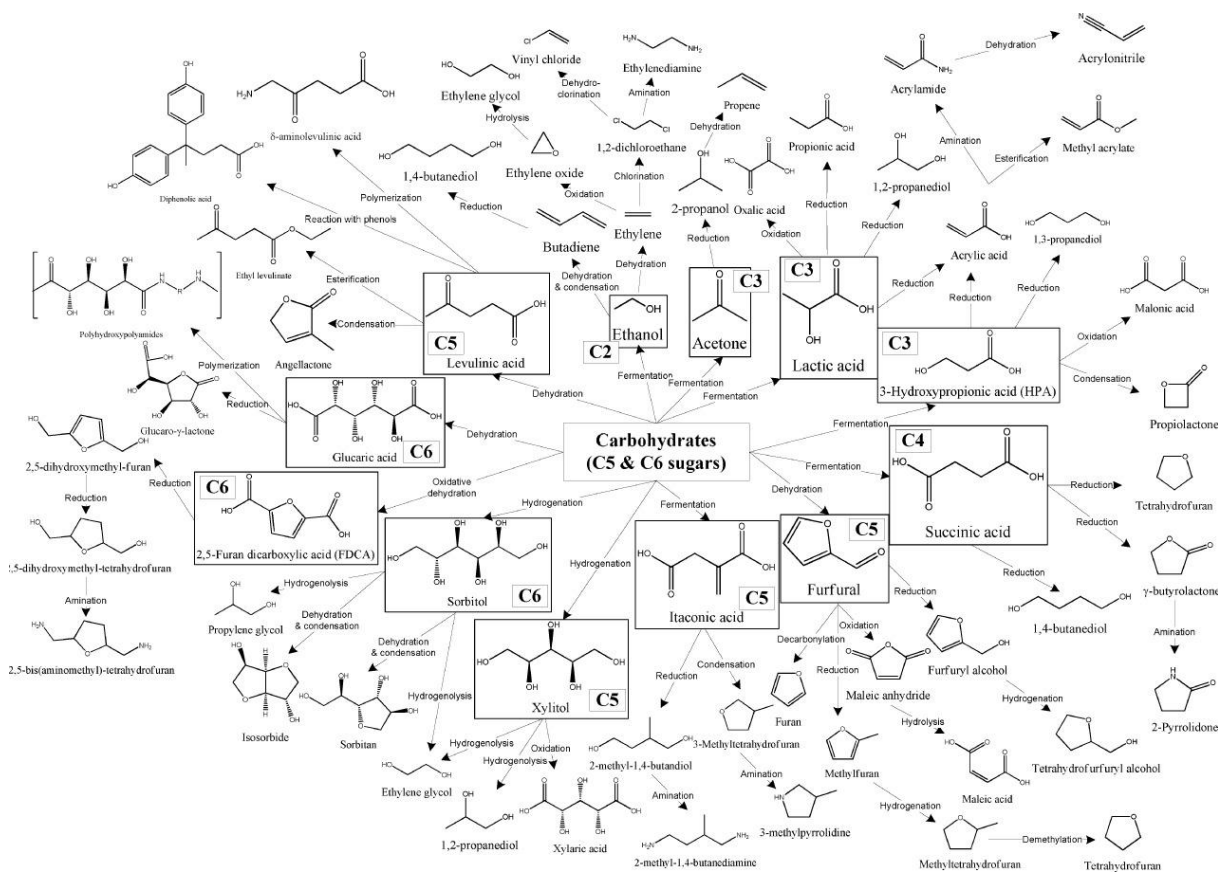
La première génération des bioraffineries lignocellulosiques est basée sur la transformation de la biomasse issue des cultures destinées à l’alimentation. La conversion des sucres fermentescibles (glucose, saccharose, etc) provenant des cultures amylacées et d’autres cultures (cannes à sucres, céréales, betteraves) génère essentiellement des biocarburants (bioéthanol et biodiesel) par fermentation alcoolique. Cependant, ces biocarburants de 1<sup>ère</sup> génération sont controversés puisque l’utilisation de ces cultures rentre en compétition directe avec l’alimentation humaine, épuise les sols et entraîne une importante consommation de pesticides pour augmenter les rendements de production (Martin, 2010).

La deuxième génération des bioraffineries lignocellulosiques est plutôt axée sur la transformation des résidus lignocellulosiques agricoles et forestiers. Comme la biomasse lignocellulosique est une matière première très abondante, moins chère que les cultures amylacées et ne rentrant pas en

compétition avec l’alimentation, la valorisation de ces résidus constitue un enjeu environnemental et économique clé pour une production renouvelable d’énergie, de biocarburants ou de composés chimiques. Les procédés de bioraffinerie de deuxième génération incluent l’hydrolyse de la lignocellulose en plus petites molécules solubles, mais aussi de très nombreux procédés de transformation de ces molécules en productions d’intérêt, qui doivent eux aussi satisfaire des exigences de rendement et de rentabilité (Naik et al., 2010).

### **1. De la lignocellulose aux carboxylates**

L’hydrolyse de la cellulose et des hémicelluloses conduit à la formation de monosaccharides en C<sub>5</sub> (xylose, arabinose) et en C<sub>6</sub> (glucose, mannose, galactose, rhamnose). Ces monosaccharides contenus dans l’hydrolysats peuvent ensuite être fermentés par l’action de souches microbiennes. La production de ces sucres fermentescibles issus de la biomasse lignocellulosique constitue ce qu’on appelle la « plateforme des sucres » et est utilisée principalement pour la production d’éthanol par une fermentation secondaire (Naik et al., 2010). La biomasse lignocellulosique offre une possibilité de production de plus de 200 molécules carbonées à haute valeur ajoutée (Figure I-12, Cherubini and Strømman, 2011). Cette voie de valorisation reste néanmoins un challenge puisqu’il faut parvenir au meilleur rendement en un produit souhaité, à l’extraire du moût de fermentation, le purifier et le transformer, le tout au plus faible coût économique, énergétique et environnemental.



**Figure I-12 : Schéma des composés chimiques et leurs dérivés obtenus à partir des carbohydrates (C<sub>5</sub>-C<sub>6</sub>) présents dans la lignocellulose. D’après Cherubini and Strømman, (2011).**

Dans ce manuscrit, nous nous intéresserons à la production d’acides gras volatils (AGV) à partir de la biomasse lignocellulosique. Les acides gras volatils sont des acides organiques à chaîne courte avec un squelette de 2 à 6 atomes de carbone: acétate (C<sub>2</sub>), propionate (C<sub>3</sub>), butyrate et isobutyrate (C<sub>4</sub>), valérate et isovalérate (C<sub>5</sub>), et hexanoate (C<sub>6</sub>). Ces acides organiques sont produits lors des étapes d’hydrolyse et d’acidogénèse de la digestion anaérobie, processus décrit dans la section suivante, et peuvent être consommés dans les phases d’acétogénèse et de méthanogénèse. Les acides gras volatils formés servent de précurseurs à de nombreux composés à haute valeur ajoutée d’intérêt industriel (bioplastiques et biocarburants) (Kim et al., 2013), lorsqu’ils sont séparés du moût de fermentation ou peuvent servir de substrat à la méthanogénèse.

### **a) La digestion anaérobie**

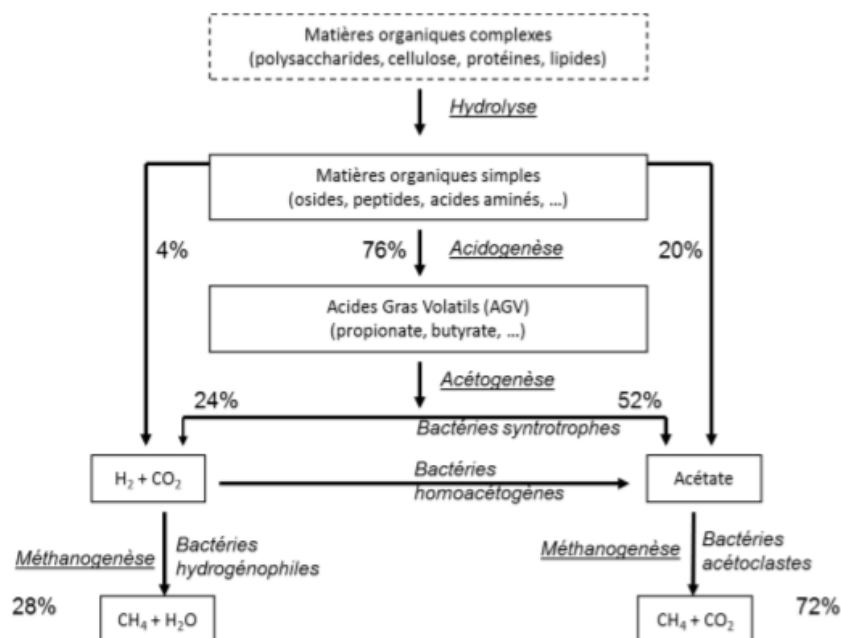
La digestion anaérobie (appelée également méthanisation), se produit naturellement dans certains environnements comme les sols non oxygénés ou le système digestif des ruminants. C’est une dégradation naturelle de la matière organique en absence d’oxygène, qui est assurée grâce à l’action concertée de micro-organismes présentant une forte diversité (bactéries et archées) et constitués en réseau trophique. Le processus de digestion anaérobie se décline en quatre étapes séquentielles (Moletta, 2011) réalisées par plusieurs groupes microbiens fonctionnant en symbiose : les bactéries hydrolytiques acidogènes, les bactéries acétogènes et les bactéries méthanogènes (acétoclastes et hydrogénophiles).

Ces quatre étapes sont :

- L’hydrolyse du substrat, première étape durant laquelle les biopolymères (protéines, lipides, carbohydrates) sont hydrolysés en monomères grâce à des enzymes sécrétées par des micro-organismes hydrolytiques ;
- L’acidogénèse où les produits de l’hydrolyse sont transformés en acides gras volatils (acétate, propionate, butyrate), en acides organiques (lactate, succinate) et en gaz (dihydrogène et dioxyde de carbone) par les bactéries hydrolytiques acidogènes ;
- L’acétogénèse, durant laquelle les produits de l’acidogénèse sont convertis en acétate. Elle est réalisée par deux types de populations microbiennes : les bactéries acétogènes (transformation des acides organiques en acétate) ou les bactéries homoacétogènes (transformation du dihydrogène et du dioxyde de carbone en acétate) ;
- La méthanogénèse, étape finale de la digestion anaérobie, où le méthane est formé soit par les méthanogènes acétotrophes à partir d’acétate, ou soit par les méthanogènes hydrogénotrophes à partir du dihydrogène et du dioxyde de carbone.

Chaque produit formé au cours d’une étape de la digestion anaérobie peut être utilisé comme substrat pour l’étape suivante. Cette chaîne trophique est présentée en Figure I-13. La biomasse lignocellulosique peut ainsi être valorisée en biogaz (méthane, CH<sub>4</sub> et dihydrogène, H<sub>2</sub>) susceptibles d’être utilisés comme combustibles, en digestat pour l’amendement organique des sols, mais aussi en synthons d’intérêts industriels (Acides Gras Volatils, AGV) en inhibant la méthanogénèse, dernière étape de la digestion anaérobie. Les acides gras volatils étant des mono-

carboxylates à chaînes courtes et linéaires, on parle également de bioconversion de la biomasse en carboxylates en qualifiant la lignocellulose de « plateforme de carboxylates » (Agler et al., 2011).



**Figure I-13 : Représentation simplifiée des étapes de la digestion anaérobie : de l’hydrolyse de la matière organique complexe à la méthanogénèse.**

(1) L’hydrolyse

L’hydrolyse est la première étape de la digestion anaérobie et consiste principalement en la transformation des molécules organiques complexes en molécules simples par des bactéries acidogènes hydrolytiques. L’holocellulose (terme englobant la cellulose, les hémicelluloses et les autres polysaccharides) est convertie en monosaccharides hydrolysables (pentoses et hexoses) par l’intermédiaire d’enzymes extracellulaires. La fraction ligneuse peut être partiellement hydrolysable en composés aromatiques mais ce processus, très rare en condition anaérobie, est encore peu compris (Khan and Ahring, 2019). Les protéines et les lipides sont respectivement convertis en acides aminés et en acides gras à longues chaînes libres ou en glycérol. La phase d’hydrolyse de la biomasse lignocellulosique peut être l’étape limitante de la digestion anaérobie en raison de la résistance de la lignocellulose et de la nécessité d’avoir à disposition une large variété d’enzymes hydrolytiques pour la réaliser (voir section I.A.3 Hydrolyse de la lignocellulose).



## (2) L’acidogénèse

L’acidogénèse est la seconde étape de la digestion anaérobie. Cette étape est rapide et peut entraîner l’acidification du milieu. En effet, elle permet, à partir des monomères issus de l’hydrolyse (sucres, acides aminés, etc), la production de gaz (dioxyde de carbone et dihydrogène) et d’acides gras volatils (AGV) tels que le formate (HCOOH), l’acétate (CH<sub>3</sub>COOH), et le propionate (CH<sub>3</sub>CH<sub>2</sub>COOH). En cas de faible activité des bactéries méthanogènes, la conversion des sucres peut également aboutir à la formation d’autres composés comme le butyrate (CH<sub>3</sub>CH<sub>2</sub>CH<sub>2</sub>COOH), le lactate (C<sub>3</sub>H<sub>6</sub>O<sub>3</sub>) et des alcools (Chynoweth and Pullammanappallil, 1996; Pavlostathis and Giraldo-Gomez, 1991). Les principales voies de dégradation d’un glucose (C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>) sont représentées dans les trois réactions suivantes :

(i)  $C_6H_{12}O_6 + 2 H_2O \rightarrow 2 CH_3COOH + CO_2 + 4 H_2$  [production d’acétate, de dioxyde de carbone et dihydrogène]

(ii)  $3 C_6H_{12}O_6 \rightarrow 2 CH_3COOH + 4 CH_3CH_2OOH + 2 CO_2 + 2 H_2O$  [production d’acétate, de propionate et de dioxyde de carbone]

(iii)  $C_6H_{12}O_6 \rightarrow CH_3CH_2CH_2COOH + 2 CO_2 + 2 H_2$  [production de butyrate, de dioxyde de carbone et de dihydrogène]

Ces étapes s’effectuent sans accepteur ou donneur d’électrons inorganiques additionnels.

Les micro-organismes acidogènes sont principalement des genres *Clostridium*, *Bacillus*, *Acetobacterium* et *Enterobacter* (Moletta, 2011).

## (3) L’acétogénèse

L’acétogénèse est l’étape de production d’acétate, de dioxyde de carbone et de dihydrogène à partir des produits d’hydrolyse (1) et d’acidogénèse (2). Elle nécessite la présence d’un accepteur d’électron inorganique. Elle est réalisée par trois groupes de bactéries anaérobies strictes :

- (i) Les bactéries syntrophiques productrices de dihydrogène, réduisent les acides à plus longues chaînes en acétate, dioxyde de carbone et dihydrogène. Le processus métabolique de formation d’acétate à partir des acides à plus longues chaînes est



appelé la  $\beta$ -oxydation (voir équations). Ces réactions sont thermodynamiquement défavorables et ne se produisent qu’à faible pression partielle de dihydrogène ( $<10^{-4}$  à  $10^{-6}$  atm). Le dihydrogène produit doit donc être consommé par un autre groupe bactérien pour que la pression partielle de dihydrogène diminue et que la  $\beta$ -oxydation soit thermodynamiquement possible.

- [Propionate]  $\text{CH}_3\text{CH}_2\text{COOH} + 2 \text{H}_2\text{O} \rightarrow \text{CH}_3\text{COOH} + 3 \text{H}_2 + \text{CO}_2$  [production d’acétate, de dioxyde de carbone et de dihydrogène]
- [Butyrate]  $\text{CH}_3\text{CH}_2\text{CH}_2\text{COOH} + 2 \text{H}_2\text{O} \rightarrow 2 \text{CH}_3\text{COOH} + 2 \text{H}_2$  [production d’acétate et de dihydrogène]

(ii) Les bactéries homoacétogènes constituées de deux groupes, réduisent soit le substrat carboné en acétate sans consommation de dihydrogène ou soit le dioxyde de carbone en acétate avec consommation de dihydrogène par la réaction suivante :

- $4 \text{H}_2 + 2 \text{CO}_2 \rightarrow \text{CH}_3\text{COOH} + 2 \text{H}_2\text{O}$  [production d’acétate]

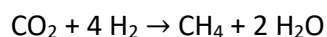
Ces dernières permettent ainsi de consommer le dihydrogène produit par les bactéries syntrophiques et contribuent à la régulation d’un niveau bas en dihydrogène dans le milieu (Siriwongrungson et al., 2007).

(iii) Les bactéries sulfato-réductrices divisées en deux groupes, peuvent soit oxyder le substrat carboné en acétate et en dioxyde de carbone ou soit oxyder les produits de l’acidogénèse en dioxyde de carbone. Elles utilisent les sulfates comme accepteur final d’électron. Elles produisent du sulfure d’hydrogène, un inhibiteur des méthanogènes.

### (4) La méthanogénèse

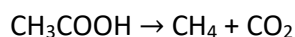
La méthanogénèse, dernière étape de la digestion anaérobie, est assurée par des micro-organismes anaérobies stricts appartenant aux *Archaea*. Les micro-organismes sont divisés en deux groupes :

(i) les méthanogènes hydrogénotrophes produisent du méthane et de l’eau à partir du dioxyde de carbone par cette équation :



Cette voie métabolique est importante pour l’étape d’acétogénèse car elle permet de maintenir une pression partielle de dihydrogène basse dans un système de digestion anaérobie.

(ii) les acétotrophes convertissent l’acétate en méthane et dioxyde de carbone par cette équation :



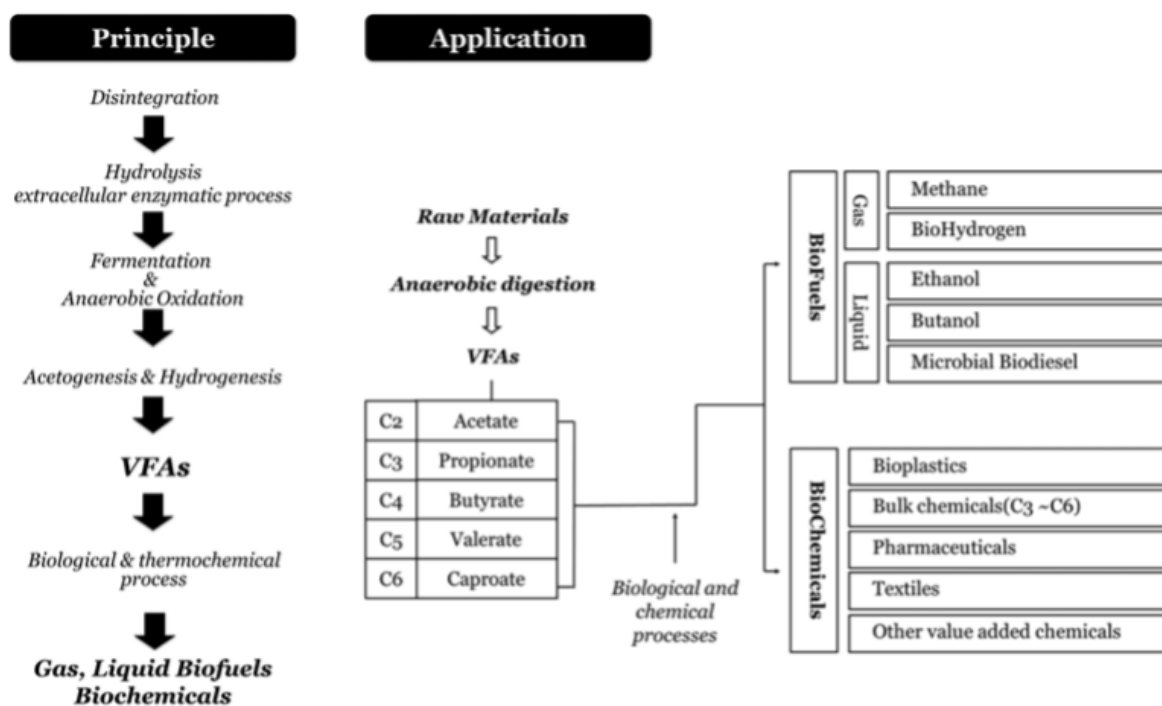
Cette voie métabolique est responsable de 65 à 70% du méthane produit (Pavlostathis and Giraldo-Gomez, 1991).

Le temps de génération des archéobactéries méthanogènes, tout comme les bactéries acétogènes, est supérieur à celui des bactéries hydrolytiques et acidogènes (de quelques heures à 7 jours). Elles sont également fortement sensibles au pH (pH optimal entre 6,8 et 7,4) et sont des bactéries anaérobies strictes.

La digestion anaérobie permet ainsi de valoriser les sucres contenus dans la biomasse lignocellulosique en biogaz : le dihydrogène ( $\text{H}_2$ ) ou le méthane ( $\text{CH}_4$ ). Le dihydrogène est intéressant comme source d’énergie car il est sans effet de serre, a un fort potentiel énergétique (122 kJ/g) et sert de carburant pour les véhicules et technologie à hydrogène, en plein essor. Plusieurs procédés permettent la production de dihydrogène par voie chimique (e.g l’électrolyse de l’eau ou la pyrolyse de la biomasse) mais lorsqu’il est produit à partir de la biomasse lignocellulosique, on parle de fermentation obscure (*dark fermentation*). Elle est réalisée par des bactéries anaérobies strictes de type *Clostridium* et *Enterobacter*. La *dark fermentation* est limitée par l’hydrolyse de la lignocellulose. Le méthane produit, peut être valorisé soit directement en tant que comburant ou soit via cogénération de chaleur et d’électricité. Plusieurs travaux de recherche tentent d’optimiser la production de biogaz à partir de la biomasse lignocellulosique (Baba et al., 2013; Ge et al., 2014; Heeg et al., 2014; Motte et al., 2014; Janke et al., 2015). Dans un but de production d’acides gras volatils, on peut constater qu’inhiber la méthanogénèse est nécessaire puisque cette inhibition permet d’éviter la conversion de l’acétate par les acétotrophes et de récupérer les acides gras volatils produits au cours de l’acidogénèse.

**b) Production des acides gras volatils**

La digestion anaérobie de la biomasse lignocellulosique produit des acides gras volatils (ou AGV) de 2 à 6 carbones grâce à l’activité des bactéries acidogènes et acétogènes. Les AGV peuvent servir de précurseurs moléculaires en chimie verte dans de nombreuses filières (Figure I-14).



**Figure I-14 : Représentation schématique du concept de la lignocellulose comme plateforme de carboxylates pour la production de biocarburants et de composés chimiques. D’après Kim et al., (2013).**

Les AGV produits à partir des substrats lignocellulosiques sont majoritairement l’acétate (C<sub>2</sub>), le propionate (C<sub>3</sub>) et le butyrate (C<sub>4</sub>) (Kivaisi and Eliapenda, 1995; Lazuka et al., 2015; Mockaitis et al., 2020), le valérate et l’isovalérate (C<sub>5</sub>) ainsi que le caproate (C<sub>6</sub>) étant plus rares. Cependant, les conditions et paramètres de la digestion anaérobie peuvent fortement impacter le profil des AGV produits (Zhou et al., 2018).

Au cours de la fermentation de la lignocellulose, le pyruvate (produit de la glycolyse à partir de glucose, ou produit de la voie des pentoses phosphates et de la glycolyse à partir de xylose) est le principal carrefour métabolique de la production des AGV (Figure I-15). L’assimilation du glucose

et des pentoses, composants principaux de la lignocellulose, suit des voies différentes. Alors que le glucose suit la voie de la glycolyse (appelée aussi voie Embden-Meyerhof-Parnas, EMP), aboutissant à la formation de 2 pyruvates par molécule de glucose (pour 2 NADH et 2 ATP produits), les pentoses suivent la voie des pentoses-phosphates : ils sont convertis en pentose-5-phosphate, puis fructose-6-phosphate et glycéraldéhyde-3-phosphate, des intermédiaires de la glycolyse. Le bilan est de 1,67 pyruvates par molécule de xylose (pour 1,67 NADH et 0,67 ATP).

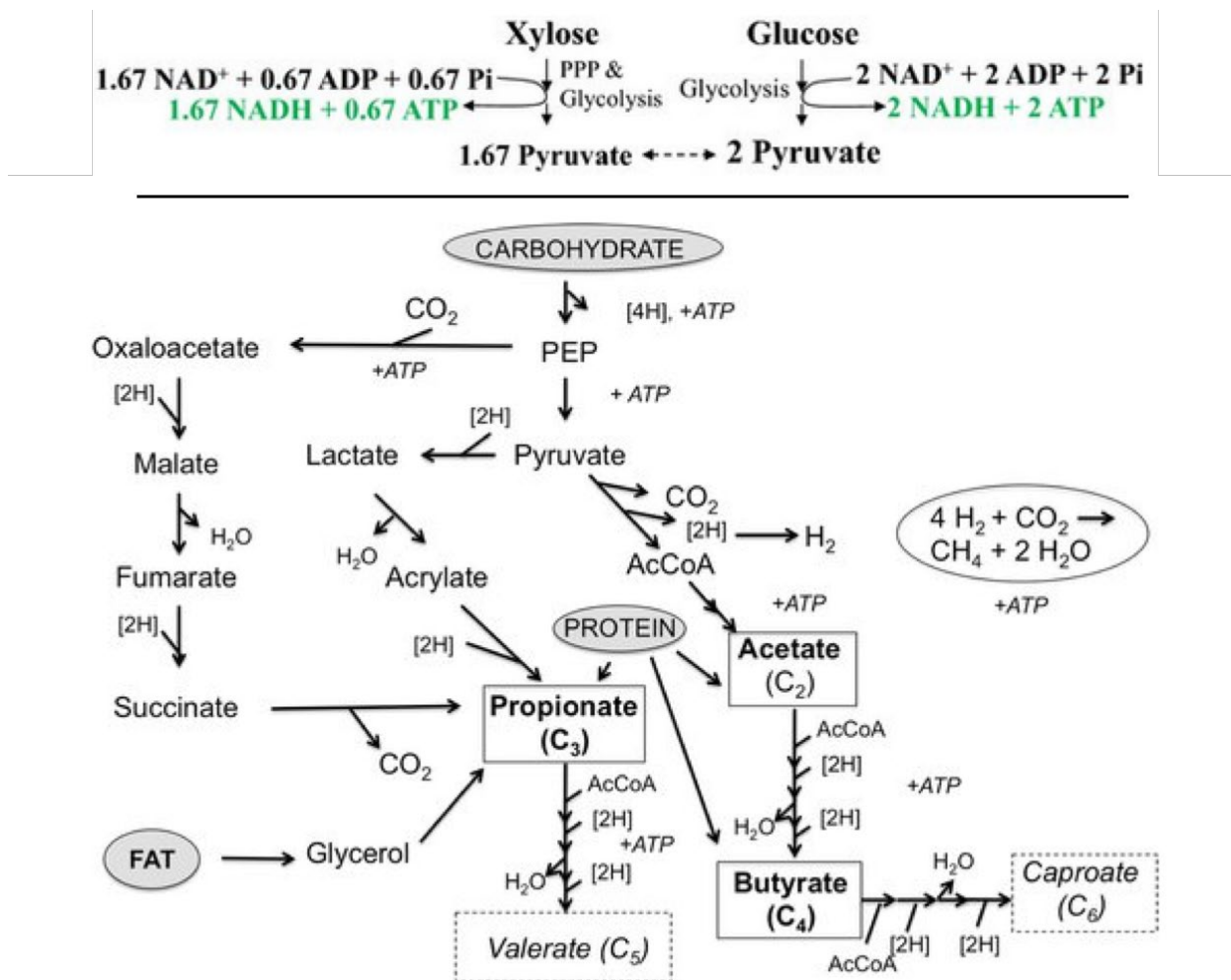


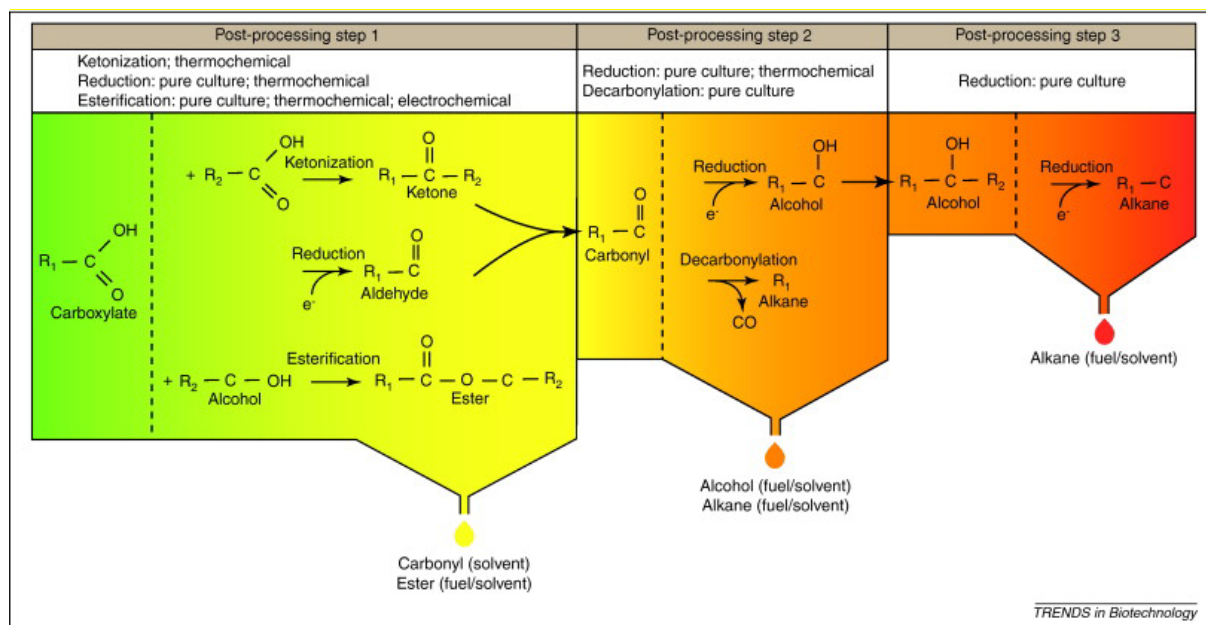
Figure I-15 : Schéma simplifié des voies métaboliques aboutissant à la production d’acides gras volatils (acétate, propionate, butyrate, valérate, caproate) et rendements en NADH, ATP et pyruvate lors de la dégradation du xylose et du glucose (modifié d’après Weimer and Kohn, (2016)).

L’acétate, AGV le plus court à 2 carbones, peut être formé à partir du pyruvate en 2 étapes avec l’acétyl-CoA comme intermédiaire. Le butyrate (C<sub>4</sub>) peut être produit à partir de 2 acétyl-CoA issus

eux aussi du pyruvate, avec une consommation de NADH et production d’ATP. D’autres voies métaboliques existent, comme la reconsommation d’acétate en acétyl-CoA par des CoA-transférases, notamment décrite chez les bactéries du genre *Roseburia* (Duncan et al., 2002). Le butyryl-CoA, intermédiaire produit lors de la formation du butyrate, peut également servir de base à la formation de caproate (C<sub>6</sub>) par  $\beta$ -oxydation inverse (consommation d’acétyl-CoA et de NADH associée à la production d’ATP). La formation de propionate (C<sub>3</sub>) se fait par plusieurs voies assez différentes de celles des AGV en C<sub>2n</sub>, impliquant des intermédiaires comme le lactate ou des métabolites du cycle de l’acide citrique via le succinate (Reichardt et al., 2014). De façon similaire à la formation de caproate à partir de butyryl-CoA, le valérate (C<sub>5</sub>) peut être produit à partir de propionyl-CoA.

### **a) Récupération et valorisation des carboxylates**

Le choix d’utiliser la lignocellulose en tant que plateforme de production de carboxylates plutôt qu’en plateforme de production de gaz présente un intérêt écologique (réduction de la consommation en ressources fossiles) mais également économique. En effet, le marché mondial évalue le méthane à 150-600 \$/tonne alors que les prix des acides acétique, propionique et butyrique sont respectivement de 500–850, 1500–2300 et 1800–1900 \$/tonne (Baumann and Westermann, 2016). L’utilisation des AGV issus de la fermentation pour la production de biocarburant, l’agro-alimentaire ou la chimie verte nécessite des étapes d’extraction et de purification complexes et font l’objet de nombreux développements (Aydin et al., 2018; Aghapour Aktij et al., 2020) permettant d’évaluer les coûts de purification à environ 15\$/m<sup>3</sup> (Ramos-Suarez et al., 2021). La séparation des acides gras volatils du moult de fermentation par extraction ou distillation est très difficile en raison de la formation d’un mélange azéotrope, surtout pour l’acétate, le propionate et le butyrate. Le valérate et le caproate sont quant à eux plus facilement extraits à l’aide de solvants. De plus, pour une utilisation comme précurseurs de biocarburant, les AGV présentent une composition élémentaire plus riche en oxygène qu’en hydrogène et carbone par rapport au pétrole. Une fois produits, la valorisation des AGV en biocarburants ou en solvants nécessite donc des étapes de post-traitements chimiques permettant de convertir les carboxylates (Figure I-16) (Agler et al., 2011, 2012).



**Figure I-16 : Exemple de procédés de post-traitements chimiques permettant de convertir les carboxylates en biocarburants ou en solvants.** Dans l’étape 1, les carboxylates sont convertis en esters par estérification, puis sont réduits ou cétonisés en carbonyles. L’étape 2 consiste en la décarbonylation des composés carbonylés afin d’être convertis en alcanes ou la réduction des carbonyles en alcools. L’étape 3 permet aux alcools d’être convertis en alcanes par réduction. Les carbonyles peuvent être utilisés en tant que précurseurs de solvants alors que les esters, les alcools et les alcanes peuvent servir de précurseurs aux biocarburants et solvants. D’après Agler et al., (2011).

En tenant compte des prix de revente des produits obtenus et des coûts d’exploitation, à partir de déchets alimentaires, la lignocellulose en tant que plateforme de carboxylates pourrait permettre de générer environ 300\$/tonne de biomasse, contre 20\$/tonne pour la méthanisation (Bastidas-Oyanedel and Schmidt, 2018). Pour la bioraffinerie lignocellulosique, ces valeurs économiques pourraient être modulées par des coûts d’exploitation plus élevés, l’efficacité de l’acidogénèse et les cours du marché des AGV. Celui-ci reste cependant prometteur : par exemple en 2012, près de 80% de la production mondiale de propionate était consommée par l’industrie agro-alimentaire (pour laquelle il sert de conservateur naturel), sa demande étant en forte croissance pour remplacer l’utilisation de molécules plus controversées (Baumann and Westermann, 2016).

## 2. Stratégies d’optimisation de la production de carboxylates

L’étape limitante de la digestion anaérobie pour la production de carboxylates est l’hydrolyse. Ainsi, le principal objectif est d’augmenter les rendements d’hydrolyse par différentes

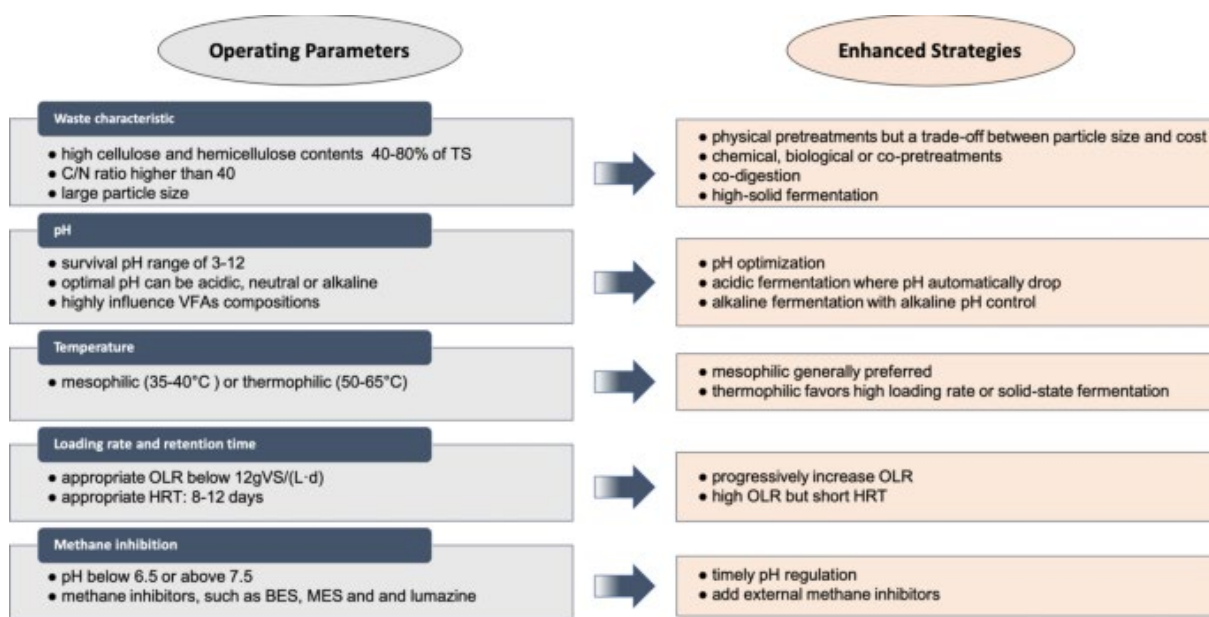
stratégies comme la définition de paramètres optimaux pour la digestion anaérobie, la recherche de sources de biomasse lignocellulosique riches en sucres fermentescibles et facilement hydrolysables, la mise en œuvre de prétraitements permettant de faciliter l’hydrolyse du substrat, la recherche de communautés bactériennes à haute activité hydrolytique et la conduite du procédé de bioraffinerie en voie sèche pour augmenter les rendements.

### ***a) Paramètres de la digestion anaérobie***

Dans la littérature, on retrouve de nombreuses stratégies d’optimisation de la digestion anaérobie pour la production de méthane et certains facteurs permettent aussi de maximiser la production des AGV (Figure I-17). Ainsi dans cet objectif, il est nécessaire d’inhiber la méthanogénèse en utilisant du 2-bromoethanesulfonate (BES, Webster et al., (2016)) pour éviter la consommation des AGV produits. Il est également possible de déterminer entre autres le substrat idéal en fonction de ses caractéristiques physico-chimiques (voir section I.B.2.a. Sources de biomasse lignocellulosique) en choisissant un substrat avec un taux élevé en cellulose et hémicelluloses et un faible ratio C/N puis d’améliorer son hydrolyse par des prétraitements (voir section I.B.2.b Prétraitements appliqués à la biomasse lignocellulosique). Le pH (et sa régulation), ainsi que la température peuvent également être optimisés en fonction de l’inoculum choisi afin de favoriser la bioconversion du substrat. Définir la teneur en substrat à hydrolyser et un court temps de rétention et opérer à forte teneur en solides (voir section I.B.2.e

La teneur en solides) peuvent aussi être des stratégies d’optimisation de production d’AGV (Eryildiz et al., 2020).





**Figure I-17 : Résumé des optimisations possibles des paramètres de la digestion anaérobie pour la production d'acides gras volatils. D'après Sun et al., (2021).**

### ***b) Sources de biomasse lignocellulosique***

Les principaux gisements de biomasse lignocellulosique sont retrouvés dans les sous-produits agricoles (élevage et culture), dans les déchets ménagers et assimilés, et dans les résidus de l'industrie agro-alimentaire et les résidus sylvicoles (Kim and Dale, 2004). Dans ce manuscrit, nous nous focaliserons uniquement sur les résidus de cultures agricoles, correspondant à la partie des plantes cultivées qui ne présente pas de valeur nutritive pour l'homme. Ces sous-produits sont principalement constitués de pailles et de cannes de culture céréalières (blé, maïs, riz, etc.) et sont généralement laissés en champ après récolte. Ils sont relativement secs (80% de matière sèche), riches en lignocellulose et pauvres en azote. La composition de la biomasse dépend de l'espèce, du stade de maturation, des tissus et des organes de la plante (racine, tige, feuille) ainsi que des conditions environnementales dans lesquelles la plante croît (Wilson, 1993). De ce fait, les études proposant des compilations de valeurs de répartition (en % de poids sec) entre cellulose, hémicelluloses et lignine (Prasad et al., 2007; Vassilev et al., 2012; Anwar et al., 2014) affichent une hétérogénéité assez forte de valeurs, parfois peu concordantes, pour un même substrat.

Certains substrats lignocellulosiques sont plus étudiés que d'autres, notamment du fait de leur disponibilité en grandes quantités dans certaines régions du monde, en fonction des particularités régionales et climatiques et donc de l'agriculture pratiquée. Les compositions de quelques



substrats parmi les plus étudiés sont détaillées dans le Tableau I-4.

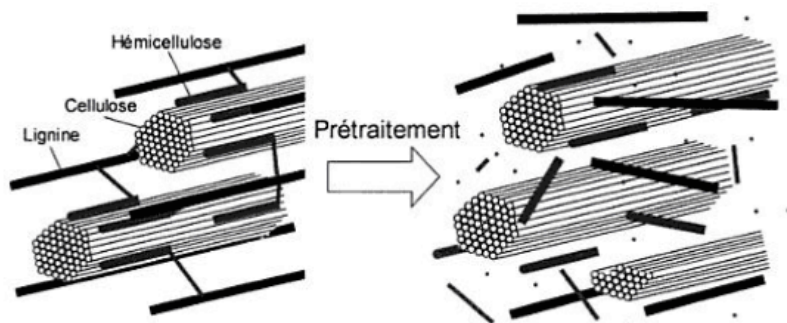
**Tableau I-4 : Compositions biochimiques des différentes fractions majeures (% cellulose, hémicelluloses, lignine exprimés en matière sèche) de biomasses lignocellulosiques issus des cultures agricoles, référencées dans la littérature**

<b>Biomasse</b>	<b>Cellulose (%)</b>	<b>Hemicelluloses (%)</b>	<b>Lignine (%)</b>	<b>Références</b>
Tige de canne à sucre	32-55	25-33	19-25	(Guimarães et al., 2009; Kim and Day, 2011; Vassilev et al., 2012)
Sorgho	45-51	25-27	21-23	(Kim and Day, 2011; Vassilev et al., 2012)
Miscanthus	38	38	25	(Barakat et al., 2013)
Rafles de maïs	45-48	35-37	15	(Prasad et al., 2007; Vassilev et al., 2012)
Résidus de maïs	38-48	26-30	19-22	(Zhu et al., 2005; Vassilev et al., 2012; Barakat et al., 2013)
Paille de riz	32-53	18-33	6-22	(Prasad et al., 2007; Vassilev et al., 2012; Barakat et al., 2013)
Paille de blé	33-45	20-33	15-22	(Prasad et al., 2007; Vassilev et al., 2012; Barakat et al., 2013)

Dans le cadre de ce travail, l’intérêt sera porté sur deux types de résidus agricoles abondants en Europe occidentale : les résidus de paille de blé, initialement utilisés comme substrat modèle au sein du laboratoire, et les résidus de maïs (constitués de feuilles, tiges et rafles de maïs), étudiés dans le cadre du projet Hi-Solids, et dont la France est le premier producteur européen. Ces deux substrats, issus de plantes monocotylédones de la famille des graminées (ou *Poaceae*) ont des compositions similaires. Leur fraction hémicellulosique est composée majoritairement d’arabinoxylane.

### **c) Prétraitements appliqués à la biomasse lignocellulosique**

La structure macromoléculaire de la lignocellulose, variable d’un substrat à l’autre, peut avoir une influence majeure dans l’hydrolyse enzymatique de la biomasse lignocellulosique (Grethlein, 1985; Tanaka et al., 1988; Mansfield et al., 1999; Meng et al., 2013; Zhao and Chen, 2013). De nombreuses méthodes de prétraitement ont été développées pour modifier ou altérer cette structure et faciliter l’hydrolyse enzymatique afin de maximiser l’accès et la récupération des sucres fermentescibles de l’holocellulose (Figure I-18).



**Figure I-18 : Représentation schématique de l’effet destructeur d’un prétraitement sur la biomasse lignocellulosique.**

L’efficacité des prétraitements est dépendante de la structure et de la composition de la biomasse (Kumar et al., 2009b). Il est donc nécessaire d’examiner les paramètres physico-chimiques de la biomasse utilisée pour la bioraffinerie de manière à sélectionner un prétraitement approprié et/ou polyvalent (da Costa Sousa et al., 2009). Selon les produits finaux recherchés, il est possible d’entreprendre un ou plusieurs prétraitements ciblant des fractions différentes de la lignocellulose (Nitsos et al., 2016; Tian et al., 2020).

Plusieurs types de prétraitements ont été étudiés et combinés pour maximiser leur efficacité. Les prétraitements peuvent être de type physique (mécanique et thermique), chimique (acide, alcalin et oxydatif), biologique, ou combiner différents types (Martins et al., 2015). Un prétraitement efficace doit satisfaire plusieurs critères afin d’être choisi : i) améliorer l’hydrolyse de la lignocellulose en sucres simples, ii) préserver les fractions holocellulosiques et limiter la perte des carbohydrates, iii) éviter la formation de produits inhibiteurs pour les activités enzymatiques ou les micro-organismes, iv) avoir un faible impact environnemental, v) être économiquement avantageux et viable. Or l’évaluation de ce dernier critère est complexe car elle implique des coûts de procédés en amont et en aval, des investissements en capitaux, ainsi que des systèmes de recyclage des déchets et produits chimiques qui sont variables (Agbor et al., 2011). Les principaux effets, avantages et inconvénients des grands types de prétraitements sont synthétisés dans le tableau ci-dessous (Tableau I-5).

**Tableau I-5 : Effets sur la lignocellulose, avantages et inconvénients des grands types de prétraitements de la biomasse lignocellulosique**

Pré-traitement	Méthode	Effets sur la lignocellulose	Avantages	Inconvénients	Référence
Physique	Broyage	Augmentation de la surface de contact disponible, diminution du degré de polymérisation et de cristallinité de la cellulose	Amélioration de l’hydrolyse du substrat : diminution du temps et augmentation du rendement ; pas de formation d’inhibiteurs	Consommations énergétiques importantes, rendement en sucre relativement faible par rapport à d’autres méthodes	(Sun and Cheng, 2002; Lazuka et al., 2017)(Lazuka et al., 2017; Sun and Cheng, 2002)
Chimique	Acide	Déstructuration complète de la lignocellulose : libération des oligomères et des monosaccharides	Solubilisation complète des hémicelluloses et de la lignine	Formation d’inhibiteurs de fermentation, difficultés techniques de la manipulation et du recyclage de solutions corrosives	(Saha et al., 2005)
	Alcalin	Rupture des liaisons lignine-carbohydrates, élimination des groupements acétyles et acides uroniques des hémicelluloses ; gonflement et dé cristallisation de la cellulose	Amélioration de l’accessibilité au substrat par augmentation de la surface interne ; peu de formation d’inhibiteurs ; rendement en sucre très élevé	Déchets à gérer après le prétraitement alcalin ; longue durée de traitement ; faible efficacité sur les substrats riche en lignine	(Guo et al., 2011; Lazuka et al., 2017)

Pré-traitement	Méthode	Effets sur la lignocellulose	Avantages	Inconvénients	Référence
Biologique	Micro-organismes	Dégradation de la lignine et des hémicelluloses	Conditions douces de prétraitement ; faibles coûts énergétiques	Traitement long, conditions de croissance exigeantes et perte de sucres pour la fermentation	(Sun and Cheng, 2002)
	Enzymes	Hydrolyse spécifique des fractions ou liaisons de la lignocellulose	Dégradation de la lignine et des hémicelluloses	Coûts très élevés ; conditions de traitement pouvant être incompatibles entre enzymes	(Vandenbossche et al., 2014)

Des prétraitements mécaniques, chimiques et enzymatiques peuvent ainsi être employés pour accélérer et favoriser l’hydrolyse (Alvira et al., 2010). Dans le contexte d’utiliser la biomasse comme « plateforme de production de carboxylates » et les communautés microbiennes, le choix d’un prétraitement de la biomasse doit être fait en respectant sa compatibilité avec la fermentation, avec une formation minimale d’inhibiteurs et de composés toxiques pour les micro-organismes. Dans la suite de cette partie, seuls les prétraitements utilisés dans le cadre de ces travaux de thèse (prétraitements appliqués à la paille de blé dans des travaux antérieurs, et prétraitements appliqués aux résidus de maïs dans le cadre du projet Hi-Solids) seront décrits.

### (1) Prétraitement mécanique

Les prétraitements mécaniques de la biomasse lignocellulosique par comminution comme le broyage ou le déchiquetage permettent principalement d’augmenter la surface spécifique d’échange et de réduire le degré de polymérisation de la biomasse lignocellulosique et la taille des particules (de 50 mm à 60  $\mu$ m). Ils accordent ainsi une meilleure accessibilité aux enzymes et aux solvants, une augmentation de la vitesse d’hydrolyse (-23 à -59%), un meilleur rendement de l’hydrolyse (+5 à +25%) et une fermentation améliorée pour la production de biogaz ou de carboxylates (Hendriks and Zeeman, 2009). Néanmoins, les effets sur la cristallinité sont encore sujet à débat, les résultats de différentes études étant parfois contradictoires (Agbor et al., 2011; Barakat et al., 2014b).

L’inconvénient principal du prétraitement mécanique est son coût énergétique, les prétraitements mécaniques étant très énergivores en fonction des caractéristiques du substrat lignocellulosique

et de la taille de particules désirée à la fin du prétraitement. Cependant, il existe de nombreuses techniques de broyage, dont une plus économe en énergie : le broyage à boulets. Celui-ci permet une augmentation significative de la surface d’accès à la cellulose, sans utilisation d’additifs chimiques. La biomasse traitée conserve la même composition chimique par rapport à la biomasse non traitée et il n’y a pas de génération de fraction liquide, de gaz ou d’inhibiteurs.

Un autre traitement mécanique considéré dans le projet ANR Hi-Solids est l’extrusion. Il sera détaillé dans la section I.B.2.c.4 L’extrusion-réactive.

### (2) Prétraitements chimiques

Les prétraitements chimiques sont les prétraitements les plus étudiés et utilisés pour la délignification des biomasses lignocellulosiques. Ils permettent une bonne récupération des sucres fermentescibles provenant de la fraction holocellulosique. Les prétraitements chimiques les plus couramment utilisés se font par utilisation d’acide, de base ou d’oxydant, mais on ne détaillera ici que les prétraitements alcalins.

L’effet principal du prétraitement alcalin est la solubilisation de la lignine au sein de la biomasse lignocellulosique en réduisant l’encombrement stérique par saponification des liaisons esters entre les xylanes et la lignine. Il provoque également un gonflement et une dé cristallisation partielle de la cellulose ainsi qu’une augmentation de la porosité du substrat. L’élimination des groupements acétyles des hémicelluloses expose la cellulose et améliore son hydrolyse enzymatique. Les prétraitements alcalins sont souvent réalisés à basse température. Ils sont employés avec de l’hydroxyde de potassium (KOH), de sodium (NaOH) ou de calcium (Ca(OH)<sub>2</sub>). Le substrat lignocellulosique doit par conséquent être rincé après traitement. Le prétraitement alcalin possède de nombreux avantages : il fonctionne à basse température, à pression atmosphérique avec des réactifs facilement disponibles, il est efficace sur les substrats à faible teneur en lignine mais il y a des risques de pertes de carbone en dioxyde de carbone et de formation de composés inhibiteurs phénoliques (Banerjee et al., 2011; Rabelo et al., 2014). L’efficacité des prétraitements alcalins (seul ou combiné) a été démontrée sur de nombreux substrats lignocellulosiques à faible teneur en lignine, dont les pailles de blé, de riz, la bagasse, le panic ou encore les résidus de maïs (Hu et al., 2008; Zhao et al., 2008). Le prétraitement alcalin peut se faire à faible teneur en eau (par imprégnation) et être combiné de manière synchrone à un prétraitement mécanique (Barakat et al., 2014a).

### (3) Prétraitements biologiques et enzymatiques

Les prétraitements biologiques permettent la dépolymérisation de la lignocellulose à l’aide de micro-organismes, le plus souvent des champignons cellulolytiques responsables de la pourriture blanche. En effet, ces derniers produisent des enzymes très actives pour la dégradation de la lignine (laccases, peroxydases...), des hémicelluloses et des polyphénols. Ce prétraitement permet d’augmenter l’accessibilité enzymatique au substrat et les vitesses de fermentation. De plus, les prétraitements biologiques sont plus efficaces, économiques et écologiques. Les inconvénients majeurs des prétraitements biologiques sont la lenteur du prétraitement (au moins 10 jours) en conditions contrôlées (aérobie) et la perte d’une partie des sucres, consommés directement par les champignons cellulolytiques (Pérez et al., 2002; Lemée et al., 2012). En raison de ces inconvénients, les prétraitements biologiques se heurtent à des barrières économiques et sont donc moins intéressants à l’échelle industrielle (Eggeman and Elander, 2005).

En revanche, l’utilisation d’enzymes, seules ou en cocktails enzymatiques en tant que prétraitement, est une méthode efficace permettant de récupérer les sucres fermentescibles dans des conditions de réactions douces et écologiques, sans traitement à l’acide ou à la soude (Wyman et al., 2005). L’efficacité de l’hydrolyse enzymatique dépend de plusieurs facteurs comme le pH, le temps de réaction et la température. Si les paramètres des enzymes sont compatibles, il est possible de réaliser l’hydrolyse enzymatique avec plusieurs types d’enzymes ciblant chacune des fractions de la biomasse lignocellulosique : les enzymes ligninolytiques agissent sur la fraction ligneuse tandis que les endoglucanases, les cellobiohydrolases et les  $\beta$ -glucosidases agissent préférentiellement sur la cellulose alors que les xylanases, les estérases et d’autres enzymes accessoires agissent sur les hémicelluloses. La majorité des enzymes utilisées dans ce contexte sont généralement produites par des espèces fongiques (*Trichoderma*, *Aspergillus*, *Penicillium*) (Yang et al., 2001; Eun et al., 2006).

### (4) L’extrusion-réactive

L’extrusion-réactive est un prétraitement prometteur qui permet de combiner les prétraitements mécaniques, et/ou thermiques, et/ou chimiques et/ou enzymatiques. Le procédé d’extrusion a été initialement conçu pour extruder des polymères et pour la transformation des aliments (Riaz, 2000). Cependant, depuis une trentaine d’années, le procédé d’extrusion est également utilisé pour le prétraitement de la biomasse lignocellulosique (Zhang et al., 2012; Vandenbossche et al., 2014; Zheng and Rehmann, 2014). L’extrudeuse se compose d’un cylindre

rotatif avec une ou plusieurs vis qui transportent le substrat en continu de l’entrée jusqu’à la sortie. Un contrôle de la température (chauffage ou refroidissement) peut être fait le long d’orifices sur la longueur de l’extrudeuse, où il est également possible d’injecter des additifs ou enzymes. Les éléments peuvent être configurés de différentes manières permettant de contrôler le mélange, les forces de cisaillement ou la pression. Les extrudeuses à double vis permettent de modifier les propriétés physiques du substrat en réduisant la taille des particules par des forces de cisaillement élevées, sans perte de matière. Elles sont de plus faciles d’utilisation et économiquement compatibles avec une industrialisation (Zheng and Rehmann, 2014) puisque l’extrusion-réactive supporte des traitements en voie sèche et peut fonctionner en système continu. On parle d’extrusion-réactive dès lors qu’on allie le procédé d’extrusion avec un autre prétraitement chimique ou enzymatique au sein de l’extrudeuse. On peut ainsi ajouter des solvants, des agents chimiques ou des enzymes au cours du procédé et adapter les conditions d’extrusion-réactive (température, pression) au sein des différentes sections de l’extrudeuse.

### **d)      *Choix de l’inoculum***

L’hydrolyse complète de la biomasse lignocellulosique nécessite l’action synergique d’une large variété d’enzymes afin de libérer les sucres nécessaires pour la bioraffinerie lignocellulosique. Plus le substrat est complexe, plus un système enzymatique varié est requis. C’est pourquoi dans une perspective industrielle, l’hydrolyse de la lignocellulose par un microorganisme en culture pure ou par des cocktails de quelques enzymes peut être rapidement limitée. Les écosystèmes dégradant naturellement la biomasse lignocellulosique sont sources de producteurs d’enzymes cellulolytiques et hémicellulolytiques et ont été longtemps exploités pour la recherche de nouvelles enzymes ou souches particulièrement actives (Tomazetto et al., 2020). Ils peuvent ainsi être utilisés en culture pour permettre de dépasser ces limitations de bioconversion. L’utilisation directe de consortia en bioréacteurs correspond au procédé dénommé CBP (*consolidated bioprocessing*) où la production d’enzymes, l’hydrolyse du substrat et la fermentation sont couplées au sein d’un même bioréacteur. Le procédé CBP permet ainsi d’améliorer les rendements économiques en augmentant l’hydrolyse du substrat par rapport à l’utilisation d’enzymes ou de souches pures, et en réduisant les coûts (Weimer et al., 2009).

#### **(1)      Les écosystèmes naturels**

Dans les écosystèmes naturels, tous les micro-organismes ne sont pas lignocellulolytiques : la dégradation de la lignocellulose est réalisée par des espèces cellulolytiques

en interaction avec des espèces non cellulolytiques. Loin d’être négative, cette interaction a pour effet d’améliorer la bioconversion de la biomasse lignocellulosique grâce à des synergies (Dehority, 1991; Sawanon and Kobayashi, 2006; Zhang et al., 2007b) et/ou des complémentarités trophiques et métaboliques, qui permettent d’éviter l’accumulation de produits à des concentrations inhibitrices (Gralka et al., 2020). Les consortia microbiens lignocellulolytiques présentent une grande diversité, en lien avec la diversité des substrats et des environnements : tourbières, sols forestiers, systèmes digestifs de Mammifères ou d’Insectes, etc. (Cragg et al., 2015). Ces environnements diffèrent sur de nombreux paramètres tels que l’apport en nutriments ou en oxygène, le brassage et les traitements mécanique et chimique du substrat, la température, le devenir des produits d’hydrolyse, les interactions entre organismes cellulolytiques et non cellulolytiques). Ces paramètres environnementaux ont un fort impact sur les communautés microbiennes, sur leurs stratégies d’hydrolyse de la biomasse et sur leurs activités d’hydrolyse (taux et vitesse). En observant les communautés lignocellulolytiques naturelles, on constate que la richesse du substrat en lignine peut être à l’origine de la structure de la communauté. Dans les sols, la cellulose est présente sous forme de matière décomposée très récalcitrante en raison d’une forte teneur en lignine. Les espèces fongiques sont particulièrement efficaces dans cet écosystème. Ainsi, les sols forestiers, à haute teneur en lignine, présentent des communautés fongiques très abondantes et variées (Buée et al., 2009) où les champignons sont les principaux micro-organismes à disposer de systèmes ligninolytiques oxydatifs extracellulaires agissant en synergie avec les cellulases extracellulaires qu’ils produisent également. A l’opposé, dans les systèmes aquatiques, la biomasse disponible est pauvre en lignine et sa dégradation est dominée par des espèces bactériennes (Mille-Lindblom and Tranvik, 2003). Les températures élevées atteintes pendant le processus de compostage favorisent les bactéries thermophiles au détriment d’autres micro-organismes (Ryckeboer et al., 2003). Enfin, dans les systèmes digestifs, les conditions contrôlées par l’hôte créent différents micro-environnements très spécifiques, souvent anaérobies. Ces derniers peuvent favoriser le développement de protistes ou héberger des symbioses complexes, et exclure la plupart des champignons (Brune, 2014). Il est donc intéressant d’étudier les propriétés et potentiels fonctionnels de plusieurs consortia microbiens afin de choisir celui qui est le plus adapté aux propriétés physico-chimiques du substrat, dans un objectif d’optimisation des rendements de bioconversion. De nombreuses études ont comparé le potentiel de différents écosystèmes en tant qu’inoculum pour la bioraffinerie lignocellulosique (Bayané and Guiot, 2011; Gu et al., 2014; Rajeswari et al., 2021), mais seul le rumen bovin, utilisé lors de ces travaux de thèse, sera détaillé ici.



### (2) Le rumen de vache

Les herbivores vivent en symbiose avec des micro-organismes qui leur permettent de digérer efficacement la lignocellulose en produisant pour l'hôte toutes les enzymes nécessaires à l'hydrolyse et la fermentation de la lignocellulose. Les ruminants comme les ovins et les bovins sont capables de digérer la biomasse lignocellulosique grâce au microbiote présent dans leur tube digestif composé de trois compartiments « pré-estomac » (rumen, bonnet, feuillet) et de la caillette (estomac). Chez le bovin adulte, le rumen est le plus volumineux des trois compartiments « pré-estomac » avec un volume d'environ 100 L. La biomasse ingérée s'écoule vers le rumen, compartiment où elle est exposée à la digestion par le microbiote ruminal. Elle peut cependant subir un ou plusieurs allers-retours vers la bouche où elle est broyée en plus petits morceaux par mastication (processus de rumination, assimilable à un prétraitement mécanique), ce qui la rend plus facilement hydrolysable pour la poursuite de la digestion. Le rumen bovin est un écosystème plutôt stable avec un pH variant de 5,5 à 7 et une température de 39°C.

Le microbiote du rumen est spécialisé dans la fermentation de fourrages riches en lignocellulose en acides gras volatils (AGV) nécessaires au métabolisme énergétique de la vache. Les AGV produits sont majoritairement de l'acétate, du propionate et du butyrate avec un ratio de l'ordre de 0.6, 0.2, 0.1 respectivement et une faible proportion partagée entre le valérate et le caproate (Weimer and Kohn, 2016). Outre les AGV produits, la digestion anaérobie au sein du rumen entraîne la production de gaz tels que le dioxyde de carbone et le dihydrogène dont l'ensemble (AGV et gaz) peut être converti en méthane. Les voies de fermentation des sucres en C<sub>5</sub> et en C<sub>6</sub> étant différentes, le type de biomasse lignocellulosique digérée engendre des ratios de production d'AGV différents. Des teneurs élevées en pectine et en hémicelluloses dans les fibres ingérées induisent un ratio acétate/propionate plus élevé par exemple (Murphy et al., 1982). La production d'acétate libère du dioxyde de carbone lors de la décarboxylation de l'acide pyruvique en acétate, et celui-ci est consommé par les méthanogènes hydrogénotrophes. Les fortes concentrations d'acétate favorisent la production de méthane (Hindrichsen et al., 2004). L'équilibre en composition élémentaire est primordial pour le fonctionnement du rumen, les régimes alimentaires trop riches en amidon peuvent conduire à des acidoses métaboliques par une production trop importante d'acide lactique notamment (Keunen et al., 2002). Le rumen bovin constitue un écosystème de choix pour produire des carboxylates et comprendre les mécanismes de dégradation de la lignocellulose. Il a donné lieu à de nombreuses études qui concernent à la fois l'optimisation de la digestion anaérobie (Hu et al., 2004; Hu and Yu, 2005), de

la production de biogaz (Kivaisi et al., 1992; Baba et al., 2013) et la compréhension enzymatique de cet écosystème (Xing et al., 2012; Béraud, 2013; Solli et al., 2014).

La communauté microbienne présente dans le rumen est composée de bactéries anaérobies facultatives ou strictes ( $10^{10}$  à  $10^{11}$  bactéries/ml de fluide ruminal), de protozoaires ( $10^6$ ), d’Archées ( $10^9$ ), de champignons ( $10^4$ ) (Kamra, 2005). Sur les 500 à 600 espèces microbiennes présentes dans le rumen, la majeure partie des bactéries appartiennent à trois grands phyla dont les abondances relatives varient selon le régime alimentaire ou d’autres facteurs : les Firmicutes particulièrement abondants, les Bacteroidetes et les Proteobacteria. Les autres phyla minoritaires sont les *Fibrobacteres*, les *Spirochaetes* et les *Verrumicrobiales* (Hess et al., 2011). Les 7 genres majoritaires déterminés par séquençage sont *Prevotella*, *Butyrivibrio*, *Ruminococcus*, les *Lachnospiraceae* non classifiés, *Ruminococcus*, *Bacteroides* et *Clostridium*. Les différents taxons bactériens présents dans le rumen présentent différentes stratégies de dégradation de la lignocellulose. L’hydrolyse de la cellulose est associée à certains phyla Firmicutes (*Ruminococcus*) et *Fibrobacteres* (*Fibrobacter*) qui sont pourtant faiblement abondants au sein du rumen (Miron et al., 2001; Ransom-Jones et al., 2012; La Reau and Suen, 2018). Les *Ruminococcus* produisent des cellulosomes pour la dégradation lignocellulosique dans le rumen. L’organisation du système cellulasique des *Fibrobacteres* est, quant à lui, beaucoup moins connu que celui des Firmicutes. Pour les *Fibrobacteres*, la dépolymérisation de la cellulose implique une adhérence au substrat suivie par son clivage par plusieurs types de cellulases (Seon Park et al., 2007; Wilson, 2009; Suen et al., 2011; Ransom-Jones et al., 2014; Neumann and Suen, 2018). *Fibrobacter succinogenes* est une bactérie connue pour s’adhérer à la cellulose et produire du succinate à partir de cette fraction. Les bactéries issues du Phylum Bacteroidetes ne sont pas considérées comme cellulolytiques mais plutôt comme saccharolytiques car elles ont le potentiel génique pour dégrader les fractions pectiques et hémicellulosiques (Naas et al., 2014; Mackenzie et al., 2015). Au sein de ce phylum, les CAZymes des *Bacteroides* sont fréquemment organisées en clusters géniques, physiquement liés et codant des ensembles de gènes codant pour l’ensemble d’une chaîne fonctionnelle impliquant liaison, dégradation et import d’une structure polysaccharidique donnée. Ces clusters de gènes sont dénommés les *Polysaccharide Utilization Loci*, ou PULs, et sont chacun dédiés à l’utilisation de polysaccharides spécifiques (voir section précédente dédiée I.A.3.f Les Polysaccharide Utilization Loci (PULs)). Bien que le genre *Prevotella* soit non cellulolytique, ce genre est présent dans d’importantes proportions au sein du rumen : il participe directement à l’hydrolyse de fibres telles que les hémicelluloses ou l’amidon. Les principales

Archées retrouvées dans le rumen sont des méthanogènes majoritairement hydrogénotrophes, notamment du genre *Methanobrevibacter* et *Methanomicrobium*. Pour compléter les procaryotes cités ci-dessus, le microbiote du rumen bovin comprend également des eucaryotes comme les protozoaires et les champignons anaérobies (classe des *Chytridiomycètes* et famille des *Neocallimasticeae*). Les protozoaires sont cellulolytiques et hémicellulolytiques, ils dégradent l’amidon et sont fortement protéolytiques. Ils participent au recyclage de l’azote dans le rumen par leur activité de prédation des autres microorganismes. Les champignons anaérobies, tels que les *Neocallimasticeae*, produisent du dioxyde de carbone, du dihydrogène et de l’acétate par hydrolyse anaérobie de la lignocellulose grâce à l’action de CAZymes libres et de complexes enzymatiques multimodulaires. Ils contribuent alors indirectement à la production de méthane.

En résumé, l’hydrolyse de la lignocellulose en sucres simples est donc réalisée dans le rumen par certains micro-organismes (*Ruminococcus*, *Fibrobacter*, *Prevotella*, *Bacteroides*, *Clostridium*) et permet à eux-mêmes et à d’autres micro-organismes de réaliser les étapes suivantes de la digestion anaérobie pour produire du dihydrogène, du dioxyde de carbone, de l’acétate, du butyrate, du propionate, du succinate, du lactate et du formate. Les gaz produits (dioxyde de carbone et dihydrogène) peuvent être consommés par les méthanogènes pour produire du méthane tandis que les AGV sont absorbés par l’hôte et entrent dans son métabolisme. Le fonctionnement général et les fonctions des taxons majoritaires sont pour la plupart bien compris et décrits (Figure I-19). Il est néanmoins important de garder à l’esprit qu’une très grande part du microbiote du rumen bovin reste à l’heure actuelle non cultivée et donc assez peu connu. Cependant, les techniques de métagénomique permettent d’identifier des « espèces métagénomiques » par reconstruction de génomes et d’identifier leurs fonctions probables (Stewart et al., 2019). Une forte redondance fonctionnelle existe au sein de cette communauté (Taxis et al., 2015; Weimer, 2015) et confère à cet écosystème des propriétés de stabilité et de robustesse face aux perturbations, ce qui est intéressant pour confirmer le choix du rumen de vache comme inoculum pour un processus industriel.

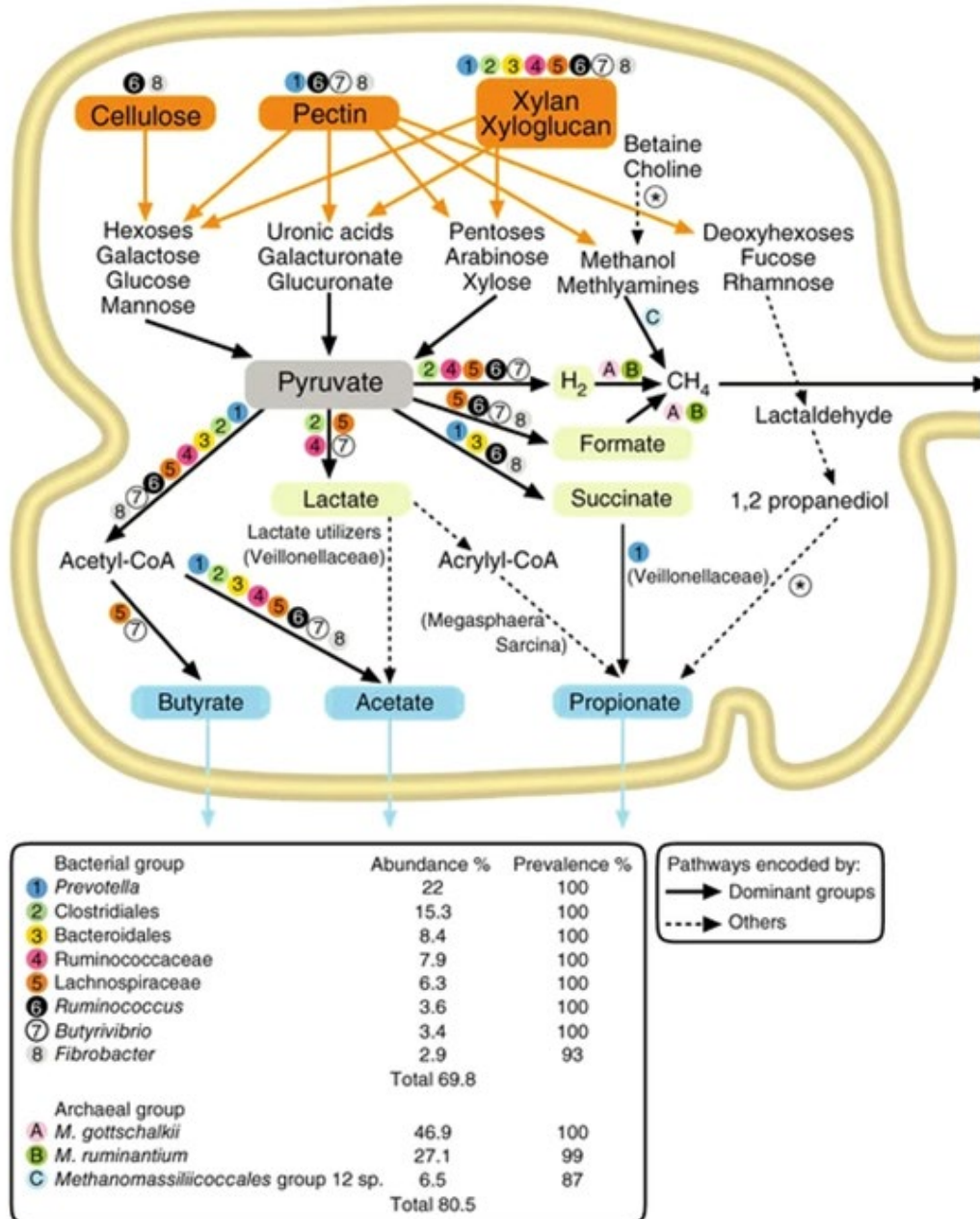


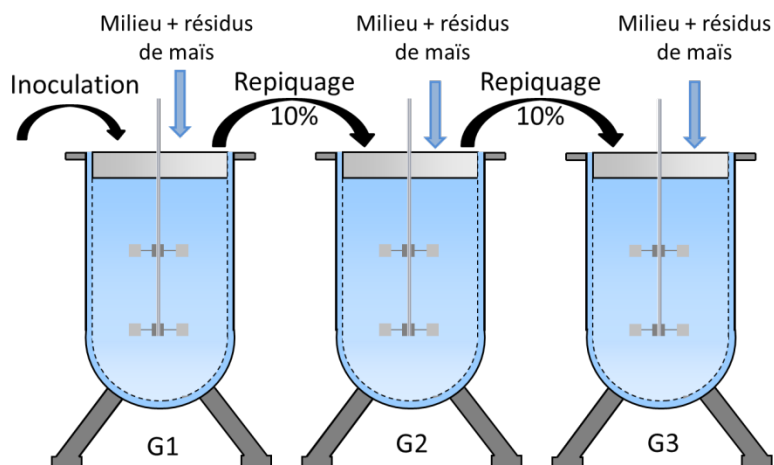
Figure I-19 : Représentation des principales voies métaboliques de dégradation de la lignocellulose dans le rumen bovin et implication des principaux groupes microbiens. D’après Seshadri et al., (2018).

(1) Enrichissement d’un consortium

Un enrichissement consiste à sélectionner une communauté microbienne initiale en lui appliquant différentes pressions de sélection (conditions de culture), dans le but de l’enrichir en micro-organismes pour une fonction voulue. Dans le contexte de la bioraffinerie lignocellulosique,

les caractéristiques souhaitées d’un inoculum après enrichissement sont : une activité hydrolytique et une production de carboxylates élevées, une absence de production de méthane, une vitesse de bioconversion élevée, et une forte stabilité du consortium en condition non stérile.

Plusieurs procédés de culture peuvent être utilisés pour réaliser un enrichissement. Dans des bioréacteurs continus, l’apport continu de biomasse et la sortie continue du contenu du bioréacteur permet d’éliminer lentement les populations microbiennes présentant une vitesse de croissance inférieure au temps de rétention paramétré, tout en évitant l’accumulation de produits à des concentrations inhibitrices. Dans des réacteurs discontinus (ou « *batch* »), l’inoculum et la biomasse lignocellulosique évoluent en système fermé, sans entrée ni sortie de matière. Des variantes existent tout de même où la biomasse peut être apportée de manière progressive (réacteur discontinu alimenté ou « *fed batch* »). Dans le procédé *SBR* (*Sequential Batch Reactor*), des réacteurs discontinus sont enchaînés successivement, en utilisant une fraction du contenu d’un bioréacteur après X temps (X à définir) comme inoculum pour le bioréacteur suivant contenant un milieu et un substrat frais. Ce procédé permet de diluer en cascade l’inoculum initial et les populations peu ou non impliquées dans la bioconversion (dans les conditions de culture fixées) peuvent être éliminées pour concentrer la communauté microbienne en populations actives adaptées aux conditions de culture et au substrat (Figure I-20). Le répertoire enzymatique des consortia microbiens peut ainsi être spécialisé pour déconstruire spécifiquement des matières premières lignocellulosiques dans des conditions industrielles. Des méthodes d’enrichissement successif sur substrat lignocellulosique permettent d’obtenir un consortium hydrolytique complexe, robuste et stable en diversité (Hollister et al., 2010; Reddy et al., 2011; Hollister et al., 2012; Lazuka et al., 2015; Auer et al., 2017).



**Figure I-20 : Principe d'enrichissement d'un consortium microbien sur substrat lignocellulosique par repiquage successif en bioréacteurs discontinus séquentiels (SBR, *Sequential Batch Reactor*)(adapté de Lazuka et al., (2015)).**

Dans plusieurs études, la source de l'inoculum est apparue comme peu impactante sur les capacités hydrolytiques obtenues puisque l'enrichissement à partir de différentes sources d'inoculum ont conduit à des communautés différentes mais à des capacités similaires (Cortes-Tolalpa et al., 2016). On parle alors « d'effet substrat » c'est-à-dire que le substrat et les conditions de culture sont les plus déterminants et conditionnent le potentiel de bioconversion, en opposition à « l'effet inoculum ». Les conditions expérimentales comme l'oxygénation, le caractéristiques du substrat et sa concentration, le milieu minéral, le pH et la température ont un impact très fort sur les populations sélectionnées au cours de l'enrichissement (Hollister et al., 2010; Yu et al., 2015; Ozbayram et al., 2018).

### **e) La teneur en solides**

La voie humide, ou voie à faible teneur en solides, désigne les cultures en bioréacteurs à moins de 15% de solides totaux (TS), tandis que les cultures semi-sèches et sèches, ou à forte teneur en solides, désignent les cultures à plus de 15% de solides totaux (De Baere, 2000). La digestion par voie sèche présente plusieurs avantages. Elle permet une concentration en substrat, donc en sucres, élevée et par conséquent, des concentrations de carboxylates obtenues élevées sans passer par des étapes de concentration de produits. Les coûts de production sont réduits puisque : des réacteurs plus petits sont nécessaires pour l'hydrolyse des sucres et la production de synthons (Ludwig et al., 2013), et la consommation en eau et les besoins énergétiques sont moins

importants (Modenbach and Nokes, 2012; Wang et al., 2012; Modenbach and Nokes, 2013) De plus, les coûts de prétraitements sont également réduits en raison de la haute concentration du substrat traité. Ainsi, le prétraitement et la fermentation à haute teneur en solides (>15% TS) pourraient permettre de rendre le procédé économiquement viable.

Néanmoins, travailler à haute teneur en solides amène à relever de nombreux challenges. En effet, les teneurs en eau et en solides sont des facteurs importants impactant les différentes réactions de la digestion anaérobie. L’eau permet la solubilisation et la diffusion des nutriments jusqu’aux cellules bactériennes et dans les étapes de la digestion anaérobie, elle est également un réactif ou un produit de nombreuses réactions de dégradation de la matière organique. Aussi, la teneur en eau a des effets importants sur la consistance mécanique du milieu de digestion (mélange et viscosité) mais aussi sur la dégradation (inhibition, cinétique, formation de co-produits), ce qui a des impacts sur le mode de fonctionnement et le design des installations. Lors du prétraitement, l’augmentation de la teneur en solides conduit certes à une augmentation des sucres solubilisés, en raison de l’hydrolyse des hémicelluloses, mais induit également une augmentation de la libération d’inhibiteurs phénoliques toxiques (Geng et al., 2015) et à une accumulation de carboxylates à des concentrations potentiellement inhibitrices. Ensuite, la réduction de l’eau libre au sein du réacteur une fois passé un certain seuil (15-20 % TS), accroît la viscosité entraînant une augmentation de l’énergie nécessaire pour l’agitation (Quiroga et al., 2010) ou l’acquisition d’un système d’agitation adapté. L’augmentation de la teneur en solides lors de l’hydrolyse enzymatique augmente aussi la concentration en sucres dans l’hydrolysats, mais diminue linéairement la conversion enzymatique malgré l’utilisation d’un rapport constant enzyme-substrat. Cette baisse est due à l’inhibition des produits intermédiaires et finaux (acides organiques ou phénoliques) en raison des concentrations élevées (adsorption non productive des enzymes) et/ou des limitations diffusionnelles qui abaissent les capacités de liaisons des produits aux enzymes. Des travaux récents indiquent que la limitation de diffusion est le principal facteur (Geng et al., 2015; Paulova et al., 2015).

En termes d’écologie microbienne et d’activités enzymatiques, il est difficile de transposer les connaissances acquises en milieu humide, notamment en raison de la faible motilité des espèces microbiennes ainsi que du manque de connaissances de leur mode d’action dans de tels milieux pauvres en eau (colonisation et croissance). En effet, en voie humide, l’hydrolyse de la lignocellulose est réalisée en partie par des enzymes extracellulaires libres sécrétées dites « *cell-*



*free* », qui sont directement sécrétées dans le milieu pour exercer leur action de manière éloignée de la bactérie productrice. Un second mécanisme d’hydrolyse emploie des enzymes qui se lient au substrat (cellulosomes) ou « *cell-associated* » et qui sont fixées à la paroi cellulaire externe afin d’exercer une action catalytique à proximité directe de la bactérie (Parawira et al., 2005; Zhang et al., 2007a). En voie sèche, comme la teneur en eau est réduite et que le transfert des enzymes solubles est fortement réduit, le second mécanisme serait alors favorisé. Les tests réalisés en laboratoires et à l’échelle industrielle montrent que la limite de travail à des fortes teneurs en solides est de l’ordre de 12 à 20% TS, ce qui est souvent considéré comme la limite maximale à laquelle la biomasse peut être mélangée et hydrolysée dans des bioréacteurs à agitation classique. En utilisant d’autres designs de bioréacteurs, l’hydrolyse enzymatique à l’échelle de laboratoire a été reportée jusqu’à une teneur en solides de 34% TS (Motte et al., 2013). Un certain nombre d’études ont été décrites avec des réacteurs en *fed-batches* en augmentant la teneur en solides progressivement (Hodge et al., 2008; Agua et al., 2015; Cardona et al., 2015).

### **C. Approches omiques pour la compréhension des consortia microbiens lignocellulolytiques**

Les écosystèmes microbiens capables de dégrader la lignocellulose se retrouvent dans des environnements très divers : le sol, les fonds aquatiques, le tube digestif des ruminants ou des xylophages. Ils présentent une très grande diversité : eucaryotes ou procaryotes, aérobies, anaérobies facultatives ou anaérobies stricts, thermophiles ou mésophiles... La majeure partie de la lignocellulose dégradée sur la planète est réalisée en aérobiose à l’air libre et à la surface des sols, mais les micro-organismes anaérobies jouent un rôle important dans la réduction de la matière organique en dioxyde de carbone et méthane dans les sols anoxiques ou dans les chaînes trophiques via des interactions mutualistes. Si quelques-uns sont bien connus et caractérisés depuis la fin du XX<sup>ème</sup> siècle, la plupart des micro-organismes restent non isolés car non cultivables et/ou non cultivés, et leur étude est de ce fait bien plus complexe. De plus, les méthodes classiques de microbiologie, difficilement applicables à ces organismes non cultivables, sont également vite limitées pour l’étude des interactions entre espèces dans leur environnement. La découverte des gènes de l’ARN ribosomal 16S et 18S, en tant que marqueurs universels permettant d’identifier et classifier les espèces à partir d’un petit fragment d’ADN, a révolutionné l’écologie microbienne et a permis d’étudier les communautés microbiennes sans passer par les traditionnelles étapes de



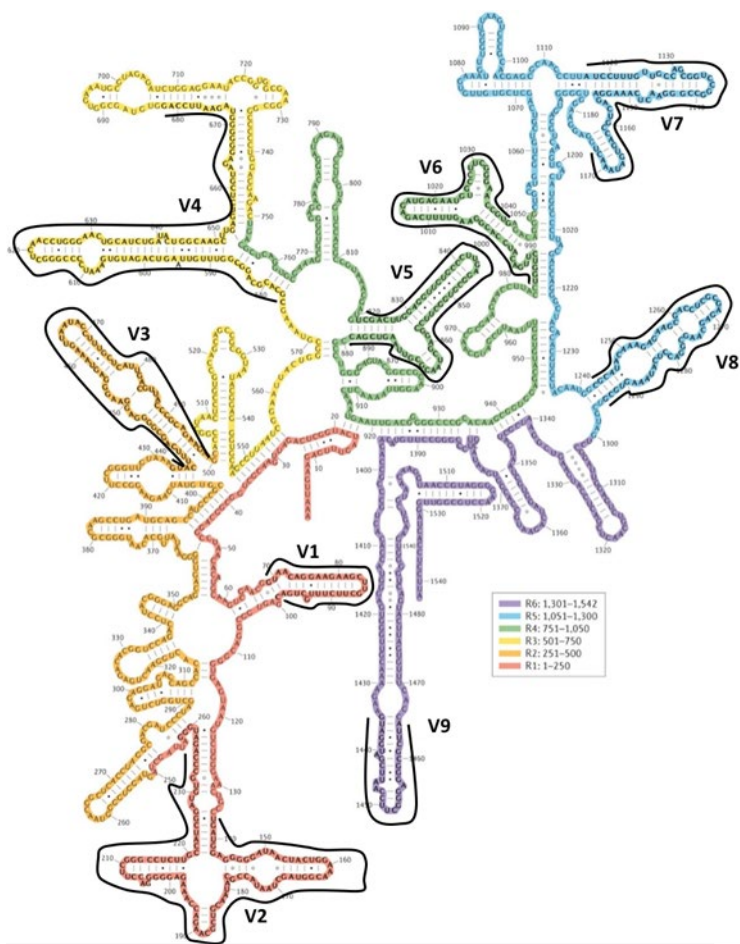
culture et d’isolement. L’essor plus récent des technologies de séquençage (et de la bioinformatique pour les analyser) et des approches « omiques » représente une seconde révolution pour l’écologie microbienne. Les nouveaux outils, ayant permis de révolutionner les concepts de diversité microbienne et d’améliorer la compréhension de la physiologie des consortia complexes, sont : (i) les techniques de séquençage de l’ARN ribosomique 16S/18S qui fournissent des informations importantes sur la composition des espèces et leur dynamique (technique aussi appelée métagénomique amplicon ou métabarcoding ciblé), (ii) les techniques de pyroséquençage et séquençage *shotgun* qui permettent la cartographie de tous les métagénomes ainsi que l’étude des profils métatranscriptomiques des consortia microbiens, (iii) les analyses métagénomiques qui permettent une étude qualitative et quantitative des protéines exprimées par un consortium. Les approches omiques (au minimum de métagénomique amplicon) sont très vite devenues incontournables (Kumar et al., 2008). Les études de métagénomique, métatranscriptomique, métagénomique ou intégrant plusieurs approches métagénomiques sont de plus en plus fréquentes pour mieux comprendre et caractériser les mécanismes des communautés lignocellulolytiques (Deusch et al., 2015; Alessi et al., 2018; Guo et al., 2018; Chukwuma et al., 2021). On détaillera ici les principes fondamentaux et les limites des techniques de métabarcoding et de métagénomique, utilisées dans ce manuscrit pour étudier les communautés microbiennes lignocellulolytiques.

### 1. Le métabarcoding 16S

La compréhension des écosystèmes microbiens s’est accélérée durant ces vingt dernières années en raison des progrès en matière de séquençage à haut débit et de l’arrivée des Nouvelles Générations de Séquençage (NGS). La technique de « métabarcoding ciblé » permet ainsi de caractériser taxonomiquement l’ensemble des micro-organismes d’un échantillon en amplifiant puis séquençant un fragment d’un gène marqueur. Pour ce faire et en résumé, à partir de l’ADN génomique extrait d’un échantillon, un fragment du gène marqueur est amplifié par PCR (*Polymerase Chain Reaction*) à l’aide d’amorces universelles. Les fragments amplifiés suivent alors le protocole de préparation de librairie adapté à la technologie de séquençage (majoritairement Illumina MiSeq) : purification, ajout d’adapteurs de séquençage, multiplexage, mélange équimolaire des échantillons séquencés simultanément. Une fois la librairie (correspondant à l’ensemble des échantillons) séquencée, elle est bioinformatiquement « démultiplexée » afin de récupérer les milliers de séquences correspondant à chacun des échantillons séquencés. Le bruit de séquençage et la variabilité intraspécifique étant non négligeables, des étapes de nettoyage et

de traitement des données sont nécessaires avant d’obtenir une table d’abondance des taxons présents dans les échantillons.

L’utilisation du gène codant pour les ARN ribosomiques 16S et 18S remonte en 1977 (Woese and Fox, 1977), et est rapidement utilisée pour l’identification de micro-organismes isolés. En écologie microbienne, avant que le séquençage sans isolement ne soit possible, ce gène marqueur a longtemps servi pour étudier la diversité des populations à l’aide de techniques de migration de fragments amplifiés, ou séquence-dépendantes (ARISA, TRFLP, DGGE ou SSCP). Chez les bactéries, le gène codant pour l’ARNr 16S est le marqueur majoritairement utilisé mais d’autres se développent en complément du 16S, dont le gène *rpoB* (codant une sous-unité de l’ARN polymérase) ou *gyrB* (codant une ADN gyrase). Le gène de l’ARNr 18S, équivalent du 16S pour les eucaryotes, est assez utilisé pour étudier les protistes. L’ITS (*Internal Transcribed Spacer*, région située entre la petite et la grande sous-unité des ribosomes) est le marqueur le plus utilisé pour étudier les champignons. Dans la suite de ce manuscrit, nous détaillerons uniquement le métabarcoding ciblé 16S.

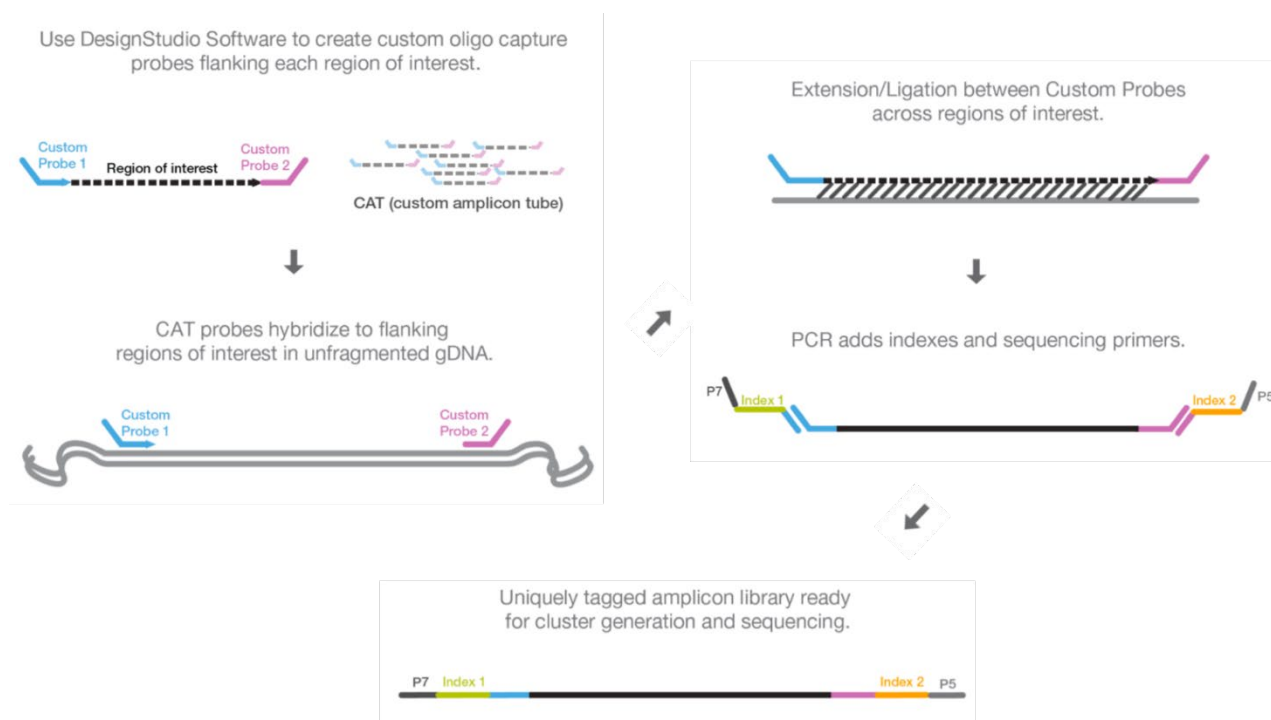


**Figure I-21 : Structure de l’ARNr 16S et position des neuf régions hypervariables (V1-V9).** D’après les travaux de Yarza et al., 2014) et des documents de formation FROGS.

En plus d’être un gène universel aux bactéries, l’ADNr 16S présente l’avantage de se diviser en 9 régions dites hypervariables et de posséder également des régions très conservées, idéales pour dessiner des amorces de PCR (Figure I-21). De nombreuses amorces universelles ont été définies pour ces régions et permettent d’amplifier chacune des régions variables. Les régions variables V3 et V4 sont flanquées de séquences très conservées qui peuvent être amplifiées par des amorces universelles. Le séquençage de ces régions permet une bonne discrimination au genre de la plupart des taxons bactériens connus pour la dégradation de la lignocellulose, et leur longueur associée est compatible avec les protocoles de séquençage (présentés par la suite). Ces régions V3-V4 sont classiquement utilisées par de nombreux auteurs, et plus particulièrement dans les travaux précédents du laboratoire (Lazuka et al., 2015; Auer et al., 2017). Néanmoins, bien qu’universelles, ces amorces excluent parfois certains taxons très divergents. De plus, les régions variables ne présentent pas une variabilité constante dans tout l’arbre phylogénétique des

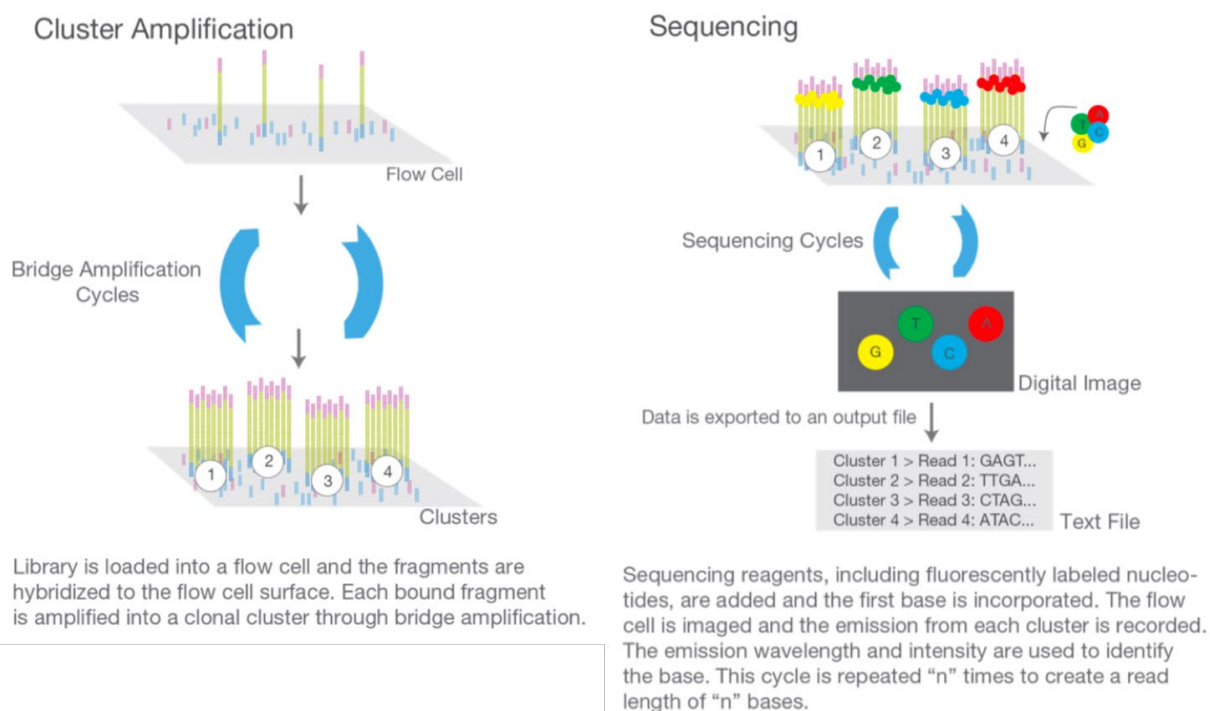
bactéries, et pour des taxons donnés, certaines régions, trop constantes, donnent une faible résolution à l’espèce ou au genre (Yang et al., 2016). Une réflexion sur les choix des régions variables à amplifier et des amorces en fonction des espèces bactériennes attendues est donc nécessaire.

Différentes techniques de séquençage existent, mais l’utilisation de l’Illumina MiSeq pour le métabarcoding ciblé est largement majoritaire du fait de sa bonne fiabilité et qualité par rapport à ses plus anciens concurrents (454, Proton) et de son faible coût par rapport aux technologies les plus récentes (PacBIO notamment). Dans le protocole de préparation « classique » des librairies Illumina, les amorces PCR permettant d’amplifier les régions variables choisies sont prolongées par des fragments complémentaires aux adaptateurs Illumina. Après une purification taille-dépendante des produits de PCR, une 2<sup>ème</sup> PCR permet d’ajouter deux éléments aux fragments amplifiés (Figure I-22) : les adaptateurs Illumina nécessaires dans les étapes ultérieures du protocole, et une étiquette de « multiplexage » composée d’une séquence de quelques nucléotides, unique à chaque échantillon séquencé. Les produits d’amplification de tous les échantillons étant alors identifiables, ils peuvent être rassemblés en une seule librairie pour être séquencés simultanément. Le mélange est réalisé de façon équimolaire afin de garantir une profondeur de séquençage similaire pour chaque échantillon.



**Figure I-22 : Schéma résumant les principales étapes de préparation des bibliothèques Illumina et la construction des amplicons séquencés.** D’après la documentation du constructeur ([www.illumina.com/technology/next-generation-sequencing.html](http://www.illumina.com/technology/next-generation-sequencing.html)).

Le séquençage Illumina est la technique de séquençage par synthèse, basée sur la détection d’une émission de lumière lors de l’incorporation de nucléotides modifiés par une polymérase (Figure I-23). Les quatre types de nucléotides sont présent simultanément lors de la réaction de synthèse. Un seul nucléotide est ajouté à la fois, et son type est déterminé par la lumière émise (différente pour A, T, C et G). Afin que l’émission de lumière générée par l’incorporation des nucléotides soit suffisamment élevée pour être détectée, le signal est amplifié en générant, sur la cellule de séquençage (ou « *flowcell* »), des groupes (ou « *clusters* ») de séquences identiques très rapprochées les unes des autres. C’est à cette étape de fixation des séquences sur la cellule et d’amplification par ponts que servent les adaptateurs Illumina ajoutés lors de la 2<sup>e</sup> PCR. Le séquençage Illumina MiSeq *pair-end* (c’est-à-dire à partir des 2 extrémités) se déroule en deux étapes : 250 à 300 cycles ont lieu dans le sens de lecture 5’-> 3’ (lecture 1 ou R1), puis le même nombre de cycles dans le sens 3’-> 5’ (lecture 2 ou R2). La région amplifiée est donc séquencée par les deux extrémités, d’une longueur de 250 ou 300 nucléotides dans chaque sens. Si des étiquettes de multiplexage ont été incorporées dans les adaptateurs, une 3<sup>e</sup> étape de lecture a lieu pour les séquencer.



**Figure I-23 : Principes généraux du séquençage Illumina : amplification des clusters par ponts (à gauche) et séquençage par synthèse (à droite).** D’après la documentation du constructeur ([www.illumina.com/technology/next-generation-sequencing.html](http://www.illumina.com/technology/next-generation-sequencing.html)).

A partir des étiquettes de multiplexage, l’ensemble des séquences R1 et R2 sont réattribuées à leur échantillon respectif, et on obtient en sortie de machine autant de fichiers R1 et R2 que d’échantillons séquencés. Ces fichiers, au format FASTQ, contiennent les séquences nucléotidiques séquencées et l’information de qualité de séquençage pour chacune des 250 (ou 300) bases de chaque lecture (*read*). La première étape de traitement bioinformatique consiste à rassembler en « contigs » les lectures R1 et R2. La région V3-V4 à séquencer faisant en général une longueur de 420 à 460 paires de bases (pb), les lectures 250 ou 300 pb issues du séquençage (R1 et R2) sont chevauchantes sur leur extrémité et sont complémentaires.

De nombreux outils permettent d’automatiser une partie des étapes bioinformatiques qui vont suivre, dont FROGS (*Find Rapidly OTUs with Galaxy Solution*, Escudié et al., 2017), outil dédié au traitement de séquençage d’ARNr 16S développé sous environnement Galaxy (Figure I-24), disposant d’une interface graphique et utilisé dans ces travaux. Plusieurs outils sont capables d’apparier les séquences R1 et R2, dont VSEARCH (Rognes et al., 2016), incorporé dans l’outil de prétraitement « preprocess » de FROGS. Lors de cette étape de prétraitement, les lectures non fusionnables sont considérées comme des erreurs (qualité trop faible, mésappariements trop

fréquents ou non correspondance des extrémités) et sont éliminées, les autres sont fusionnées en contigs. Ceux-ci passent à travers d’autres filtres successifs de qualité et sont éliminés s’ils présentent des tailles non attendues, si les amorces PCR ne sont pas retrouvées, ou s’ils comportent des bases indéterminées (N).

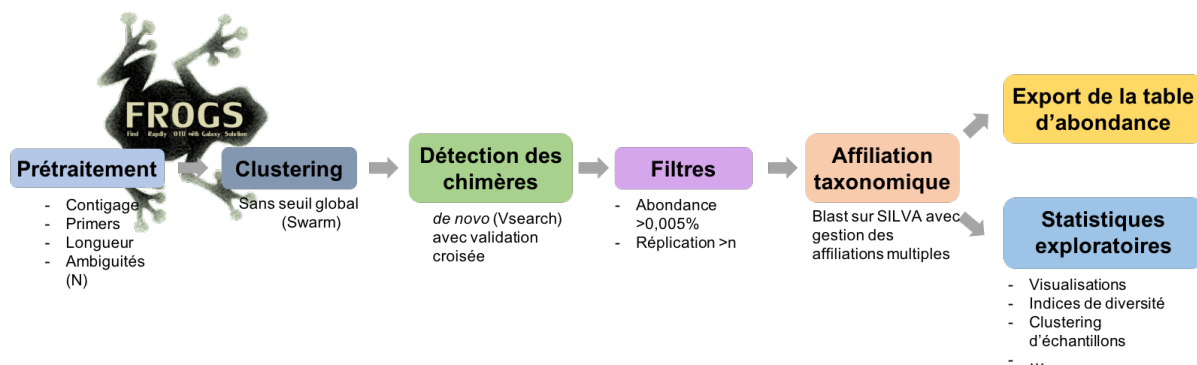


Figure I-24 : Schéma général du pipeline d’analyse FROGS.

Les amplifications successives (1<sup>ère</sup> et 2<sup>ème</sup> PCR, amplification par ponts) et le séquençage lui-même introduisent des erreurs dans les lectures (*reads*), et il peut par ailleurs exister de la variabilité intra-génomique entre les différentes copies du gène 16S. Les données de séquençage obtenues sont donc très bruitées. Afin de réduire cette variabilité, on réalise classiquement une étape de « clustering » pour regrouper en unités taxonomiques opérationnelles (OTU) les séquences présentant moins de 3% de différences entre elles, ce seuil étant basé sur la différence moyenne entre les séquences 16S de deux espèces différentes. Cependant, la définition de ce seuil et les modalités de son application sont parfois contestés et d’autres approches sans seuil ont émergé, dont l’outil de clustering proposé dans FROGS, Swarm (Mahé et al., 2014). Celui-ci permet d’éviter les problèmes liés à l’utilisation d’un seuil fixe mais génère en contrepartie beaucoup de *singletons* (cluster isolé contenant un seul contig) qui doivent être filtrés par la suite. Les étapes de PCR induisent également la formation de chimères. Celles-ci sont composées d’au moins 2 séquences réelles et sont produites lorsqu’un fragment généré lors d’une amplification non complète sur une séquence mère, s’hybride à une séquence parent différente lors d’un cycle ultérieur. Selon les conditions de PCR et la composition taxonomique des échantillons, les séquences chimériques peuvent représenter jusqu’à 10 voire 20% des séquences obtenues, et doivent être éliminées. Différents outils existent pour les détecter et FROGS repose sur l’utilisation d’un outil *de novo*,

VSEARCH (Rognes et al., 2016) et d’un contrôle par validation croisée des chimères à travers les échantillons analysés simultanément.

Le bruit de séquençage résiduel peut représenter quelques pourcentages des séquences totales mais est composé de séquences rares et souvent propres à un échantillon. La combinaison d’un filtre sur les très faibles abondances (conseillé à 0,005% par Bokulich et al., (2013)) et les OTUs faiblement répliqués entre échantillons (présents dans moins de  $n$  échantillons, à définir en fonction du plan expérimental) permet d’obtenir avec FROGS des données extrêmement proches des données attendues ( résultats sur données simulées, Escudié et al., 2017). Les séquences représentatives de chaque cluster peuvent alors être affiliées taxonomiquement en utilisant les banques de données de séquences ribosomales 16S, notamment SILVA, la plus couramment utilisée (Quast et al., 2013) ou la Ribosomal Database Project (RDP, Cole et al., (2014)). Différentes séquences d’une banque de données peuvent être strictement identiques sur la région variable choisie et leurs annotations taxonomiques entrent donc en conflit. Alors que la plupart des outils existant choisissent arbitrairement une des annotations (souvent la première dans la base ou dans l’ordre alphabétique), l’outil FROGS produit dans ce cas une annotation de « multi-affiliation », indiquant qu’il n’est pas possible, sur la seule base de la séquence obtenue, de différencier les taxons concernés.

Enfin, la table d’abondance (effectif de chaque OTU pour chaque échantillon) est exportable à la fin du pipeline d’analyse FROGS. Par la suite, les figures permettant la visualisation des données ainsi que quelques métriques, dont des indices de diversité, peuvent être générées sur FROGS ou sur l’outil statistique R. Ceux-ci permettent de quantifier la diversité d’un échantillon individuellement (indices alpha), ou de quantifier la ressemblance entre deux échantillons (indices bêta). Les indices  $\alpha$  courants les plus simples sont la richesse observée (nombre d’OTU détecté dans un échantillon), les indices de Simpson (en réalité inverse de Simpson) et Shannon qui intègrent tous deux la notion d’équitabilité des abondances de chaque espèce. Plus la distribution est homogène (toutes les espèces sont équi-abondantes), plus la diversité est élevée. La différence entre ces deux indices réside dans le poids attribué aux OTUs rares, l’indice de Simpson repose sur la probabilité d’obtenir deux fois le même OTU en tirant au hasard alors que l’indice de Shannon implique une transformation logarithmique. Les indices de diversité  $\beta$  sont très variés mais peuvent être assimilés à des distances. Ces distances peuvent être basées sur de la présence-absence (Jaccard), intégrer les abondances des taxons pour évaluer la dissimilarité entre deux



échantillons donnés (Bray-Curtis), ou encore être pondérées par les distances phylogénétiques entre OTUs (UniFrac et weighted-UniFrac). La plus couramment utilisée, très majoritaire, est la distance de Bray-Curtis. Elle est calculée en faisant le rapport entre la somme des abondances spécifiques et l’abondance totale. Le calcul des indices de diversité  $\beta$  se fait 2 à 2 entre échantillons et génère donc des matrices de distances. Celles-ci peuvent être utilisées pour visualiser les ressemblances entre échantillons via des méthodes de projection (MDS, NMDS ou PCoA) couramment utilisées dans les études d’écologie microbienne.

### 2. L’analyse métagénomique

Bien que le métabarcoding 16S nous permet de décrire la structure de la communauté microbienne et peut nous donner accès à une prédiction de son potentiel génomique, la métagénomique nous permet de caractériser l’ensemble des protéines exprimées par la communauté à un instant donné, de connaître leur abondance, leur affiliation taxonomique et leur fonction. En 2004, les travaux de Wilmes and Bond ont permis de définir la métagénomique comme « la caractérisation à grande échelle de l’ensemble des protéines d’un écosystème environnemental à un point donné dans le temps » (Wilmes and Bond, 2004). Les progrès rapides durant ces dix dernières années ont permis de mettre au point des technologies à haut débit de type « -omiques » afin de réaliser des études de protéomique environnementale, de métagénomique, d’analyses protéomiques ou de protéogénomiques communautaires (Armengaud et al., 2013).

Une analyse métagénomique *shotgun* et *label-free* (sans marquage) dite « *bottom-up* »\* correspond à l’identification sans *a priori* des peptides issus de la digestion enzymatique des protéines et comprend généralement huit étapes : i) l’extraction des protéines présentes dans l’échantillon, ii) la purification et la concentration protéique, iii) la dénaturation, la réduction et la digestion enzymatique par trypsine des protéines en peptides, iv) la purification des peptides, v) la séparation des peptides en chromatographie liquide ultra haute performance (UHPLC) et l’analyse en spectrométrie de masse en tandem (MS/MS), vi) le traitement informatique permettant l’identification des peptides à partir des spectres obtenus, vii) le regroupement des

---

\* L’autre approche de métagénomique dite « *top-down* » correspond à l’analyse des protéines entières par fragmentation en MS/MS sans protéolyse.

peptides identifiés en groupe de protéines, viii) la quantification des peptides, des protéines et des groupes de protéines ainsi que leur annotation fonctionnelle et taxonomique.

L’avantage de cette approche sans marquage est qu’elle peut être utilisée sur tous types d’échantillons, y compris environnementaux mais cette approche doit relever des défis supplémentaires inhérents à la complexité de ces échantillons : sur ou sous-représentation des organismes/protéines, forte similarité entre séquences protéiques d’espèces proches, hétérogénéité des contaminants.

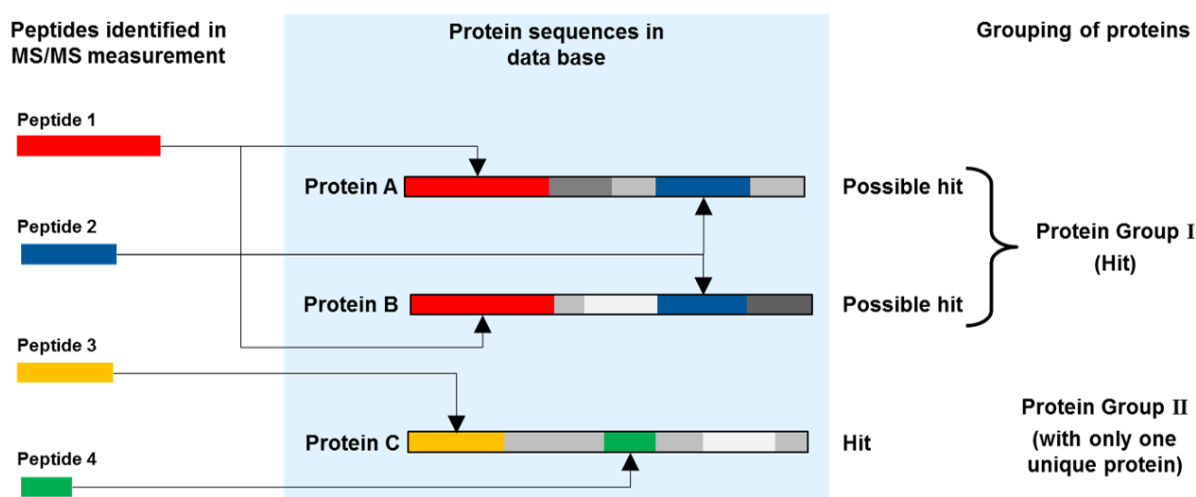
La métaprotéomique *shotgun* sans marquage est réalisée par chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS). La spectrométrie de masse en tandem (MS/MS) a émergé comme un outil très fiable pour identifier les protéines et est souvent utilisée dans la plupart des analyses métaprotéomiques d’échantillons environnementaux. Les peptides sont élués par chromatographie liquide (UHPLC) puis chargés au moyen de l’ionisation par électrospray. La méthode repose ensuite sur la fragmentation des ions issus des peptides ionisés et la quantification en MS/MS du rapport masse sur charge ( $m/z$ ) des ions fragmentés générés par la fragmentation. Les spectres de masse obtenus sont comparés aux spectres de masse théoriques calculés par les algorithmes comme Mascot (Perkins et al., 1999), X!Tandem (Craig and Beavis, 2004) ou Sequest (Eng et al., 2008), algorithme choisi dans ce manuscrit. Ces algorithmes permettent l’analyse à haut débit de milliers de spectres issus de la spectrométrie de masse en tandem qui sont générés en un seul run. L’identification des peptides est réalisée en s’appuyant sur la prédiction des peptides issus de la digestion *in silico* des protéines contenues dans une base de données en prenant en compte quelques paramètres comme l’enzyme utilisée pour la digestion, le nombre de sites de coupure manqués, les modifications chimiques et post-traductionnelles autorisées, l’erreur tolérée en masse pour les ions. Cette base de données peut être issue du métagénome de l’échantillon, ou être issue d’une banque de protéines ou de génomes assez large (base de données non redondantes et annotées automatiquement comme UniprotKB/TrEMBL par exemple ou annotées manuellement comme UniprotKB/Swiss-Prot) dans laquelle il est préférable de sélectionner les taxons susceptibles d’être présents. L’analyse métaprotéomique d’échantillons environnementaux est donc hautement dépendante de la qualité du métagénome s’il est disponible, ou de la richesse et de l’exactitude des informations classées dans les banques de données. L’algorithme Sequest compare les spectres expérimentaux avec les spectres théoriques et attribue un score de similarité  $Xcorr$ . Si un peptide obtient un  $Xcorr$

élevé pour un spectre, alors il est retenu. Enfin, l’identification des peptides est validée par le calcul du taux de faux positifs (FDR, *False Discovery Rate*) en créant une banque de protéines chimères (appelée *decoy database*, Elias and Gygi, (2007)). Cette méthode est la plus utilisée pour valider le match entre spectres et peptides (PSM, *Peptide-Spectrum Matches*). Elle consiste en la création d’une banque de données d’une taille similaire à la banque de données choisie pour l’identification, qui contient des protéines chimères inversées ou formées aléatoirement à partir des séquences protéiques de la banque choisie. Les spectres sont donc comparés à la base de données *decoy* et permettent de calculer le taux de faux positif avec cette formule (Käll et al., 2008) :

$$FDR = \frac{\text{Nombre de PSM avec la banque } decoy}{\text{Nombre de PSM avec la banque de données choisie}}$$

En général, un taux de faux positif de 1 est toléré.

Une séquence peptidique identifiée à partir des spectres peut ensuite correspondre à une protéine ou être partagée par différentes protéines et/ou espèces microbiennes. Si un peptide est commun à plusieurs protéines alors celles-ci sont regroupées en un groupe de protéines. Il suffit d’un peptide unique pour distinguer une protéine d’un groupe de protéines (Figure I-25).



**Figure I-25 : Processus d’identification des protéines et de la création des groupes de protéines à partir des peptides identifiées en spectrométrie de masse en tandem (MS/MS). Le peptide 1 (rouge) et le peptide 2 (bleu) sont communs aux protéines A et B et en raison de**

l’absence de peptide unique détecté lors des mesures en MS/MS, ces deux protéines sont rassemblées dans le « groupe de protéines » 1. Le peptide 3 (jaune) et le peptide 4 (vert) sont des peptides uniques qui permettent de bien caractériser la protéine C et de créer le « groupe de protéines » 2 contenant une seule protéine. Le peptide blanc étant commun aux protéines B et C et le peptide gris clair étant unique à la protéine A, s’il n’y avait pas eu d’autres peptides détectés, les deux protéines B et C auraient constituées ensemble un groupe de protéines tandis que la protéine A aurait constitué à elle seule un autre groupe de protéines.

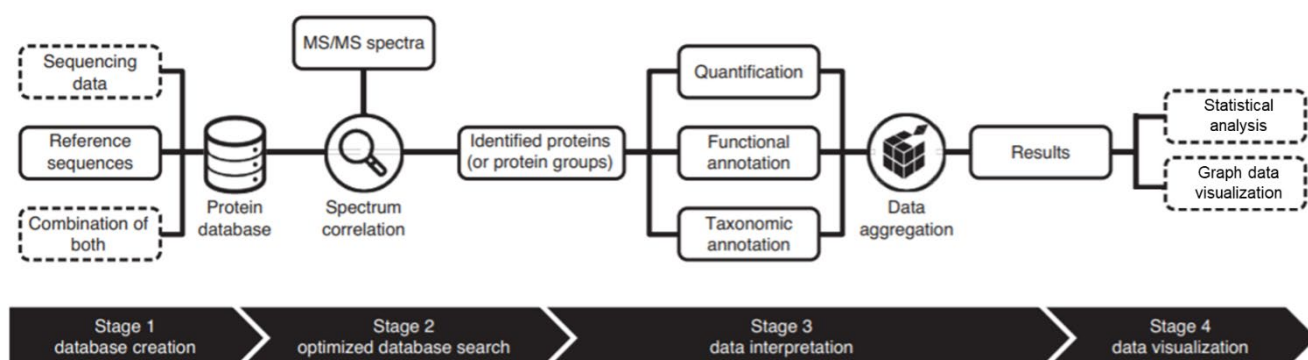
L’approche métagénomique *label free* est basée sur plusieurs méthodes de quantification dont la mesure sous l’aire des pics du signal MS/MS (AUC, *area under the curve*) ou le comptage des spectres de fragments de peptides (Denef et al., 2010; Hawley et al., 2014), technique utilisée dans ces travaux. La méthode du calcul du facteur d’abondance des spectres normalisé repose sur le fait que les protéines plus grandes ou plus abondantes ont tendance à générer plus de peptides en spectrométrie de masse que les petites ou les moins abondantes. Le nombre de spectres acquis pour une protéine est donc divisé par la longueur de cette protéine donnant ainsi un facteur d’abondance spectrale (SAF) pour chaque protéine. Le SAF est ensuite divisé par la somme de tous les SAF obtenus pour cet échantillon, ce qui permet d’obtenir une valeur de NSAF pour chaque protéine dans chaque échantillon comprise entre 0 et 1 (*Normalized Spectral Abundance Factor*).

$$NSAF_i = \frac{\left(\frac{SpC}{L}\right)_i}{\sum_{i=1}^N \left(\frac{SpC}{L}\right)_i}$$

Avec N : le nombre de protéines identifiées ; SpC : le nombre de spectres identifiés par protéine *i* ; L : la longueur de la protéine *i*.

Pour calculer le NSAF d’un groupe de protéines, l’ensemble des spectres correspondant aux peptides de ce groupe est pris en compte. Cette valeur permet une comparaison des abondances de protéines entre différents réplicats techniques ou échantillons biologiques (Zybailov et al., 2006). En l’absence de normalisation interne, elle exige cependant une standardisation des protocoles d’extraction, de purification et de préparation très rigoureuse afin d’éviter toute variabilité technique. D’autres approches avec marquage existent et sont plutôt employées dans des contextes de comparaison d’échantillons et/ou de traitements deux à deux (Pan and Banfield, 2014). L’approche de marquage isotopique des acides aminés dans les cellules par méthode SILAC (*stable isotope labelling by amino acids in cell culture*) requière de cultiver une communauté

référence sur un milieu enrichi en azote  $^{15}\text{N}$  ou en carbone  $^{13}\text{C}$ . L’approche ILC (*Isobaric chemical labelling of peptides*) implique que différents échantillons soient marqués avec des étiquettes différentes (jusqu’à 8 actuellement) puis mélangés pour être analysés simultanément. Ces deux méthodes sont plus contraignantes mais permettent des comparaisons plus fines que le *label-free*, avec lequel il est difficile de pouvoir observer de petites variations d’abondances protéiques (Li et al., 2012). Des outils d’analyses bioinformatiques en libre accès, rapides et simples à utiliser comme Prophan et MetaProteomeAnalyzer permettent ensuite l’annotation taxonomique et fonctionnelle des protéines avec plusieurs bases de données (NCBI, Uniprot, EggNOG, PFAM, TIGRFAM et dbCAN) au travers d’interfaces utilisateurs intuitives et graphiques (Schiebenhoefer et al., 2020). L’ensemble des étapes de bioinformatique peut être résumé en quatre grandes étapes : le choix et la création de la base de données, l’identification des spectres et des peptides et la création de groupe de protéines, l’annotation taxonomique et fonctionnelle des protéines et leur quantification puis la visualisation graphique et statistique des données.



**Figure I-26 : Suite des étapes de bioinformatique pour l’analyse métabolomique *shotgun label-free*.** Adapté de Schiebenhoefer et al., (2020).

La métabolomique a déjà montré son grand potentiel pour l’étude du fonctionnement des communautés lignocellulolytiques en bioréacteur. L’analyse comparée du métabolome extracellulaire d’un consortium issu de compost avec trois différents substrats de complexité croissante (xylose, xylan et résidus de maïs) a révélé qu’au cours de la dégradation, la part des CAZymes parmi les protéines extracellulaires augmentaient de 2 à 20% (Zhu et al., 2016). Majoritairement exprimées par les Firmicutes (le consortium ne comportant qu’une faible proportion de Bacteroidetes), ces protéines lignocellulolytiques extracellulaires ont montré une

grande variabilité d’abondance en fonction des substrats et du temps, démontrant une complémentarité entre taxons Firmicutes et une déconstruction séquentielle de la lignocellulose par le consortium étudié. La stratégie choisie de n’extraire que les protéines extracellulaires/sécrétées a permis de concentrer le signal obtenu sur des protéines potentiellement impliquées dans l’hydrolyse de la lignocellulose. Au vu des résultats présentés, la métaprotéomique semble très prometteuse sur ce point, même si par ce choix, elle empêche d’étudier d’autres aspects de la bioconversion, comme les voies métaboliques impliquées dans la production des carboxylates qui sont internalisées. De plus, aucune donnée n’est disponible sur le comportement des micro-organismes ou des protéines attachés au substrat lors de la centrifugation, et l’activité hydrolytique de certains taxons comme les *Fibrobacter* ou les taxons exprimant des cellulosomes (*Clostridium*, *Ruminococcus*) pourrait être invisible avec une telle approche. Bien qu’une analyse métaprotéomique a assez tôt permis d’étudier l’abondance, l’activité et l’origine phylogénétique des enzymes impliquées dans la dégradation de la biomasse (Schneider and Riedel, 2010), les connaissances sur la régulation/dynamique des communautés lignocellulolytiques sont encore peu fournies. L’analyse métaprotéomique a déjà été utilisée pour étudier l’influence des facteurs environnementaux et de nutriments sur la structure et la décomposition de litière de sols (Schneider et al., 2012) ou pour caractériser le rumen (Deusch et al., 2017; Snelling and Wallace, 2017), mais elle a été surtout appliquée pour rechercher de nouvelles enzymes (Sethupathy et al., 2021). A ce jour, seuls quelques auteurs ont appliqué des approches métaprotéomiques à des consortia enrichis pour la bioconversion de lignocellulose, dans des publications qui restent très descriptives et ne prenant pas vraiment en compte la dynamique temporelle de la dégradation (D’haeseleer et al., 2013; Zhu et al., 2016; Singh et al., 2019). L’identification des protéines microbiennes dans un environnement ainsi que leur origine phylogénétique et leur dynamique temporelle devraient fournir de nouvelles perspectives sur le fonctionnement des différents taxons impliqués dans les processus de bioconversion et leurs interactions, aussi bien pour l’hydrolyse que pour la production des AGV. Cependant, des approches de métaprotéomique comparative s’appuyant sur des conditions ou des prétraitements de substrats différents semblent nécessaires pour dépasser la simple description des fonctions des protéines présentes et leurs origines, que les approches de métagénomique et métatranscriptomiques pouvaient déjà en partie fournir.

### 3. Limite des techniques omiques

La première limitation des omiques concerne leur niveau de dépendance aux banques de données, variable selon les techniques. Selon les environnements, il peut y avoir de nombreuses espèces inconnues qui n’auront de ce fait aucune affiliation taxonomique, ou seulement une annotation de séquence environnementale (déjà détectée dans d’autres études mais jamais identifiée ni cultivée). Par ailleurs, lors de la PCR d’amplification de la région de l’ADNr 16S, le choix des amorces « universelles » est basé sur les séquences conservées connues, et conditionne les OTUs (Unités Taxonomiques Opérationnelles) détectables. De ce fait, certains taxons peu connus sont systématiquement peu amplifiés car ils présentent trop de différences avec les amorces dites universelles, et le problème est sans doute plus important pour des taxons qui seraient complètement inconnus à ce jour (Baker et al., 2003). La métagénomique et la métatranscriptomique, quand elles s’appuient sur des protocoles impliquant une amplification aléatoire, évitent cet écueil. En effet, les algorithmes d’assemblage *de novo* sont en capacité de fournir des séquences complètement inédites et absentes des bases de données. Cependant, certains génomes riches en GC sont très réfractaires à l’amplification et sont pour l’heure impossible à séquencer (Chen et al., 2013). La métaprotéomique *shotgun* elle, est extrêmement dépendante d’une base référence pour l’identification des protéines. Les algorithmes de reconstruction des protéines à partir des peptides identifiés en spectrométrie de masse s’appuient obligatoirement sur une base de peptides prédits à partir de protéines connues dans les banques, et toute protéine inconnue sera de fait complètement invisible à l’analyse. De plus, il est important de ne pas avoir une base de données trop exhaustive sous peine de réduire la robustesse des algorithmes d’identification, surtout lorsqu’on utilise une base de données *decoy* (qui sera d’une taille équivalente à la base de données choisi) pour calculer le taux de faux positif (Muth et al., 2015). La seule façon d’éviter cet important biais des bases de données est de reconstruire le métagénome correspondant aux échantillons et de se servir des gènes qu’il contient comme référence.

Par ailleurs, d’autres biais ont pu être identifiés concernant les approches quantitatives. Le gène de l’ARNr 16S est présent en multiples copies dans les génomes, et ce nombre de copies peut être très variable d’un organisme à l’autre. Les techniques de séquençage ne peuvent que compter l’occurrence des différentes séquences présentes dans un échantillon, ce qui donne une image biaisée des effectifs de population : à abondance égale, un taxon avec un nombre élevé de copies apparaîtra plus abondant qu’un taxon avec peu de copies. Des méthodes de correction

existent, mais elles dépendent de la connaissance des génomes en présence, ce qui est très limitant dans le cas de l’étude d’environnements peu connus, et leur efficacité n’est pas avérée (Starke et al., 2021). Par ailleurs, les amorces universelles n’ont pas une efficacité d’amplification constante quels que soient les taxons (Parada et al., 2016), et les résultats obtenus avec des couples d’amorces ciblant des régions différentes de l’ADNr 16S présentent des différences parfois très importantes à partir d’un même échantillon (Engelbrektsen et al., 2010). De plus, les approches *shotgun* sont connues pour introduire un autre biais : sur-représenter des espèces/protéines abondantes et sous-représenter des espèces/protéines rares. Ce biais est également observable en métaprotéomique mais dans une moindre mesure (Kleiner et al., 2017).

Enfin, les approches de métabarcoding et la métaprotéomique *label-free* fournissent toutes deux des données relatives, qu’elles soient sous forme de comptage ou déjà divisées par le total de l’échantillon. De ce fait, les données obtenues sont compositionnelles : si une variable augmente fortement, toutes les autres diminuent en proportion (Gloor et al., 2017). Les effets compositionnels sont d’autant plus forts que le nombre de variables est petit et les abondances faibles, aussi les données de métaprotéomiques (plusieurs milliers de protéines aux abondances largement inférieures à 1/1000<sup>e</sup>) sont moins sensibles aux effets compositionnels que les données amplicon (quelques centaines d’OTU dont des abondances supérieures à 10%). Cette caractéristique des données omiques commence à être prise en compte dans l’analyse des communautés microbiennes. En effet, si la problématique est considérée depuis plusieurs dizaines d’années dans d’autres domaines comme la géochimie et que des méthodes statistiques prenant en compte cette compositionnalité ont été développées (Aitchison, 1982; Aitchison et al., 2000), elle n’a émergé que plus récemment en écologie microbienne, et les transformations qu’elles impliquent rendent les données omiques assez complexes à interpréter (Hawinkel et al., 2020).





## II. Chapitre 2 – Matériel et méthodes

### A. Substrats utilisés

#### 1. Les résidus de paille de blé bruts

Les expérimentations à la base des chapitres 3 et 4 ont été menées avec de la paille de blé tendre de variété Koréli issue d'une ferme expérimentale d'INRAE située sur la parcelle argilolimoneuse de l'Etang (Commune de Boissy-le-Repos – 51210 France), semée en 2006 et récoltée en Août 2011. La semence est certifiée traitée par un traitement fongicide, Celest XL® (Fludioxonil 25 g/L ; Mefenoxam 10 g/L), selon un dosage réglementaire établi par le Ministère de l'Agriculture de 0.1 L/Q. La paille, coupée au broyeur à couteaux (Retsch SM 100, Germany) et tamisée à 2 mm, a été conditionnée à l'abri de la chaleur et de l'humidité au laboratoire. La paille n'a pas été stérilisée afin de ne pas induire de modifications structurelles dues à la chaleur dans le cas de l'étude de la bioconversion de la paille de blé par le consortium enrichi dérivé de rumen de vache RWS. Dans le cas de l'étude de la bioconversion de la paille de blé avec le consortium enrichi dérivé d'intestins de termites TWS, la paille de blé brute a été autoclavée 20 minutes à 121°C à une pression de 1.2 bar. Le substrat obtenu est appelé « raw wheat straw » (paille de blé brute) dans le chapitre 3 ou « A – 2 mm » dans le chapitre 4.

#### 2. Les résidus de paille de blé prétraités

Trois prétraitements ont été réalisés sur le substrat « A – 2 mm » dans les laboratoires d'INRAE de Montpellier, au Laboratoire Agro-polymère et Technologies émergentes (IATE) par l'équipe du Dr. Abdellatif Barakat. Les différentes conditions de prétraitement pour chaque échantillon sont représentées dans le Tableau II-1. Les prétraitements B et D consistent en un traitement mécanique (broyage) pour réduire la taille des particules à 0.1 mm. Les substrats C et D ont été obtenus après un traitement alcalin à la soude en voie sèche. Des cristaux de soude (NaOH) ont été dissouts dans de l'eau distillée. La solution alcaline a été pulvérisée sur la paille de blé avec une quantité d'eau ajustée pour obtenir une concentration massique de soude à 5% (5 g de catalyseur pour 100 g de paille) et une teneur en eau de 30% (sur matière sèche). Le traitement chimique à la soude a été réalisé avec un ratio biomasse : liquide de 5:1 soit une concentration de 5 KgMS/L. La température de traitement était de 25°C et le temps de séjour de 5 heures. La biomasse traitée a ensuite été séchée à 105°C jusqu'à obtention d'un taux d'humidité de 8 à 10%.

Une fraction des substrats non traité (A – 2 mm) et traité chimiquement (2 mm - NaOH), ont ensuite été broyés par broyeur à impact (Hosokawa-alpine, type UPZ, Augsburg, Germany) équipé d'un tamis 0.1 mm, opéré à température ambiante et à la vitesse de 18000 rpm pour donner respectivement les substrats B (0.1 mm) et D (0.1 mm – NaOH). Les résidus prétraités de paille de blé n'ont pas été stérilisés pour ne pas induire de modifications structurelles supplémentaires occasionnées par l'autoclavage.

**Tableau II-1 : Résumé des conditions des prétraitements mécaniques et chimiques pour chacun des substrats de paille de blé A-B-C-D utilisés dans ces travaux de thèse (MS : matière sèche).**

<b>Substrat</b>	<b>Nom</b>	<b>Prétraitement mécanique (PM)</b>	<b>Prétraitement chimique (PC)</b>
<b>A</b>	A – 2 mm	Broyage 2 mm	-
<b>B</b>	B – 0.1 mm	Broyage 2 mm + Broyage 100 µm	-
<b>C</b>	C – 2 mm – NaOH	Broyage 2 mm	NaOH 5%, 5 kgMS/L
<b>D</b>	D – 0.1 mm - NaOH	Broyage 2 mm + Broyage 100 µm	NaOH 5%, 5 kgMS/L

### **3. Les résidus de maïs bruts**

Les expérimentations du projet Hi-Solids (chapitre 5) ont été réalisées à partir de paille de maïs de la variété P9900 (semis tardif), provenant d'une parcelle expérimentale (INRAE, St-Martin-de-Hinx, France). Les résidus de maïs se composent de feuilles, de tiges, et de rafles laissées dans un champ après la récolte du maïs. Une fois la récolte de 60 kg de paille effectuée, le lot a été séché à 45-50°C en étuve jusqu'à l'obtention d'une humidité résiduelle de 10%. Les résidus de maïs ont ensuite été coupés au broyeur à couteaux (Retsch SM 100, Germany), tamisés à 4 mm, reséchés à l'étuve (45-50°C) jusqu'à atteindre une humidité résiduelle de 10% puis conditionnés en sachets à l'abri de la chaleur et de l'humidité au laboratoire par l'équipe du Dr Abdellatif

Barakat du laboratoire d'IATE à Montpellier. Ce substrat est appelé « substrat brut » (*raw corn stover* en anglais). Les résidus de maïs n'ont pas été stérilisés pour ne pas induire de modifications structurales par un « traitement thermique » lié à l'autoclavage.

### **B. Origine des consortia microbiens**

#### **1. Inocula dérivés de rumen bovin**

Le consortium microbien RWS, caractérisé en métaprotéomique dans le chapitre 3, provient d'un enrichissement, sur paille de blé non stérile, de micro-organismes issus de rumen de vache. Les vaches laitières de race Holstein ont été élevées à l'Unité Mixte de Recherche 1213 (INRAE, Saint-Genes-Champanelle, France), selon la norme nationale fixée par la législation sur la protection des animaux (Certificat d'autorisation d'expérimenter sur animaux vivants n°004495, Ministère de l'Agriculture, France). Les bovins ont été nourris une fois par jour le matin, à partir de rations normalisées pour vaches laitières, composées d'ensilage de maïs (64% MS), de foin (6% MS) et de concentré (30% MS). Le contenu ruminal provenant de deux vaches (Vaches n°3042 et 3059) équipées d'une canule de prélèvement, a été prélevé en dehors d'une période de lactation, homogénéisé, sous-échantillonné et stocké à -80°C, avant de servir d'inoculum pour l'enrichissement d'un consortium microbien sur paille de blé (Lazuka et al., 2015). Les propriétés de bioconversion du consortium microbien RWS obtenu après enrichissement en réacteur séquentiel discontinu (SBR) ont été déterminés dans des travaux antérieurs (Lazuka et al., 2015).

Dans le cas des expériences avec des résidus de maïs, le prélèvement du rumen a également été effectué avant la ration du matin sur trois vaches laitières de l'UMR 1213, équipées d'une canule de prélèvement (Vaches n°3002, 3042 et 4065). Des sous-échantillons de 50 g ont été immédiatement congelés à l'azote liquide, stockés à -80°C avant d'être envoyés au LISBP dans de la carboglace par transporteur et restockés à -80°C.

#### **2. Inoculum issu d'intestin de termite supérieur**

Le consortium microbien TWS, caractérisé en métaprotéomique dans le chapitre 3, provient d'un enrichissement, sur paille de blé stérile, de micro-organismes issus d'intestin de termites. Les termites supérieurs *Nasutitermes ephratae* provenaient de colonies élevées à l'IRD de Bondy, en chambre contrôlée (27°C et 60% d'humidité relative). Un enrichissement en réacteur séquentiel discontinu (SBR) a été mené à partir d'intestins de cinq cent termites ouvriers. Les conditions de

cet enrichissement et la caractérisation macro-cinétique du consortium TWS ont été décrits dans deux publications (Auer et al., 2017; Lazuka et al., 2018).

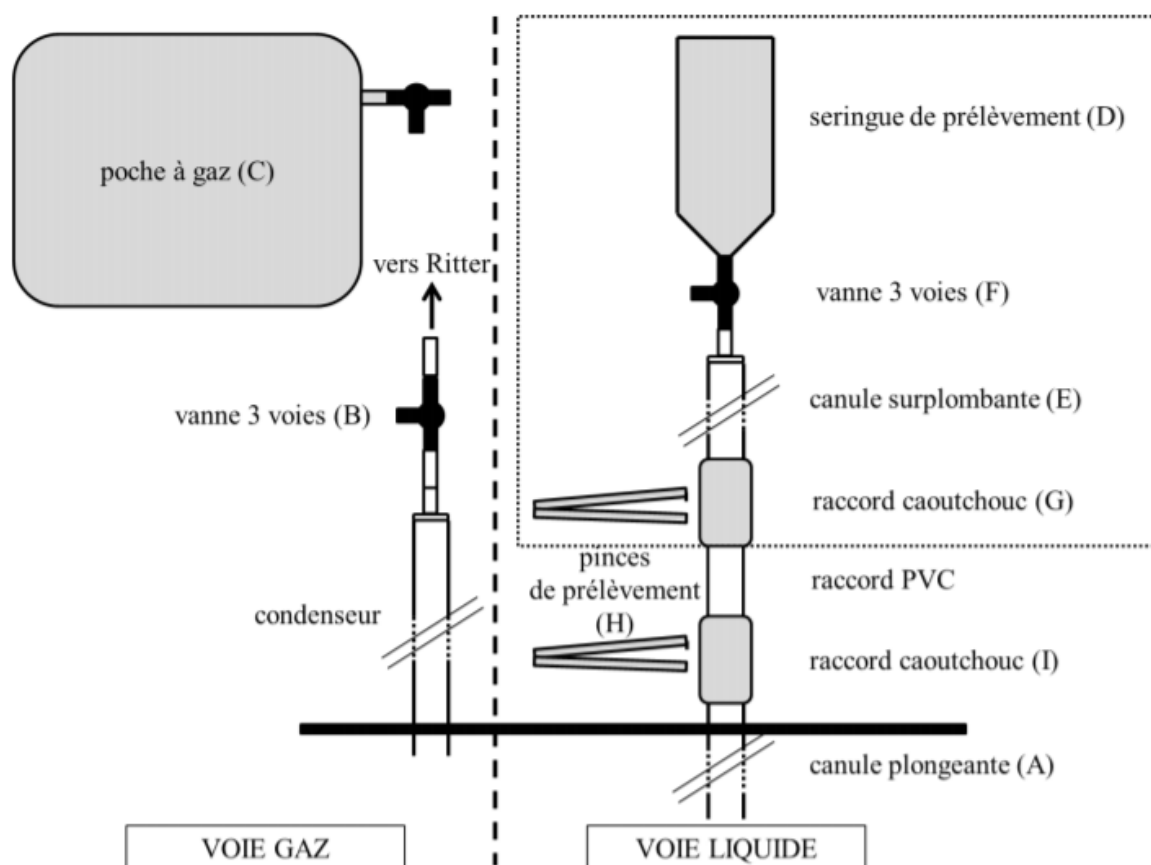
### C. Mise en œuvre des bioréacteurs

La mise en oeuvre des réacteurs des cinétiques RWS et TWS, permettant l'analyse métabolomique des chapitres 3 et 4, a été décrit dans les publications correspondantes (Lazuka et al., 2015, 2017, 2018).

Les cultures (enrichissement et cinétiques du chapitre 5) ont été menées en anaérobiose dans des réacteurs discontinus (*batch*) contenant des résidus de maïs broyés à 4 mm ( $20 \text{ g.L}^{-1}$ ) comme seule source de carbone. Les réacteurs (2 L, BIOSTAT® A+, Sartorius, Allemagne) présentaient un volume utile de 2,4 L et un volume gazeux de 0,6 L. Les expérimentations ont été conduites en condition mésophile et légèrement acidophile avec un pH régulé à  $6,15 (\pm 0,05)$  via l'ajout d'une solution d'hydroxyde de sodium (1N) et avec une température régulée à  $35^\circ\text{C}$ . Un condenseur réfrigéré permettait d'éviter l'évaporation de la fraction liquide. Les choix du pH et de la température d'incubation résultent des conditions retrouvées dans le rumen bovin. La température est légèrement réduite à  $35^\circ\text{C}$  de manière à rendre le procédé moins couteux en énergie et le pH acide permet de minimiser la croissance des archées méthanogènes. Du bromoéthanesulfonate (BES, Sigma Aldrich) était cependant utilisé en tant qu'inhibiteur de méthanogénèse dans des concentrations comprises entre 0,1 et 10 mM. A chaque démarrage de culture, du BES était ajouté pour atteindre une concentration finale en réacteur de 1 mM. Si au cours des expériences, du méthane était détecté dans la phase gazeuse, l'inhibiteur était ajouté jusqu'à une concentration maximale finale en réacteur de l'ordre de 10 mM. L'anaérobiose était assurée par un balayage à l'azote gazeux du milieu de culture au début de chaque expérience. La présence d'oxygène dans le bioréacteur était détectée en continu par une sonde polarographique (AppliSens) qui mesurait l'oxygène dissous dans le moût de fermentation. Si des traces de dioxygène étaient détectées dans le ciel gazeux ou dans le moût de fermentation au cours de l'incubation, alors le bioréacteur était à nouveau flushé par de l'azote gazeux jusqu'à disparition de l'oxygène au sein du bioréacteur. Un compteur gaz à tambour de type Ritter TG 5 permettait de mesurer le volume de gaz produit. La vitesse d'agitation a été fixée à 350 rpm de manière à garantir une homogénéisation correcte de la fraction solide. L'agitation était assurée par un axe muni de deux turbines de Rushton à disque à pales verticales (RT4) et une turbine de Rushton à disque à pales verticales (RT8).

Le milieu de culture est un milieu minéral mimant la composition de la salive des ruminants. Il est adapté pour ne pas présenter d'autre source de carbone que le substrat inséré. Il est constitué d'un inhibiteur de méthanogénèse (BES), d'une solution de vitamines (V7, Pfennig and Trüper, (1992)) et d'une solution d'oligo-éléments (SL12, adaptée d'Overmann et al., (1992)). Ces deux dernières solutions ont été stérilisées par filtration (0,2  $\mu\text{m}$ ). La composition finale du milieu minéral MM est la suivante : (pour 1 L)  $\text{KH}_2\text{PO}_4$  : 3,42 g;  $\text{K}_2\text{HPO}_4$  : 0,846 g;  $\text{NH}_4\text{Cl}$  : 5,58 g;  $\text{NaCl}$  : 0,9 g;  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  : 0,1599 g;  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  : 0,09 g; solution V7 : Biotine : 0,002 mg; p-aminobenzoate : 0,01 mg; thiamine : 0,01 mg; pantothénate : 0,005 mg; pyridoxamine : 0,05 mg; vitamine B12 : 0,02 mg; nicotinate : 0,02 mg; solution SL12 :  $\text{H}_3\text{BO}_3$  : 0,6 mg;  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  : 1,1 mg;  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  : 0,19 mg;  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  : 0,05 mg;  $\text{ZnCl}_2$  : 0,042 mg;  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$  : 0,024 mg;  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  : 0,018 mg;  $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$  : 0,002 mg.

Les prélèvements ont été réalisés grâce à un système de prélèvement, conçu durant les travaux de thèse d'A. Lazuka au laboratoire, qui permet de maintenir l'anaérobiose dans le bioréacteur.



**Figure II-1 : Présentation du système de prélèvement utilisé lors de la conduite des bioréacteurs en anaérobiose (Repris des travaux d'A. Lazuka, 2018)**

A chaque prélèvement, une ponction de 20 mL des gaz présents dans le ciel gazeux était faite grâce à une seringue composée d'une vanne 2 voies (ouverture/fermeture) et s'effectuait au niveau de la vanne 3 voies en B située à la sortie du condenseur (Figure II-1). Cette vanne 3 voies permettait, durant toute la durée du prélèvement, de fermer la connexion entre le réacteur et le Ritter. La mise en place d'une poche remplie à l'azote gazeux au niveau de la vanne 3 voies en B après le condenseur permettait de compenser la dépression occasionnée par ce prélèvement gaz. Durant toute la durée des prélèvements de la fraction liquide, la voie gaz restait connectée à la poche d'azote et déconnectée du Ritter. Pour réaliser les prélèvements de la fraction liquide, une canule plongeante (A - Figure II-1) jusqu'au tiers inférieur du réacteur et de diamètre interne de 1.3 cm permettait de prélever de façon homogène la fraction solide (résidus de substrat) et la fraction liquide (moût de fermentation) grâce à une seringue de prélèvement (D - Figure II-1). Cette seringue était également composée d'une seringue 3 voies (F - Figure II-1) et était reliée à la

canule plongeante avec deux raccords mous en silicone (G et I - Figure II-1), un raccord plus rigide en PVC et une canule surplombante (E - Figure II-1). Ce système permettait d'aspirer et de refouler le moût de fermentation tout en maintenant l'anaérobiose. Pour ce faire, à chaque prélèvement, une première manipulation consistait à aspirer le gaz présent dans le système de prélèvement et à le chasser pour le remplacer par de l'azote gazeux. Le moût de fermentation était alors aspiré jusqu'à atteindre la vanne 3 voies (F - Figure II-1), puis celle-ci était alors positionnée de sorte à sceller le réacteur et à éjecter le gaz aspiré et contenu dans la seringue de prélèvement (D - Figure II-1). Par l'intermédiaire de la vanne 3 voies (F - Figure II-1), la seringue de prélèvement était remplie d'azote gazeux. La vanne 3 voies (F - Figure II-1) pouvait alors être réouverte sur le réacteur de manière à injecter l'azote gazeux afin de repousser le moût dans le réacteur. Plusieurs aspiration-refoulements étaient à présent possibles et étaient nécessaires pour assurer un prélèvement homogène, la paille décantant en quelques dizaines de secondes. La canule surplombante (E - Figure II-1) était graduée afin d'obtenir des volumes de prélèvement précis. Chaque prélèvement était mesuré entre la graduation voulue et le raccord en silicone sous-jacent (G - Figure II-1). Aussitôt le moût était aspiré dans le système de prélèvement au volume désiré, une pince de prélèvement (H - Figure II-1) était clipsée au niveau du raccord en silicone (G - Figure II-1). Une seconde pince était rapidement positionnée au niveau du second raccord en silicone (I - Figure II-1) de manière à sceller le réacteur. Le haut du système de prélèvement (canule surplombante et seringue de prélèvement) pouvait alors être détaché au niveau du raccord G et l'échantillon pouvait être déposé dans le contenant approprié. La suite consistait à reconnecter le haut du système de prélèvement au raccord rigide en PVC et de recommencer le protocole de prélèvement afin de chasser à nouveau l'air présent à l'intérieur du système de prélèvement, de le remplacer par de l'azote gazeux et de refaire les aspirations-refoulements pour un nouveau prélèvement au volume désiré. Une fois l'ensemble des prélèvements effectué, la procédure d'éjection de l'air contenu dans la canule de prélèvement était une dernière fois réalisée puis un clip était fixé au niveau du raccord en silicone (I - Figure II-1). L'ensemble du système de prélèvement restait connecté au réacteur. Un second prélèvement gaz au niveau de la vanne 3 voies (B - Figure II-1) permettait de mesurer les volumes perdus et les transferts de gaz dans la poche d'azote lors du prélèvement et de vérifier l'absence d'oxygène dans le ciel gazeux du réacteur. Enfin la vanne 3 voies (B - Figure II-1) était remise en position ouverte entre le réacteur et le Ritter.



### **1. Enrichissement d'un consortium microbien**

L'enrichissement des consortia microbiens RWS et TWS, analysés par métabolomique dans les chapitres 3 et 4, a été décrit dans les publications correspondantes (Lazuka et al., 2015, 2018).

L'enrichissement d'un consortium dérivé de rumen de vache, pour la bioconversion de résidus de maïs broyés à 4 mm, a été conduit en anaérobiose par cultures successives en duplicats de bioréacteurs discontinus (SBR – *Sequencing Batch Reactor*) dans les conditions de culture décrites précédemment. Avant le démarrage de l'enrichissement, le contenu ruminal de trois vaches a été décongelé doucement sur la nuit à 4°C avant d'être homogénéisé à poids humides équivalents pour constituer l'inoculum du 1<sup>er</sup> cycle d'enrichissement (présenté en chapitre 5). Deux bioréacteurs ont ainsi été inoculés à 2% (m/v, soit 48 g d'inoculum de rumen de vache) et le 1<sup>er</sup> cycle de chaque réplicat a duré 14 jours afin de permettre l'enrichissement de bactéries fibrolytiques à croissance lente, retrouvées naturellement dans l'écosystème du rumen de vache. Le cycle d'enrichissement suivant a été inoculé à partir de 10% (v/v, soit 240 mL) du réacteur précédent (contenant aussi bien la fraction liquide de fermentation que solide). Les cycles suivants ont été conduits jusqu'à atteindre une concentration d'AGV stable, fixant ainsi le temps de culture de chaque cycle d'enrichissement (période variable de 9 à 14 jours). L'enrichissement s'est terminé au bout de 8 cycles. Après le 8<sup>ème</sup> cycle, des échantillons de 240 mL du contenu du réacteur (fractions liquide et solide) ont été congelés sans cryo-conservateur à l'azote liquide et stockés à -80°C pour servir d'inoculum pour les expériences suivantes.

### **2. Essais de stabilité face à la cryo-conservation et mise à l'échelle**

L'impact de la cryo-conservation sur la communauté microbienne enrichie a été déterminée. Pour ce faire, un 9<sup>ème</sup> cycle de culture a été réalisé en duplicats de bioréacteurs (2 L BIOSTAT® A+, Sartorius, Germany) en prenant comme inoculum des échantillons congelés à -80°C du consortium enrichi présentant les meilleures caractéristiques de production d'AGV à partir de résidus de maïs broyés à 4 mm. Les échantillons ont été décongelés doucement sur la nuit à 4°C. Les bioréacteurs de 2,4 L ont été inoculés à 10% (v/v). Les bioréacteurs ont été conduits dans les mêmes conditions expérimentales que celles décrites précédemment pour l'enrichissement et la durée d'incubation a été fixée à 14 jours.

Afin de mesurer la stabilité du consortium enrichi RCS dans des conditions industrielles, une mise à l'échelle du bioréacteur a été faite et une expérimentation a été menée dans un réacteur de 30 L

(30 L BIOSTAT® C+, Sartorius, Germany) avec un volume utile de 24 L. Les conditions de culture sont les mêmes que celles décrites précédemment pour l'enrichissement, en respectant les proportions de substrat (2% TS) et les concentrations du milieu. Ce réacteur a été opéré pendant 14 jours.

### **3. Cinétiques de caractérisation du consortium microbien**

La dynamique de bioconversion des résidus de maïs par le consortium microbien enrichi RCS, dérivé de rumen de vache, a été caractérisée au cours d'une cinétique de 21 jours en bioréacteurs (2 L BIOSTAT® A+, Sartorius, Germany). Les bioréacteurs conduits en duplicats ont été inoculés à 10% (v/v, soit 240 mL) avec le consortium RCS (congelé à -80°C après enrichissement et décongelé doucement sur la nuit à 4°C), puis ils ont été conduits dans les mêmes conditions de culture décrites pour l'enrichissement, l'essai de cryo-conservation et la mise à l'échelle avec une teneur en substrat de 2% TS (m/v, soit 20 g/L). Des prélèvements quotidiens de phase liquide-solide et de phase gazeuse ont été réalisés au cours des 9<sup>ers</sup> jours afin de pouvoir effectuer un suivi fin de la dynamique des composantes macrocinétiques (% de dégradation de la biomasse, production d'AGV et de gaz), enzymatiques (xylanase) et métaprotéomiques, ainsi que l'analyse de la diversité bactérienne. Des prélèvements tous les 3 jours ont eu lieu du 9<sup>ème</sup> jour au 21<sup>ème</sup> jour. Un « réacteur contrôle » non inoculé a permis de mesurer la dynamique de bioconversion du substrat par l'inoculum endogène présent naturellement sur le substrat non stérilisé.

## **D. Analyses macrocinétiques**

### **1. Dosage des acides gras volatils par chromatographie à phase gazeuse**

Un prélèvement de 6 mL du moût de fermentation est centrifugé afin de récupérer le surnageant filtré à 0,2 µm. Les acides gras volatils présents dans le surnageant filtré sont dosés par chromatographie à phase gazeuse (CPG) selon le protocole décrit par Cavallé et al., (2013). Ce dosage s'effectue par la méthode de l'étalon interne (acide éthyl-2-butyrique à 1 g/L dans une solution d'acide orthophosphorique à 5%). La CPG est un modèle Varian 430 composé d'un injecteur 1177 splitless et d'un four à température programmable équipé d'une colonne Agilent CP-Wax 58 FFAP CB. Le gaz vecteur est de l'azote à un débit de 25 L/min, la température de l'injecteur est à 250°C et la température du détecteur est à 275°C. Un gradient de température permet la séparation des différents composés : rampe de 90 à 115 °C (30°C/min), rampe de 115 à 130 °C (1,5 °C/min) et rampe de 115 à 230 °C (70 °C/min). Les acides gras volatils mesurés sont

l'acide acétique (C<sub>2</sub>), l'acide propionique (C<sub>3</sub>), l'acide isobutyrique (C<sub>4</sub>), l'acide butyrique (C<sub>4</sub>), l'acide isovalérique (C<sub>5</sub>), l'acide valérique (C<sub>5</sub>) et l'acide hexanoïque (C<sub>6</sub>). La quantification des AGV est réalisée à l'aide d'une gamme étalon entre 0 et 1 g/L.

### 2. Suivi des gaz produits

La composition des gaz produits (CH<sub>4</sub>, CO<sub>2</sub>, N<sub>2</sub> et H<sub>2</sub>) et l'absence d'O<sub>2</sub> est déterminée par analyse en chromatographie gazeuse grâce à un prélèvement de 10 mL du ciel gazeux du réacteur. Le volume prélevé est compensé par une poche à gaz remplie d'azote. La GC est un modèle HP 5890 composé d'un détecteur à sonde de conductivité, d'une colonne Hayesep D avec tamis moléculaire de 5A. Le gaz vecteur est l'argon avec un débit de 100 mL/min. Les températures respectives de l'injecteur, du four et du détecteur sont 100°C, 60°C et 140°C.

### 3. Dosage de l'activité xylanase

Le dosage de l'activité xylanase est réalisé à partir d'un prélèvement de 3 échantillons totaux de volume connu X (X = 6-7 mL). Les échantillons subissent une centrifugation (7197 x g, 10 minutes, 4°C), permettant d'obtenir d'une part un surnageant contenant les enzymes extracellulaires dites libres ou « *cell-free* » et un culot contenant les enzymes attachées au substrat dites « *cell-attached* ». Le culot a été resuspendu dans 6 mL de tampon acétate de sodium (50 mM – pH 6 ; contenant de l'azide de sodium à 0.05% (m/v)) et soniqué sur glace au travers de 4 cycles de 20 secondes (avec 20 secondes sur glace entre chaque cycle de sonication) avec un sonicateur HD 2070 fixé à une puissance de 40% et équipé d'une sonde MS73 (Bandelin Sonoplus). Cette étape de sonication permet de détacher les enzymes de la matrice lignocellulosique. Chaque échantillon (surnageant ou culot) est ensuite dosé en dupliquat. L'activité xylanase a été mesurée par un dosage colorimétrique à l'acide 3,5-dinitrosalicylique (DNS). L'activité xylanase a été dosée en utilisant du xylan de hêtre (Sigma Aldrich) dissolu à 1% dans du tampon acétate de sodium (50 mM – pH 6 ; contenant de l'azide de sodium à 0.05% (m/v)). L'activité xylanase a été estimée par la mesure de la libération de sucres réducteurs après incubation de 200 µL d'échantillons avec 400 µL de substrat à 35°C pendant exactement 1 heure. Un volume de 600 µL d'une solution de DNS a permis de stopper la réaction enzymatique. Tous les échantillons ont ensuite été chauffés à 100°C pendant exactement 5 minutes afin de réduire les sucres libérés puis refroidis sur glace à l'abri de la lumière pendant 10 minutes. L'absorbance mesurée à 570 nm avec un spectrophotomètre UV/Vis (Multiskan Ascent, Thermo Scientific) a permis de doser la quantité de sucres libérés dans chaque échantillon à l'aide de gammes étalon de xylose entre 0 et 1 g/L. Une

unité d'activité était définie comme la quantité d'enzymes permettant la libération de 1  $\mu\text{mol}$  de sucres réducteurs par minute.

Tous les échantillons étaient dosés contre des témoins : (i) ne contenant pas de substrat (volume de substrat remplacé par un volume équivalent de tampon acétate de sodium) afin de connaître l'apport de sucres par l'échantillon à doser pendant l'incubation, (ii) ne contenant pas d'échantillon afin de connaître l'apport de sucres par le substrat pendant l'incubation (volume d'échantillon remplacé par un volume équivalent de tampon acétate de sodium, (iii) ne contenant ni le substrat ni l'échantillon (volume total de réaction en tampon acétate de sodium) afin de connaître l'absorbance du tampon acétate de sodium et de la solution DNS.

### **E. Analyses omiques**

#### **1. Caractérisation de la diversité microbienne par métabarcoding 16S**

##### ***a) Echantillons totaux***

La caractérisation de la diversité bactérienne a été réalisée à partir d'échantillons de 1 mL (prélevés en quintuplicat), conservés sur glace le temps du prélèvement puis immédiatement ultra-centrifugés (13 000  $\times g$ , 5 minutes, 4°C). Le surnageant était alors éliminé et le culot (contenant toute la biomasse lignocellulosique et bactérienne) a été congelé à l'azote liquide et stocké à -80°C jusqu'à co-extraction des ADN/ARN, détaillée plus loin.

##### ***b) Echantillons fractionnés***

Un suivi de la diversité bactérienne a été réalisé indépendamment sur trois fractions de bioréacteurs : libre, faiblement attachée et attachée. De la même façon que pour les échantillons totaux, des échantillons de 1 mL ont été prélevés (en triplicat) et conservés sur glace le temps du prélèvement. Ils ont ensuite été centrifugés à plus faible vitesse (150  $\times g$ , 10 min, 4°C), ce qui a permis de précipiter la biomasse lignocellulosique dans le culot. Le surnageant a été récupéré dans un autre eppendorf et ultra-centrifugé (13 000  $\times g$ , 5 min, 4°C) afin de précipiter la biomasse bactérienne présente (fraction libre) dans le culot, alors congelé dans l'azote liquide et conservé à -80°C. Le culot de biomasse lignocellulosique obtenu par la 1<sup>ère</sup> centrifugation à faible vitesse a été resuspendu dans 1 mL de NaCl à 0.9% puis centrifugé une seconde fois à faible vitesse (150  $\times g$ , 10 min, 4°C) afin de séparer la biomasse faiblement attaché. Le surnageant a été retiré dans un nouvel eppendorf, ultra-centrifugé (13 000  $\times g$ , 5 min, 4°C) pour obtenir un culot puis

congelé comme la fraction libre. La fraction attachée correspond au culot obtenu après la 2<sup>ème</sup> centrifugation à basse vitesse, congelé dans l'azote et conservé à -80°C.

### *c) Co-extraction ADN/ARN*

La co-extraction ADN/ARN a été réalisée à l'aide du kit PowerMicrobiome RNA isolation (MoBio), en suivant le protocole constructeur dans sa version sans traitement DNase. La lyse cellulaire d'un échantillon congelé a été effectuée en 4 cycles de broyage de 30 secondes à l'aide de microbilles via un équipement FastPrep (MP Biomedicals) opérant à 6 m.s<sup>-1</sup>. Les ADN et ARN obtenus ont été séparés en suivant le protocole fournisseur du kit AllPrep DNA/RNA MiniKit (Qiagen). Les qualités et quantités extraites ont été vérifiées par électrophorèse sur gel d'agarose 1% et mesures d'absorbance à 260 et 280 nm (Nanodrop 1000, ThermoScientific). Toute trace potentielle d'ADN dans les échantillons d'ARN purifié a été éliminée par traitement DNase en suivant le protocole fournisseur du kit TURBO DNA-free (Ambion). Les ARN ont été rétro-transcrits en ADN complémentaire en suivant le protocole fournisseur du kit M-MLV Reverse Transcriptase (Promega) avec des hexamères aléatoires (Roche) comme amorces en suivant le protocole fournisseur. Les ADNc ont été conservés à -80°C jusqu'à amplification PCR.

### *d) Séquençage MiSeq*

La diversité bactérienne a été analysée par séquençage Illumina MiSeq en séquençant les régions hypervariables V3 et V4 du gène codant l'ARNr 16S à la plateforme de séquençage GenoToul Genomics and Transcriptomics (GeT PlaGe, Auzeville, France). La région V3-V4 du gène de l'ARNr 16S a été amplifiée par PCR en utilisant des amorces bactériennes 343F et 784R modifiées (incorporant des adaptateurs de la PCR2 décrite plus loin):

343F = 5'-CTT-TCC-CTA-CAC-GAC-GCT-CTT-CCG-ATC-TAC-GGR-AGG-CAG-CAG-3'

784R = 5'-GGA-GTT-CAG-ACG-TGT-GCT-CTT-CCG-ATC-TTA-CCA-GGG-TAT-CTA-ATC-CT-3'

La PCR1 a été réalisée dans 50 µL de milieu réactionnel à 1X tampon et 2,5 U de MTP Taq DNA Polymerase (Sigma), 0,2 µL de chaque dNTP, 0,5 µM de chaque amorce et 2 ng d'ADN génomique (ou 0,1 ng d'ADNc). L'amplification a été réalisée en 30 cycles (1 min : 94°C ; 1 min : 65°C ; 1 min : 70°C). Les produits de PCR obtenus ont été congelés et transférés à la plateforme de séquençage GeT-PlaGe. Après purifications sur billes magnétiques, une 2<sup>ème</sup> PCR (permettant d'ajouter les

adaptateurs de séquençage Illumina ainsi qu'un index unique pour chaque échantillon) a été réalisée en utilisant les amorces suivantes :

FP2=5'-AAT-GAT-ACG-GCG-ACC-ACC-GAG-ATC-TAC-ACT-CTT-TCC-CTA-CAC-GAC-3'

RP2=5'-CAA-GCA-GAA-GAC-GGC-ATA-CGA-GAT-index-GTG-ACT-GGA-GTT-CAG-ACG-TGT-3'

La PCR2 a été réalisée dans 50 µL de milieu réactionnel à 1X tampon et 2,5U de MTP Taq DNA Polymerase (Sigma), 0,2 µL de chaque dNTP, 0,5 µM de chaque amorce et 15 ng de produit purifié issu de la PCR1. L'amplification a été réalisée en 12 cycles (1 min : 94°C ; 1 min : 65°C ; 1 min : 72°C). Les produits de PCR obtenus ont à nouveau été purifiés en utilisant des billes magnétiques avant d'être quantifiés par un spectrophotomètre Nanodrop 1000. La qualité et la mesure de quantité des échantillons de faible concentration au Nanodrop ont été vérifiées en utilisant un BioAnalyzer 2100 et un kit High Sensivity DNA Analysis (Agilent). La librairie de séquençage a été préparée par mélange équimolaire de tous les échantillons à séquencer (~300 par run de séquençage), et dosée par qPCR. Le séquençage a été réalisé à la plateforme de séquençage GenoToul Genomics and Transcriptomics (GeT, Auzeville, France) en chargeant 1 mL de librairie à 7 pM sur une cartouche Illumina MiSeq et un kit de réactifs v2 (*pair-end* 2x250pb). Les séquences passant les filtres de qualité interne au séquenceur ont été démultiplexées (attribuées à leur échantillon d'origine) en se basant sur les index attribués à chaque échantillon en PCR2.

### **e) FROGS**

Les données issues de séquençage ont été analysées à l'aide des outils intégrés dans FROGS (v3.0, Escudié et al., (2017)) en suivant les recommandations des auteurs. Les séquences R1 (forward) et R2 (reverse) ont été combinées en utilisant VSEARCH intégré dans FROGS (Rognes et al., 2016), avec un seuil de mésappariement de 10%. Les séquences obtenues ont été filtrées pour ne garder que les séquences d'une longueur de 380 à 460 paires de bases (pb), possédant les séquences des amorces (sans erreur tolérée) et exemptes de bases indéterminées (N). Le « clustering » a été réalisé avec une distance d'agrégation de  $d=3$  après étape de *denoising* en utilisant Swarm intégré dans FROGS (Mahé et al., 2014). Les chimères ont été détectées et éliminées en utilisant l'implémentation FROGS de VSEARCH, et les OTUs obtenus ont été filtrés s'ils présentaient moins de X séquences dans l'ensemble du jeu de donnée (valeur X calibrée à l'aide d'une communauté synthétique de composition connue ajoutée au run de séquençage : 10 séquences dans le cas du

chapitre 5) ou avec une valeur seuil de 0.005% dans le cas du chapitre 3 (Bokulich et al., 2013). L'assignation taxonomique des séquences représentatives de chaque OTU a été réalisée par Blast en utilisant la base de données Silva 132 avec un seuil de *pintail score* fixé à 100 (Quast et al., 2013).

### **f) Visualisation et analyse statistique des données ADNr 16S**

L'analyse statistique des données a été effectuée sur R (v3.5.0). La table d'abondance a été manipulée en utilisant les fonctions de base de R ainsi que les packages *dplyr* et *tidyr*. Les données ont été normalisées en suivant les recommandations de Cao et al., (2016), à savoir :

- Une normalisation par Total Sum Scaling (division de chaque comptage par le total des comptages par échantillon), correspondant au passage en abondance relative et permettant de tenir compte des différences de profondeur de séquençage entre échantillons ;
- Agrégation des OTUs au niveau taxonomique désiré par somme des abondances ;
- Agrégation des taxons de faible abondance en une catégorie « Autres » ;
- Transformation CLR (*Centered Log Ratio*) pour tenir compte de la compositionnalité des données TSS, avec un *offset* de 1 (constante ajoutée pour éviter la présence de zéros au moment de la transformation log).

Les indices de diversité (richesse, Shannon et inverse de Simpson) et la distance de Bray-Curtis entre échantillons ont été calculés en utilisant le package R *vegan* (v2.5-7), ainsi que les ordinations NMDS (*Non-Metric Multidimensional Scaling*) en utilisant la fonction *metaMDS*. Les arbres de clustering hiérarchique ont été obtenus en utilisant les distances de Bray-Curtis entre échantillons ainsi obtenues et la fonction R *hclust*. Les courbes, diagrammes en bâtons (*barplots*) et « boîtes à moustache » (*boxplots*) ont été générées à l'aide du package *ggplot2* (v3.3.5) et les courbes lissées ont été tracées avec la fonction *stat\_smooth*. La significativité des différences entre échantillons (ou groupes) a été évaluée par test non-paramétrique de Wilcoxon, les p-valeurs obtenues ont été ajustées par la méthode FDR (*False Discovery Rate*) à l'aide du package *ggpubr* (v0.4.0). Enfin, l'analyse longitudinale intégrative des taxons et des acides gras volatils a été réalisée à l'aide du package *timeOmics* (v1.4.0) en suivant les recommandations par défaut (Bodein et al., 2019).

## **2. Caractérisation de la diversité fonctionnelle et taxonomique des protéines par analyse métaprotéomique**

L'analyse de diversité fonctionnelle et taxonomique des protéines exprimées par analyse métaprotéomique s'est déroulée en quatre étapes principales i) extraction et séparation des protéines, ii) digestion en peptides et purification de ces derniers, iii) séparation et détermination des peptides par chromatographie liquide suivi d'une spectrométrie de masse en tandem (LC-MS/MS) et iv) analyse et visualisation des données. Dans le cadre du doctorat d'A. Lazuka, ce protocole avait été défini en collaboration avec le laboratoire partenaire, le Centre Helmholtz for Environmental research - UFZ Leipzig en Allemagne, puis transféré en partie au LISBP à Toulouse pour les étapes 1, 2 et 4. Après transfert des procédures d'extraction, séparation, digestion et purification en France, les échantillons étaient toutefois analysés à l'UFZ (LC-MS/MS et affiliation contre la base des données peptidiques), puis les données étaient transférées et traitées par nos soins.

### ***a) Echantillonnage, extraction et séparation des protéines***

L'extraction des protéines était effectuée en utilisant un tampon phénol comme solvant et selon la procédure dédiée aux échantillons complexes de sédiments détaillée par Lin et al., (2016). Pour chaque échantillon, trois répliquats techniques ont été prélevés et stockés à -80°C jusqu'à extraction. Pour l'extraction, 3 mL d'échantillon homogénéisé ont été resuspendus dans 5 mL de tampon SDS (SDS 1.25%, DTT 20 mM, Tris-HCl 0.1 M pH 6.8) et le mélange a été agité doucement pendant 1 heure à température ambiante. Ensuite, une sonication douce a été effectuée sur glace (2x40 secondes avec 1 minute de repos – sonde MS72, puissance de 40 W). Une centrifugation (7197 x g, 10 min, 4°C) a permis de récupérer le surnageant contenant les protéines, alors mélangé avec 3 mL de solvant phénol (10 g de phénol dans 1 mL d'eau ultrapure – pH 8.0) et incubé 15 minutes à température ambiante sous agitation douce. Le surnageant était alors récupéré (centrifugation 7197 x g, 10 min, 4°C) et l'étape précédente répétée. Les protéines contenues dans le surnageant obtenu après double extraction au phénol étaient précipitées sur la nuit à -20°C par ajout de cinq volumes d'acétate d'ammonium 0,1 M dissout dans du méthanol. Après précipitation, les protéines étaient récupérées par centrifugation (7197 x g, 20 min, 4°C) et les culots étaient rincés deux fois au méthanol froid (1 mL) et une fois à l'acétone froid (1 mL). Chaque étape de rinçage incluait une agitation par vortex et une élimination du surnageant après centrifugation (12000 x g, 10 min, 4°C). Finalement, les échantillons étaient séchés sous hotte à



flux laminaire avant séparation par SDS-PAGE. Pour la séparation, les culots secs étaient alors dissouts dans 100  $\mu$ L de tampon de charge Laemmli (2% SDS, 2 mM  $\beta$ -mercaptoethanol, 4% glycerol, 40 mM Tris-HCl pH 6.8, 0.01% bromophenol blue) et chauffés 5 minutes à 90°C. Le gel de séparation par SDS-PAGE se répartissait en un gel de collection (environ 1,5 cm – 4% polyacrylamide) et le gel de séparation (12% polyacrylamide). Les échantillons ont été chargés sur le gel (10  $\mu$ L) avec deux étalons de poids moléculaire connu (Bio-Rad) et l'électrophorèse a été effectuée à 20 mA pendant environ 30-40 minutes jusqu'à migration d'environ 1 cm dans le gel de séparation. Le gel a été coloré au bleu de Coomassie (Coomassie brilliant blue G-250 - Merck) et chaque bande (correspondant à un échantillon) a été découpée en petits cubes (d'environ 1 mm<sup>2</sup>) qui ont été transférés en Eppendorf de 0,5 mL avant digestion à la trypsine.

### ***b) Digestion des protéines et purification des peptides***

Pour chaque échantillon, la digestion à la trypsine s'effectuait sur gel. Pour cela les petits morceaux de gels étaient décolorés trois fois par ajout de 200  $\mu$ L de bicarbonate d'ammonium (10 mM) dissout dans un mélange eau/acétonitrile (60/40 v/v), puis finalement lavés avec 200  $\mu$ L d'acétonitrile pendant 5 minutes. Les morceaux de gels lavés étaient alors séchés au Speed-Vac, puis les protéines contenues étaient réduites au dithiothreitol (DTT) (30 minutes d'incubation à température ambiante dans une solution de 60  $\mu$ L à 10 mM de DTT) et alkylées à l'iodoacetamide (IAA) (30 minutes d'incubation à température ambiante et au noir dans une solution de 60  $\mu$ L à 100 mM d'IAA). Par la suite, chaque échantillon a été déshydraté à l'acétonitrile (200  $\mu$ L, 5 minutes, agitation douce) puis lavés au bicarbonate d'ammonium (200  $\mu$ L à 10 mM, 10 minutes, agitation douce) et enfin une dernière fois désydraté à l'acétonitrile et séché au Speed-Vac. La digestion à la trypsine a été effectuée sur la nuit à 37°C en suivant les recommandations du fournisseur (Promega, Madison, WI, USA), puis les morceaux de gels ont été lavés pour extraire les peptides avec 50  $\mu$ L de bicarbonate d'ammonium à 5 mM pendant 10 minutes à température ambiante. La phase liquide contenant les peptides a été collectée et l'extraction des peptides a été répétée avec 50  $\mu$ L d'un mélange acétonitrile/eau/acide formique à 85% (50/44/0.06 v/v/v). Les peptides extraits ont été complètement séchés au Speed-Vac (3 heures) et désalés au moyen de pointes équipées de résine de séparation des peptides ZipTip- $\mu$ C18 tips (Millipore, Billerica, MA, USA). Enfin, les peptides désalés ont été récoltés dans une solution d'acétonitrile (80%) et d'acide formique (0,1%), évaporés au Speed-Vac et le lysat peptidique a été resuspendu dans une solution d'acide formique à 0,1% pour l'analyse par LC-MS/MS.

***c) Chromatographie liquide et spectrométrie de masse en tandem***

Les lysats peptidiques ont été séparés par chromatographie ultra haute performance (UHPLC - Ultimate 3000, Dionex/Thermo Fisher Scientific, Idstein, Germany). Un volume de 5  $\mu$ L d'échantillon a été chargé sur la pré-colonne ( $\mu$ -precolumn, Acclaim PepMap, 75  $\mu$ m de diamètre interne, 2 cm, C18, Thermo Scientific) puis séparé et élué de la colonne analytique (PepMap Acclaim C18 LC Column, 25 cm, taille de particule de 3  $\mu$ m, Thermo Scientific).

Pour la pré-colonne, la phase mobile était constituée d'un mélange à 4% (v/v) d'une solution A à 80% d'acétonitrile et 0,08% d'acide formique, et à 96% d'une solution B à 0,01% d'acide formique, toutes deux préparées en solution d'eau nanopure. Pour la colonne analytique et l'éluion des peptides, un gradient non linéaire d'acétonitrile était effectué à partir des solutions A et B, augmentant la fraction de la phase B de 4 à 55% en 150 minutes.

La spectrométrie de masse (MS) a été effectuée sur un appareillage Q Exactive HF MS (Thermo Fisher Scientific, Waltham, MA, USA) équipé d'une source TriVersa NanoMate (Advion, Ltd., Harlow, UK) couplée à la UHPLC. La MS était paramétrée en mode HCD (higher energy collision dissociation) à une énergie de collision normalisée de 30%. Les scans ont été mesurés avec une résolution de 120000 dans la gamme de 350-1600 m/z. La cible de quantification des ions était fixée à 3.106 avec un temps d'injection de 80 ms. Pour la fragmentation MS/MS, les ions étaient isolés sur le quadrupole avec une fenêtre de 1,6 Da et mesurés avec une résolution de 15000 dans la gamme de 200-2000 m/z. La durée d'exclusion dynamique était fixée à 30 secondes avec une tolérance de 10 ppm, alors que le contrôle du gain automatique était de 2.105 avec un temps d'injection de 120 ms et un ratio de sous-remplissage de 1% (Starke et al., 2017).

***d) Identification des spectres et création des groupes de protéines sous Proteome Discoverer***

Les données brutes ont été traitées avec le logiciel Proteome Discoverer (v1.4.1.14; Thermo Fisher Scientific, Waltham, MA, USA). A partir des spectres acquis en MS/MS, des recherches peptidiques ont été effectués sous l'algorithme Sequest HT (Eng et al., 2008) contre la base de données UniprotKB/TrEMBL (mise à jour le 6 Avril 2016). Les paramètres de recherche fixés étaient i) le clivage à la trypsine avec une permission de sites de coupure manqués de deux erreurs maximum, ii) une tolérance de 10 ppm et 0,05 Da sur les ions peptidiques, iii) l'oxydation des résidus

méthionine et la carbamylation des résidus lysines et arginines et iv) la carbamidométhylation des résidus cystéines. De plus, seuls les peptides avec un taux de fausse découverte (false discovery rate -FDR) inférieur à 1% (calculé avec l'algorithme Percolator, Käll et al., (2007)) et un rang peptidique de 1 ont été considérés. Ces recherches peptidiques ont abouti à la génération de groupes de protéines (par la suite appelés protéines pour des raisons de clarté).

Les peptides ont été filtrés en conservant uniquement les peptides de « haute confiance » (*high peptide confidence* sous Proteome Discoverer) et ont été gardées seulement les protéines différenciables et distinctes.

### ***e) Identification taxonomique et fonctionnelle des groupes de protéines avec le pipeline Prophan***

Par la suite, le pipeline Prophan (Schneider et al., 2011) a été utilisé en libre accès pour déterminer la classification taxonomique (basée sur les bases de données NCBI et BlastP) et la prédiction fonctionnelle par catégorisation des clusters de groupes orthologues (Cluster of Orthologous Groups - COG) avec RPSBLAST. Seules les protéines identifiées ont été conservées et pour chaque réplicat le facteur d'abondance spectrale normalisé (normalized spectral abundance factor - NSAF) a été calculé pour compenser le biais quantitatif induit par la différence de taille entre protéines (Zybailov et al., 2006). Le traitement des triplicats consistait par la suite à ne conserver que les protéines présentes dans au moins 2 réplicats sur 3. En d'autres termes, les protéines présentes dans seulement un triplicat étaient considérées comme des faux positifs et écartées, les autres étant conservées et considérées selon leur valeur moyenne. Les échantillons des duplicats biologiques (réacteurs) étaient alors regroupés par temps d'incubation (4 temps d'échantillonnage considérés par réacteurs), générant moyenne et écart-type pour chaque protéines.

### ***f) Recherche des CAZymes***

Pour obtenir plus d'informations sur la structure du metaprotéome, les protéines identifiées ont été soumises à l'assignation des familles de CAZy via le pipeline d'annotation utilisé pour la mise à jour quotidienne de la base de données CAZy comme décrit précédemment (Svartström et al., 2017; Delannoy-Bruno et al., 2021).

### **g) Visualisation et analyse statistique des données métaprotéomiques**

Les données métaprotéomiques ont toutes été analysées avec l'outil de statistiques R (v3.5.0). La variabilité entre réplicats techniques a été évaluée par une analyse de corrélation de Pearson à l'aide de la fonction R *cor*. Les réplicats techniques avec une corrélation de Pearson inférieure à 0.7 ont été écartés du jeu de données. Les tables d'abondances annotées ont été manipulées à l'aide des packages R *dplyr* et *tidyr*. Les protéines ont été analysées à différents niveaux (taxonomiques, modules CAZy, COG affiliés aux enzymes de production d'AGV) après agrégation au niveau souhaité et leurs abondances NSAF ont été sommées avant transformation CLR (*Centered Log Ratio*) avec un offset de 1. Au moment de l'agrégation taxonomique, les phylums minoritaires représentant moins de 2% d'abondance ont été regroupés dans une catégorie « Autres », ainsi que les genres correspondants. Les genres représentant moins de 2% et appartenant à un phylum non minoritaire ont été regroupés dans les catégories « Phylum X – Autres » correspondant. Les diagrammes en bâtons et « boîtes à moustache » ont été générés à l'aide du package R *ggplot2* (v3.3.5). Les différences entre conditions ont été testées en utilisant des tests de Wilcoxon non paramétriques, ajustés pour comparaisons multiples par la méthode FDR, le tout à l'aide du package R *ggpubr* (v0.4.0). Les ordinations NMDS ont été réalisées avec le package R *vegan* (v2.5-7), ainsi que les analyses de PERMANOVA (*Permutational Multivariate Analysis of Variance* ; 9999 permutations) appliquées aux distances euclidiennes basées sur les données CLR-transformées. Pour les analyses fonctionnelles des familles CAZymes et des protéines en lien avec les acides gras volatils, l'indice d'allocation (AI) a été défini comme le rapport entre l'abondance relative genre-spécifique d'une fonction donnée (par exemple une famille CAZyme) et l'abondance relative totale de ce genre. Autrement dit, l'indice d'allocation reflète pour un genre donné la proportion de ses protéines consacrée à la fonction étudiée. Une variation significative d'AI reflète ainsi un changement dans l'expression des protéines en lien avec cette fonction. Enfin, les analyses multivariées (Analyses en composantes principales, ACP et Analyse discriminante par les moindres carrés partiels, PLS-DA) ont été réalisées en utilisant le package R *mixOmics* (v6.12.2, Rohart et al., (2017)) à partir des données transformées CLR.



## III. Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

### A. Introduction

Ce chapitre vise à élucider le fonctionnement des consortia microbiens lignocellulolytiques, à comprendre les rôles fonctionnels des espèces les constituant et à identifier leurs complémentarités. Pour cela, nous avons étudié par une approche métabolomique deux consortia microbiens précédemment enrichis dans notre laboratoire. L'analyse métabolomique a été initiée lors du doctorat d'Adèle Lazuka et a été réalisée en partenariat avec l'UFZ de Leipzig (Centre de Recherche Environnementale Helmholtz, Leipzig, Allemagne), plus particulièrement avec l'équipe de métabolomique environnementale et de biologie des microbiomes du Dr. Nico Jehmlich, intégrée au Département de biologie moléculaire des systèmes dirigé par le Prof. Dr. Martin von Bergen. L'annotation des modules CAZymes a été réalisée par le Dr. Vincent Lombard et le Dr. Bernard Henrissat du Laboratoire d'Architecture et Fonction des Macromolécules Biologiques (AFMB, Marseille, France).

Les deux consortia microbiens étudiés ici, dénommés RWS et TWS (pour « Rumen Wheat Straw » et « Termite Wheat Straw ») sont dérivés de deux inocula d'origine différente : le rumen bovin et l'intestin de termite (*Nasutitermes ephratae*). Les deux consortia ont été cultivés, enrichis et stabilisés sur de la paille de blé comme seule source de carbone, en absence d'oxygène et en inhibant l'étape de méthanogénèse de la digestion anaérobie. Leur capacité à dégrader la paille de blé en acides gras volatils, ainsi que la diversité bactérienne de ces communautés déduite à partir du séquençage du gène codant l'ARNr 16S ont été décrites précédemment (Lazuka et al., 2015, 2018). Cependant, même si l'étude taxonomique de ces deux communautés bactériennes a permis de révéler qu'elles étaient relativement proches au niveau de leur composition taxonomique (genres majoritaires identiques), des différences compositionnelles et des variations dans les dynamiques des populations existent au cours de la dégradation. L'analyse des séquences de l'ARNr 16S a permis d'émettre des hypothèses concernant leur fonctionnement, notamment la complémentarité fonctionnelle entre *Bacteroides* et *Clostridium*, en s'appuyant sur les

### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

---

connaissances disponibles sur ces espèces. L'étude de la dynamique de ces consortia par analyse métabolomique fut une solution pour valider ces hypothèses fonctionnelles.

Pour caractériser la dynamique fonctionnelle des communautés RWS et TWS au cours de la bioconversion de la paille de blé non prétraitée, des bioréacteurs ont été menés en duplicat biologique pour chacun des consortia. Sur ces expériences, quatre temps d'incubation ont été sélectionnés présentant des niveaux croissants de dégradation de la biomasse, équivalents entre les deux systèmes. Chaque temps choisi a été analysé en triplicat technique pour les deux replicats biologiques de chaque consortium, résultant en un total de 48 échantillons analysés en chromatographie liquide couplée à la spectrométrie de masse en tandem. Le protocole mis en place consistait en i) l'extraction des protéines contenues dans les échantillons et leur séparation sur gel, ii) la digestion des protéines en peptides et leur purification, iii) la détermination de ces derniers par LC-MS/MS, iv) l'identification des spectres obtenus contre la base de données UniprotKB/TrEMBL et la création de « *protein group* », v) le calcul de l'abondance et l'annotation fonctionnelle et taxonomique de chaque « *protein group* », vi) l'annotation des modules CAZy et, vii) le traitement, l'analyse statistique et la visualisation des données obtenues.

L'analyse du profil d'expression des protéines impliquées dans la dégradation des carbohydrates (*Carbohydrate Active enZymes* et *Carbohydrate Binding Modules*) chez RWS et TWS, nécessaires pour l'hydrolyse de la lignocellulose, devrait nous aider à mieux comprendre les mécanismes d'hydrolyse mis en œuvre par ces communautés bactériennes enrichies sur la paille de blé. Elle devrait notamment permettre de confirmer la complémentarité hydrolytique entre les deux genres majoritaires *Bacteroides* et *Clostridium*, comme suggérée par les résultats précédents issus de l'analyse 16S. De la même façon, l'analyse métabolomique devrait permettre le suivi des enzymes impliquées dans la production d'acides gras volatils (acétate, butyrate, propionate), et expliquer les différences macro-cinétiques observées entre les deux consortia microbiens (vitesse et % maximal de dégradation de la paille, profil et vitesse de production des AGV).

## **B. Horizontal metaproteomics and CAZymes analysis of lignocellulose decomposition by microbial consortia derived from cow rumen and termite gut**

Emeline Auer<sup>1</sup>, Adèle Lazuka<sup>1</sup>, Nico Jehmlich<sup>2</sup>, Sébastien Déjean<sup>3</sup>, Vincent Lombard<sup>4,5</sup>, Bernard Henrissat<sup>4,5,6</sup>, Michael O'Donohue<sup>1</sup>, Guillermina Hernandez-Raquet<sup>1\*</sup>

<sup>1</sup>) Toulouse Biotechnology Institute - TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, F-31077 Toulouse, France.

<sup>2</sup>) Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, Germany.

<sup>3</sup>) Toulouse Mathematics Institute, UMR 5219 CNRS, Toulouse University, Toulouse, France

<sup>4</sup>) CNRS UMR 7257, Aix-Marseille University, F-13288, Marseille, France.

<sup>5</sup>) USC 1408 AFMB, INRAE, F-13288, Marseille, France.

<sup>6</sup>) Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

AUER Emeline emeline.auer@inrae.fr

LAZUKA Adèle adele.lazuka@live.fr

JEHMLICH Nico nico.jehmlich@ufz.de

DEJEAN Sebastien sebastien.dejean@math.univ-toulouse.fr

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

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

O'DONOHUE Michael michael.odonohue@inrae.fr



## 1. Abstract

### Background

Microbial consortia derived from natural environments can be used for biomass bioconversion to produce biofuels and commodity chemicals. The functional potency of microbial consortia can be enhanced using enrichment, targeting a specific function. This strategy reduce microbial biodiversity and amplifies the community fraction, for example, responsible for lignocellulose degradation. Studying enriched microbial consortia is an interesting approach to decipher the specific roles of community members and investigate the temporal dynamics of proteins involved in biomass degradation.

### Results

Using metaproteomics, 16S rRNA gene amplicon sequencing and multivariate discriminant analysis, we investigated the biological mechanisms underlying anaerobic transformation of lignocellulose by two robust microbial consortia enriched from cow rumen (RWS) and *Nasutitermes* termite gut (TWS) microbiomes. This revealed at unprecedented depth the temporal dynamics of proteins in both anaerobic consortia when growing on unpretreated wheat straw and producing carboxylates. The 16S rRNA diversity and metaproteomics analyses revealed a dominance of Bacteroidetes, Firmicutes and Proteobacteria phyla, displaying functional complementarities, irrespective of the inoculum origin. The metaproteomics led to the identification of 10,342 protein groups (referred to as proteins), 423 of which contained at least one carbohydrate-active enzyme (CAZyme) domain, including 40 families of glycoside hydrolases involved in biomass deconstruction. Furthermore, 346 proteins could be assigned to short chain fatty acids biosynthesis functions. Multivariate partial least squares discriminant analysis of the CAZyme profiles of the two consortia revealed key similarities during the early stage of lignocellulose degradation, with overall function in both cases being the utilization of easily accessible oligosaccharides. In contrast, at later stages of the biomass degradation, the CAZymes profiles were clearly different between RWS and TWS. Our results provide important insights into

the specific functional roles and complementarities that exist between Bacteroidetes and Firmicutes members of these lignocellulolytic consortia enriched from cow rumen and termite gut in anaerobic bioreactors and their potential utilization for biorefinery applications.

### **Conclusions**

Combining 16S rRNA gene amplicon sequencing, metaproteomics and multivariate discriminant analysis, we described the main features of wheat straw degradation by microbial consortia enriched from cow rumen and termite gut in anaerobic bioreactors. Our results provide important insights into the specific functional roles and complementarities that exist between Bacteroidetes and Firmicutes members of these lignocellulolytic microbial consortia and their potential utilization as a source of enzyme cocktails for biorefinery applications.

### **Keywords**

Lignocellulose bioconversion; anaerobic consortium; carboxylate production; CAZyme; metaproteomics; termite gut; cow rumen

## 2. Introduction

Lignocellulose (LC) is the major component of plant cell walls and the primary terrestrial source of renewable carbon. Therefore, lignocellulose decomposition represents a major carbon recycling process in the biosphere. In natural ecosystems, lignocellulose recycling is performed by complex microbial communities that produce appropriate enzyme arsenals to achieve this. Examples of biomass deconstruction by microbial communities include the composting process (Liu et al., 2015; Wang et al., 2016), cellulose decomposition in soils (Eichorst and Kuske, 2012; López-Mondéjar et al., 2016b), leaf-litter decomposition (Schneider et al., 2012), and anaerobic digestion in the rumen of herbivores such as cows (Hess et al., 2011) and termites (Warnecke et al., 2007; Dumond et al., 2021). In the case of herbivores, biomass decomposition is achieved through a complex symbiosis between the animal and its digestive microbiome, which provides nutrients to the host in the form of short-chain volatile fatty acids (VFA) such as acetate, butyrate, propionate, and microbial proteins. In the technosphere, lignocellulose is also of great interest because it constitutes the most promising raw material for biofuels and commodity chemicals production. However, in the lignocellulose biorefinery schemes biomass hydrolysis is very challenging, because lignocellulose components are locked into very stable, insoluble heteropolymeric matrix consisting mainly of cellulose, hemicellulose and lignin (Cragg et al., 2015). Therefore, to improve the economic feasibility of lignocellulose biorefining, new more effective strategies are required to breakdown lignocellulosic matrix and release fermentable sugars. To achieve this, it is relevant to look at natural lignocellulose degrading ecosystems, since these can provide Nature-inspired solutions deployable in controlled bioreactor conditions.

In Nature, biological biomass deconstruction is a complex process that marshalls a large diversity of enzymes that are rarely produced by a single microbial species (Singh et al., 2014). A major category of lignocellulose-degrading enzymes are carbohydrate-active enzymes (CAZymes), which form the most abundant category of enzymes in living organisms. More specifically, CAZymes designate all enzymes that break down, modify or build glycosidic bonds (Lombard et al., 2014). In the CAZy classification system, these enzymes are divided into 5 main categories, namely glycoside hydrolases (GH), glycosyltransferases (GT), carbohydrate esterases (CE), polysaccharide lyases (PL) and auxiliary activities (AA), and their appended non-catalytic carbohydrate-binding modules (CBM) (Lombard et al., 2014). Other relevant proteins domains include cohesins (COH) and dockerins (DOC) that are necessary to incorporate CAZymes into large complexes such as cellulosomes and S-layer homology (SLH) domains that link enzymes to the cell surfaces (Bayer et

al., 2008). Microorganisms present in lignocellulose degrading environments such as the digestive tract of termites or cow rumen provide vast reservoirs of CAZymes, particularly glycoside hydrolases (GH), which are sources of new potentially potent lignocellulolytic enzyme cocktails for biomass deconstruction (Scharf, 2015).

To better understand the biological mechanisms that underpin lignocellulose degradation, metagenomics technologies (metagenomics, metatranscriptomics, metaproteomics, metabolomics) are being used to characterize both the structure and function of lignocellulose-active microbial communities. In particular, such investigations aim at elucidating how microbial symbiosis and enzymatic synergy contribute to LC deconstruction (Rosnow et al., 2017). 16S rRNA gene amplicon sequencing is used to probe the taxonomical composition of microbial communities and provide the means to better pinpoint the microbial populations that are more specifically involved in biomass deconstruction. In this way, diversity analysis of complex microbial ecosystems such as rumen, soils and compost have revealed cooperative behavior between various microbial species for polysaccharide hydrolysis (Eichorst et al., 2013; Deusch et al., 2017). Likewise, shotgun metagenomics has provided information on both the genetic, metabolic and hydrolytic potential of various lignocellulolytic microbial communities (Warnecke et al., 2007; Brulc et al., 2009; Hess et al., 2011; Singh et al., 2014) and on the synergistic action between members of these communities (Wang et al., 2016). However, these approaches do not inform on the actual metabolic activities of community members, nor do they provide details of the effective role of genes in ecosystem functioning. To overcome this constraint, metaproteomics (defined “as the large-scale characterization of the entire protein component of microbial communities at a given point in time”) is used to link proteins to specific microorganisms and correlate their presence with metabolic activity (Wilmes and Bond, 2004). Additionally, this powerful approach provides both taxonomic and functional information that are useful to reveal metabolic networks and symbiotic interactions that underpin the lignocellulose-deconstruction potential of microbial ecosystems.

Previous metaproteomics studies on bovine rumen ecosystems have shown that the most abundant proteins are affiliated to Bacteroidetes, Firmicutes and Proteobacteria (Hart et al., 2018). Considering the main metabolic pathways expressed by these phyla, it is possible to conclude that lignocellulose degradation is associated with high redundancy of key enzyme activities, an observation that is consistent with the high taxon diversity. Regarding the termite gut microbiome, metaproteomic analysis of *Nasutitermes corniger* showed that among 197

identified proteins with known functions, 48 proteins are directly related to glycoside hydrolysis (Burnum et al., 2011). However, partly due to the high complexity of these natural ecosystems and the limited number of proteins detected, these metaproteomic studies neither revealed the specific roles of individual microbial taxa, nor the temporal dynamic of proteins involved in lignocellulose breakdown.

To gain insight into how lignocellulolytic ecosystems function, ecosystem enrichment under controlled conditions is often used to reduce microbial complexity and hone community functions for use in bioprocesses (Lazuka et al., 2015; Cortes-Tolalpa et al., 2016; Lazuka et al., 2018; Tomazetto et al., 2020). Although interesting results have been obtained, there is still a sparsity of information related to microbial dynamics and interplay in these simplified systems and very little information is available to describe the diversity and temporal dynamics of CAZymes involved in lignocellulose deconstruction. As mentioned, metaproteomics is a good option to study proteins produced by enriched lignocellulolytic microbial communities. However, previous studies have revealed only a small number of proteins (D'haeseleer et al., 2013; Jiménez et al., 2015, 2016; Alessi et al., 2017; Tomazetto et al., 2020), with measurements being mostly limited to the end-time point of the bioprocess. Nevertheless, one study did consider protein expression dynamics, but only focused on a subset of enzymes (Zhu et al., 2016). Clearly, biological lignocellulose degradation is a dynamic process that involves changes in both taxonomic and enzyme diversity. Therefore, to fully elucidate how microbial communities degrade lignocellulose, it is vital to fully capture the temporal dynamics of all expressed proteins.

In previous studies, we reported the enrichment of two robust anaerobic lignocellulolytic microbial consortia, the first derived from cow rumen (RWS) and the second, from the termite gut microbiome of *Nasutitermes ephratae* (TWS) (Lazuka et al., 2015, 2018). These anaerobic consortia were enriched and simplified by a sequential batch reactor process, using untreated wheat straw as sole carbon source. The resultant communities displayed high lignocellulolytic activity and good ability to produce carboxylates (mainly VFA such as acetate, butyrate and propionate), which represent target compounds for the production of added-value chemicals (e.g. polyhydroxyalkanoate bioplastics) and liquid biofuels (Agler et al., 2011; Torella et al., 2013). The kinetic characteristics of these enriched consortia were determined by measuring lignocellulose degradation rate, xylanase activity and carboxylate (VFA) production. Herein, we expand on previous work, performing shotgun metaproteomic analysis on wheat straw hydrolysis reactions

catalysed by RWS and TWS consortia, respectively. To improve on previous studies, we sampled at different time points over the reaction period. Moreover, to better understand the different functional roles and complementarities of microbial community members, we have correlated kinetics and microbial diversity analysis.

### **3. Materials and methods**

#### ***a) Lignocellulose degradation by RWS and TWS microbial consortia***

The kinetic behaviors of the enriched lignocellulolytic microbial communities derived from cow rumen (RWS) and termite gut (TWS), summarized in Table III-1, have been described previously (Lazuka et al., 2015, 2018). Two identical anaerobic bioreactors were carried out for each consortium, using a mineral medium containing wheat straw (variety Koreli, milled to 2 mm) as the sole carbon source (20 g.L<sup>-1</sup>). Bioreactors were operated under mesophilic conditions (35°C) under stirring (400 rpm) and pH control (6.15), as detailed in Additional file 1: Supplementary Materials and Methods.

To assess the temporal dynamics of microorganisms and expressed proteins in RWS and TWS along lignocellulose degradation, 16S rRNA gene sequencing and shotgun metaproteomic analysis were performed on four time points for each biological replicate bioreactor of each consortium. The criteria for time point selection were based on wheat straw degradation, VFA production and xylanase activity profiles (Table III-1). The first point (T1) corresponds to an early phase where xylanase activity and lignocellulose degradation are low. Time points 2 and 3 (T2 and T3 respectively) correlate with the start and end of maximal xylanase activity and peak lignocellulose degradation rate. Finally, the fourth time point (T4) captures the final phase, when wheat straw degradation, VFA production and xylanase activity stagnate.

**Table III-1 : RWS and TWS sample characteristics**

	Dry matter degradation (%)	VFA (mgC/L)	Acetate:Propionate:Butyrate (C molar ratio)	Xylanase (mUA/L)
RWS-1* T1	7.6 ± 7.1	552.20 ± 23.1	57:24:18	174.6 ± 22.8
RWS-2* T1	10.1 ± 3.2	569.4 ± 42.0	58:25:17	158.4 ± 17.3
RWS-1 T2	14.2 ± 0.5	828.4 ± 56.7	55:25:20	543.0 ± 18.1
RWS-2 T2	20.5 ± 5.4	862.7 ± 20.1	62:24:14	369.8 ± 19.1
RWS-1 T3	33.2 ± 3.1	1408.3 ± 64.4	53:23:24	852.3 ± 48.3
RWS-2 T3	30.5 ± 0.4	1254.6 ± 130.9	64:23:13	865.4 ± 17.9
RWS-1 T4	40.7 ± 1.5	1842.8 ± 46.5	54:23:22	879.6 ± 40.9
RWS-2 T4	45.6 ± 3.8	2174.6 ± 67.4	65:21:14	937.5 ± 58.0
TWS-1 T1	11.3 ± 4.5	344.7 ± 10.4	62:16:22	297.9 ± 167.4
TWS-2 T1	16.3 ± 4.4	359.3 ± 6.3	61:13:26	327.9 ± 56.7
TWS-1 T2	30.7 ± 3.5	1896.1 ± 18.0	36:21:42	2561.4 ± 292.3
TWS-2 T2	31.1 ± 6.5	1455.3 ± 22.7	41:23:36	2955.7 ± 355.0
TWS-1 T3	34.6 ± 3.3	1901.5 ± 8.3	41:22:38	2467.8 ± 251.2
TWS-2 T3	38.9 ± 5.2	1887.3 ± 8.3	45:23:33	2177.5 ± 204.3
TWS-1 T4	46.2 ± 3.1	2131.6 ± 110.3	44:21:35	2070.5 ± 187.6
TWS-2 T4	45.0 ± 3.1	2075.3 ± 14.7	46:22:32	1477.8 ± 129.5

Note: \*1 and 2 indicate the biological replicate bioreactors. Data are means ± SD.



### ***b) Microbial diversity analysis***

Bacterial diversity analysis was performed on the pellet fraction of 1.5 mL samples withdrawn of the duplicate bioreactors of each consortium (13,000 x g, 5 min, 4°C). Simultaneous DNA/RNA extraction followed by DNA purification was done in biological duplicates per sampling point, according to the procedures described previously (Auer et al., 2017). The hypervariable V3-V4 region of the 16S rRNA gene was amplified by Illumina MiSeq sequencing using the GenoToul Genomics and Transcriptomics platform (Toulouse, France) and conditions and primers previously described (Auer et al., 2017). A minimum of 25,000 high quality reads were generated for each sample. Joined-pair reads of all samples were simultaneously imported in Find Rapidly OTUs with Galaxy solution (FROGS) pipeline according to its guidelines (Escudié et al., 2017). Briefly, data preprocessing removed sequences outside a 380-460 bp range, presenting mismatch in primers or N calls. The aggregation distance was set to 3 and, after denoising, clustering was performed using the Swarm algorithm (Mahé et al., 2014). OTUs with a total abundance lower than 0.005% of the overall dataset were discarded (Bokulich et al., 2013) and chimeras were removed using VSEARCH (Rognes et al., 2016) using default FROGS parameters. Taxonomic assignment was performed by Blast using Silva 132 16S with a pintail of 100 as reference database (Quast et al., 2013). A phylogenetic tree was produced using the FROGS dedicated tool and default parameters. R CRAN software (v4.0.0) was used for further analysis on the obtained abundance file. Relative abundances and alpha diversity index were obtained after rarefying randomly all samples to 15,000 sequences using R's Phyloseq package (v1.32.0) (McMurdie and Holmes, 2013). 16S rRNA sequencing data have been deposited to the sequence read archive (SRA) of the National Center for Biotechnology Information (NCBI) under accession number PRJNA729464.

### ***c) Metaproteomics analysis***

#### ***(1) Protein extraction and separation***

For each duplicate bioreactor, three technical replicate samples (3 mL) were collected from four time points and stored at -80°C. Protein extraction was carried out using a phenol buffer following the procedure for complex sediment samples (Lin et al., 2016). For protein extraction, samples were suspended in 5 mL of SDS buffer (SDS 1.25%, DTT 20 mM, Tris-HCl 0.1 M pH 6.8). The mixture was shaken at 140 rpm, 1 h at room temperature. After soft ultra-sonication on ice (2 x 40 sec with 1 min break – MS 72 probe, power 40%) and centrifugation (7,197 x g, 10 min, 4°C), the sample supernatant was mixed with liquid phenol (10 g phenol in 1 mL water, pH 8) and shaken (100 rpm,

15 min, 20°C). The aqueous and organic phases were separated by centrifugation (7,197 x *g*, 10 min, 4°C); the aqueous phase was recovered, and the previous step was repeated. The proteins in the organic phenol phase were precipitated overnight with fivefold volume of 0.1 M ammonium acetate dissolved in methanol at -20°C. Samples were then centrifuged (7,197 x *g*, 20 min, 4°C) and the protein pellet was successively washed in 1 mL ice-cold methanol and twice with 1 mL ice-cold acetone, each washing step included vortexing and subsequent centrifugation (12,000 x *g*, 10 min, 4°C). Finally, samples were dried in a laminar flow hood before SDS-PAGE separation.

To separate the extracted proteins, dried pellets were dissolved in 100 µL buffer solution (2% w/v SDS, 2 mM beta-mercaptoethanol, 4% v/v glycerol, 40 mM Tris-HCl pH 6.8, 0.01% w/v bromophenol blue) and heat-treated at 90°C for 5 min. Ten microliters of this mixture were loaded on SDS-PAGE (4% stacking gel, 12% separating gel) with Bio-Rad's molecular weight electrophoresis standards (Bio-Rad, Munich, Germany) and electrophoresis was performed until the sample had migrated 1 cm into the separating gel. The PAGE gel was stained with colloidal Coomassie brilliant blue G-250 (Merck, Darmstadt, Germany) before each lane was cut into small squares (~ 1 mm<sup>2</sup>) and transferred into 0.5 mL Eppendorf tubes.

## (2) Proteolytic digestion with trypsin and purification

For each sample, the pool of gel pieces was destained three times in 200 µL of 10 mM ammonium bicarbonate in water/acetonitrile (60/40 v/v), washed with 200 µL acetonitrile and incubated at room temperature for 5 min with shaking at 600 rpm. Each sample pool was dried using a SpeedVac concentrator (Eppendorf AG, Hamburg, Germany). Subsequent steps were carried out at room temperature. The protein lysates were reduced with 60 µL 10 mM dithiothreitol (DTT) solution (shaking at 600 rpm for 30 min) and alkylated with 60 µL 100 mM iodoacetamide (IAA) (30 min in the dark under agitation at 600 rpm). Each sample was washed with 200 µL acetonitrile for 5 min at 600 rpm, washed with 10 mM ammonium bicarbonate (10 min, shaking at 600 rpm) and finally washed with 200 µL acetonitrile (5 min, shaking at 600 rpm). After dehydration of gel pieces with acetonitrile washing and drying using a SpeedVac, trypsin proteolysis (Promega, Fitchburg, WI, USA) was performed overnight at 37°C. The pieces were washed with 50 µL 5 mM ammonium bicarbonate solution for 10 min at 800 rpm. The liquid phase (with peptides) of each sample was collected in a 0.5 mL Eppendorf tube. Gel pieces were rewashed with 50 µL acetonitrile/water/85% formic acid solution (50/44/0.06 v/v/v) and the liquid phase was withdrawn. Extracted peptides were completely dried using a SpeedVac for 180 min, and desalted

using ZipTips (C18, Merck, Millipore, Billerica, MA, USA) with 80% acetonitrile/0.1% formic acid solution. The samples were dried using a SpeedVac and stored at -20°C.

### (3) Liquid chromatography-tandem mass spectrometry analysis

Before LC/MS-MS analysis, the samples were resuspended in 0.1% formic acid. Peptides lysates were separated on a UHPLC system (Ultimate 3000, Dionex/Thermo Fisher Scientific, Idstein, Germany). In total, 5 µL samples were first loaded for 5 min on the precolumn (µ-precolum, Acclaim PepMap, 75 µm inner diameter, 2 cm, C18, Thermo Scientific) at 4% mobile phase B (80% acetonitrile in nanopure water with 0.08% formic acid) and 96% mobile phase A (nanopure water with 0.1% formic acid), then eluted from the analytical column (PepMap Acclaim C18 LC Column, 25 cm, 3 µm particle size, Thermo Scientific) over a 150 min non-linear gradient of mobile phase B (4-55% B).

Mass spectrometry was performed on a Q Exactive HF MS (Thermo Fisher Scientific, Waltham, MA, USA) with a TriVersa NanoMate (Advion, Ltd., Harlow, UK) source in LC chip coupling mode. The MS was set on loop count of 20 for MS/MS scan with higher energy collision dissociation (HCD) at normalized collision energy of 30%. MS scans were measured at a resolution of 120,000 in the scan range of 350-1,600 *m/z*. MS ion count target was set to  $3 \times 10^6$  at an injection time of 80 ms. Ions for MS/MS scans were isolated in the quadrupole with an isolation window of 1.6 Da and were measured with a resolution of 15,000 in the scan range of 200-2,000 *m/z*. The dynamic exclusion duration was set to 30 s with a 10 ppm tolerance. Automatic gain control target was set to  $2 \times 10^5$  with an injection time of 120 ms using the underfill ratio of 1% (Starke et al., 2017).

### (4) Spectra identification

The mass spectrometry data have been deposited into the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the proteomics identification (PRIDE) partner repository (Perez-Riverol et al., 2019). The generated MS data files were processed for peptide identification with the Thermo Proteome Discoverer software (v1.4; Thermo Fisher Scientific, Waltham, MA, USA). MS/MS spectra were searched using Sequest HT against the UniprotKB/TrEMBL database (release date of 6<sup>th</sup> April 2016) using 10 ppm peptide ion tolerance and 0.05 Da MS/MS tolerances. To improve the protein identification, MS/MS spectra were searched against bacterial sequences only (Tanca et al., 2016). Oxidation at methionine and

carbamylation at lysine and arginine residues were selected as variable modifications and carbamidomethylation of cysteine as the static modification. Peptide identification considered a false discovery rate (FDR) below 1% calculated by Percolator (Käll et al., 2007), a minimum length of six amino acids, and a peptide rank of one. A multi consensus report with the whole dataset was created with Thermo Proteome Discoverer software (v2.2; Thermo Fisher Scientific, Waltham, MA, USA). Protein matches were only accepted if they were identified by a minimum of one unique peptide and a high confidence. Formation of “Protein Groups” (hereafter referred to as proteins) was enabled with strict parsimony principle and using the highest scoring protein in the group as the confident representative protein (master candidate protein).

#### (5) Protein abundance, taxonomic and functional annotations

The bioinformatics pipeline “PROteomics result Pruning and Homology group Annotation Engine” (PROPHANE, [www.prophane.de](http://www.prophane.de)) (Schneider et al., 2011) was used for protein identification and taxonomic annotations using the highest sequence similarity to the UniProtKB/TrEMBL database and BlastP. The protein database was filtered for Bacteria proteins only. Fungal proteins were not included as no anaerobic fungi were detected by qPCR in RWS (data not shown) and no fungi are present in the gut of higher termites (Brune, 2014). Functional predictions of cluster of orthologous groups proteins (COGs) and Pfam (Protein families) domains were obtained with, respectively, RPS-BLAST algorithm and the COG collection (release 22.03.2003) (Tatusov et al., 2003), considering the first hit for each protein (e-value  $\leq 0.01$ ), and the Hidden Markov Model profiles with HMMER3 algorithm, considering the first hit for each protein (gathering cut-off) (Punta et al., 2012). Only identified proteins were kept for further analysis. For each sample, PROPHANE estimated protein abundance based on the normalized spectral abundance factor (NSAF) (Zybailov et al., 2006). NSAF is calculated as the number of spectral counts (SpC) identifying a protein, divided by the protein’s length (L), divided by the sum of SpC/L for all proteins in the experiment. NSAF allows the comparison of abundance of individual proteins in multiple independent samples and has been applied to quantify the protein expression changes in various microbial communities (Schneider et al., 2012; Hawley et al., 2014). Replicate-to-replicate variation was assessed by Pearson correlation analysis using the *cor* R function. Technical replicates with Pearson correlation value below than 0.7 were removed from metaproteomics data. Proteins present in only one of the three technical replicates were considered as false positives and suppressed, whereas remaining proteins were expressed as mean values. After

computational annotations of protein functions, assignments were carefully checked, completed, and manually curated.

#### (6) CAZyme annotation

The identified proteins were assigned to CAZyme families using a homemade script used for the day-to-day updates of the CAZy database as described previously (Svartström et al., 2017; Delannoy-Bruno et al., 2021). This homemade script enabled to appropriately handle the modular structure of CAZymes (which consists on a catalytic domain with a variable number of modules) for a precise protein classification by determining the secure parameters to consider a module as part of a family, or not. Briefly, amino acid sequences were analyzed on a 384-nodes computer cluster. The protein sequences were compared to the full length sequence of previously annotated proteins, stored internally in the CAZy database, using BlastP (version 2.3.0+). All remaining sequences with no hit were compared with BlastP to a library of individual modules (catalytic or auxilliary) and a HMMER3 searched against a collection of hidden Markov models based on each CAZy family (Delannoy-Bruno et al., 2021).

#### d) Data analysis and visualization

Hierarchical clustering was used to group the samples according to Euclidean distance and Ward as agglomeration criterion with *dist* and *hclust* R functions.

Plots were constructed using R's ggplot2 package (v3.3.2). Relative abundances of bacterial OTUs or proteins were aggregated at the phylum or genus level for stacked bar plot representation. Low abundance phyla and genera, showing relative abundances lower than 1% and 2%, respectively, were gathered as "Other".

To determine statistical significance of data comparison, non-parametric Wilcoxon rank sum test was applied (also called Mann–Whitney test) (Xia and Sun, 2017). The p-values resulting from pairwise comparisons were adjusted by the Benjamini–Hochberg correction. For multivariate analysis, to take the compositional properties of our data into account, we applied centered log-ratio transformation of variables (CLR transformation) (Calle, 2019). Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed using R packages mixOmics (v6.12.2) (Rohart et al., 2017) and factoextra (v1.0.7). PLS-DA was used to identify CAZymes that best discriminate the microbial consortia and degradation phases (Lê Cao et al.,

2011). Discriminant factors used in PLS-DA analysis were validated with permutation multivariate analysis of variance (PERMANOVA), a permutation-based version of the multivariate analysis of variance, which was performed using R's vegan package (v2.5-6) to test the statistical significant differences (Anderson, 2001).

## **4. Results**

### ***a) Data overview***

Cow rumen (RWS) and termite gut (TWS) derived lignocellulolytic consortia were used to generate 16S rRNA gene sequencing and metaproteomics data (Additional file 2: Supplementary Data 1: Metaproteomics data). 16S rRNA sequencing of RWS and TWS samples procured more than 800,329 high quality sequences and rarefying samples to 15,000 sequences captured most of the microbial diversity (Additional file 4: Figure S1). For metaproteomics data, Pearson correlation analysis (greater than 0.7) confirmed the similarity between technical replicates (Additional file 5: Supplementary Table S1). Only the replicate RWS2\_T1\_C was removed (correlation lower than 0.5 within technical replicates). A similar number of proteins (from 2,241 to 3,784 proteins) was found in each sample (Table III-2) irrespective of the consortium or time point considered, suggesting a similar coverage for all samples. RWS and TWS consortia possessed 7,149 and 6,682 proteins, respectively, having in common only 33.7% (3,489 proteins) of the total proteins (10,342 proteins), while about 35.4% (3,660) and 30.9% (3,193) of proteins were specific to the rumen-derived and the termite-derived consortia, respectively.

**Table III-2 : Diversity of species (OTUs) and proteins detected in RWS and TWS samples**

	16S rRNA		Metaproteomics			
	Richness (OTUs)	Shannon index	Richness (Protein groups)			Shannon index
RWS-1* T1	166	7.97	3016	3936	7149	3.65
RWS-1 T2	175	5.63	2524			4.00
RWS-1 T3	176	4.29	2389			3.68
RWS-1 T4	182	5.90	2463			3.63
RWS-2* T1	172	5.15	2329	4777		3.50
RWS-2 T2	177	5.19	2741			3.56
RWS-2 T3	179	6.51	2241			3.62
RWS-2 T4	176	5.99	2617			3.68
TWS-1 T1	101	2.79	3142	4802	6682	3.34
TWS-1 T2	103	2.07	3155			3.17
TWS-1 T3	134	2.04	2890			3.07
TWS-1 T4	134	2.06	2606			2.92
TWS-2 T1	120	2.92	3784	5724		3.58
TWS-2 T2	115	3.11	3159			3.44
TWS-2 T3	128	2.66	2878			3.31
TWS-2 T4	134	2.95	2742			3.21

Note: \*1 and 2 indicate the biological replicate bioreactors.

Moreover, the comparison of overall proteins of RWS and TWS by hierarchical clustering (Additional file 4: Figure S2) revealed consortia-specific profiles, evolving as a function of lignocellulose degradation. For RWS, although some variations were observed for each replicate, samples were globally ordered according to the incubation time. For both consortia, T1 points were separated from other time points. This probably reflects the more generic nature of the microbial community's ground states, unchallenged by lignocellulose and corresponds to low lignocellulose degradation and xylanase activity levels (Table III-1). This is supported by the fact that the other time point profiles were quite similar.



**b) Taxonomic and functional profiles of RWS and TWS throughout wheat straw degradation**

The diversity index and richness calculated from 16S rRNA gene sequencing data and phylogenetic annotation of metaproteomic data, revealed similar levels of diversity in both consortia (Table III-2). However, TWS was characterized by a less diverse community compared to RWS, with lower richness ( $122 \pm 13$  OTUs for TWS and  $176 \pm 5$  OTUs for RWS) and Shannon diversity index ( $2.6 \pm 0.4$  for TWS and  $5.8 \pm 1.0$  for RWS). Nevertheless, the Shannon diversity indices of the two consortia, estimated on metaproteomic-based phylogeny (Liu et al., 2017), were very close, with an average of  $3.26 \pm 0.21$  and  $3.67 \pm 0.15$  for TWS and RWS, respectively.

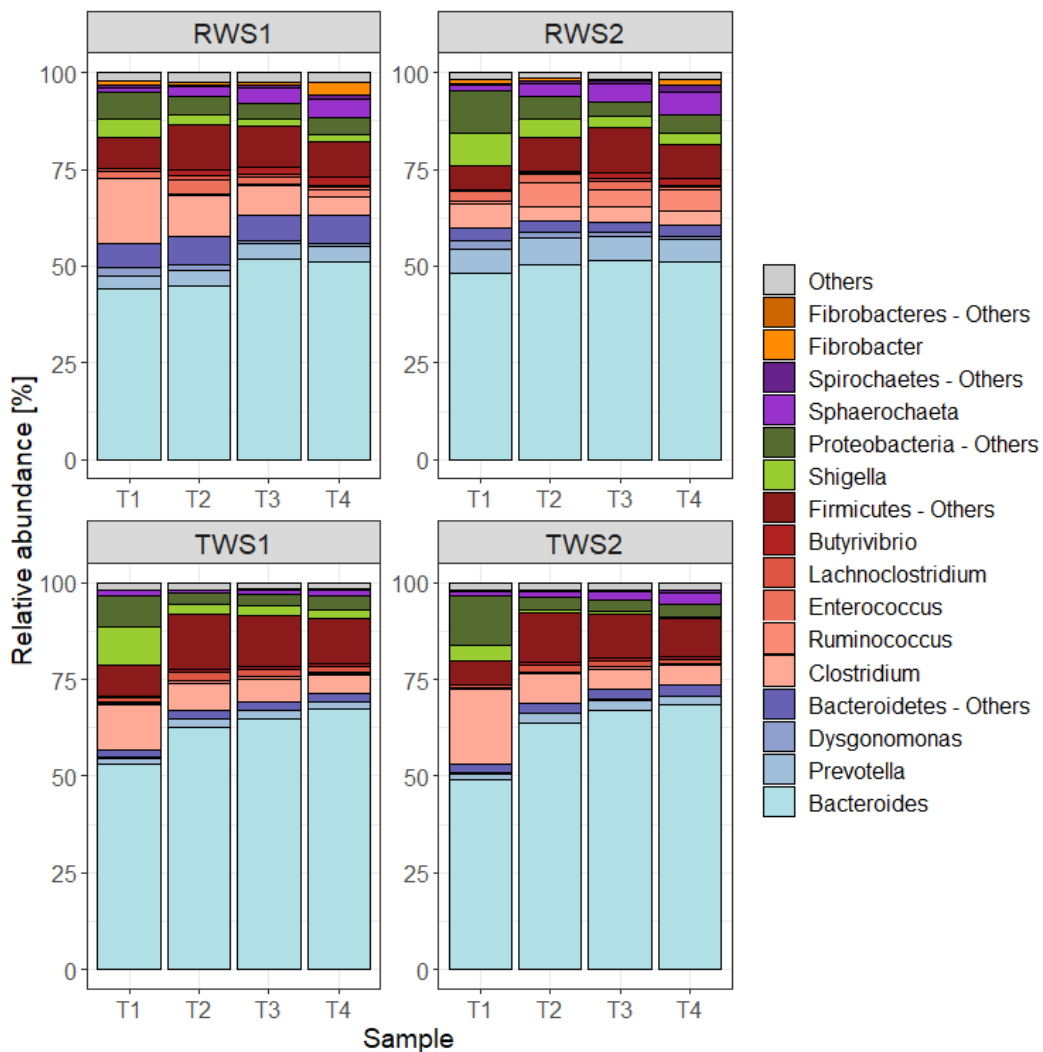
Taxonomic affiliation of proteins was in agreement with 16S rRNA gene sequencing data (Additional file 4: Figure S3) and revealed a dominance of species belonging to Bacteroidetes ( $60.4 \pm 2.6\%$  for RWS and  $66.5 \pm 7.6\%$  for TWS) in both consortia, followed by those related to Firmicutes (RWS:  $22.7 \pm 4.3\%$  and TWS:  $21.8 \pm 3.2\%$ ) and to a lesser extent to Proteobacteria phyla (RWS:  $9.4 \pm 4.4\%$  and TWS:  $8.0 \pm 5.9\%$ ). A minor fraction of species affiliated to Spirochaetes was present in both communities ( $4.3 \pm 2.0\%$  and  $1.9 \pm 0.9\%$  for RWS and TWS, respectively) while a minor fraction of Fibrobacteres members was exclusively observed in RWS ( $1.1 \pm 0.9\%$ ), consistent with previous data that associate this phylum with rumen environments (Hess et al., 2011).

Affiliation of proteins at the genus level provided a more detailed description of the active populations in both consortia throughout the duration of the experiments (Figure III-1A). Most of the proteins were affiliated to *Bacteroides* genus in all samples for both consortia ( $49.1 \pm 3.0\%$  for RWS and  $62.0 \pm 7.0\%$  for TWS) and in TWS, this genus remained abundant throughout the experiment. An opposite trend is observed for *Clostridium* (RWS:  $7.1 \pm 4.6\%$  and TWS:  $8.3 \pm 5.0\%$ ), because for both consortia the abundance of proteins affiliated to this genus decreased over time. It is noteworthy that in both consortia, proteins affiliated to *Bacteroidetes graminisolvans* were the most abundant ( $43.3 \pm 3.4\%$  for RWS and  $49.3 \pm 6.8\%$  for TWS) over the whole reaction period. This organism has previously been identified as a xylanolytic anaerobe (Nishiyama et al., 2009). Considering all proteins and their taxonomic affiliations, PCA analysis revealed a clear difference between the RWS and TWS communities as a function of a set of variables discriminated by the x-axis (Figure III-1B). This observation was confirmed by a Wilcoxon rank-sum test (Additional file 4: Figure S4). The experiments performed on the TWS community revealed a stronger abundance of proteins affiliated to *Bacteroides* and *Lachnoclostridium* and to minor genera gathered as

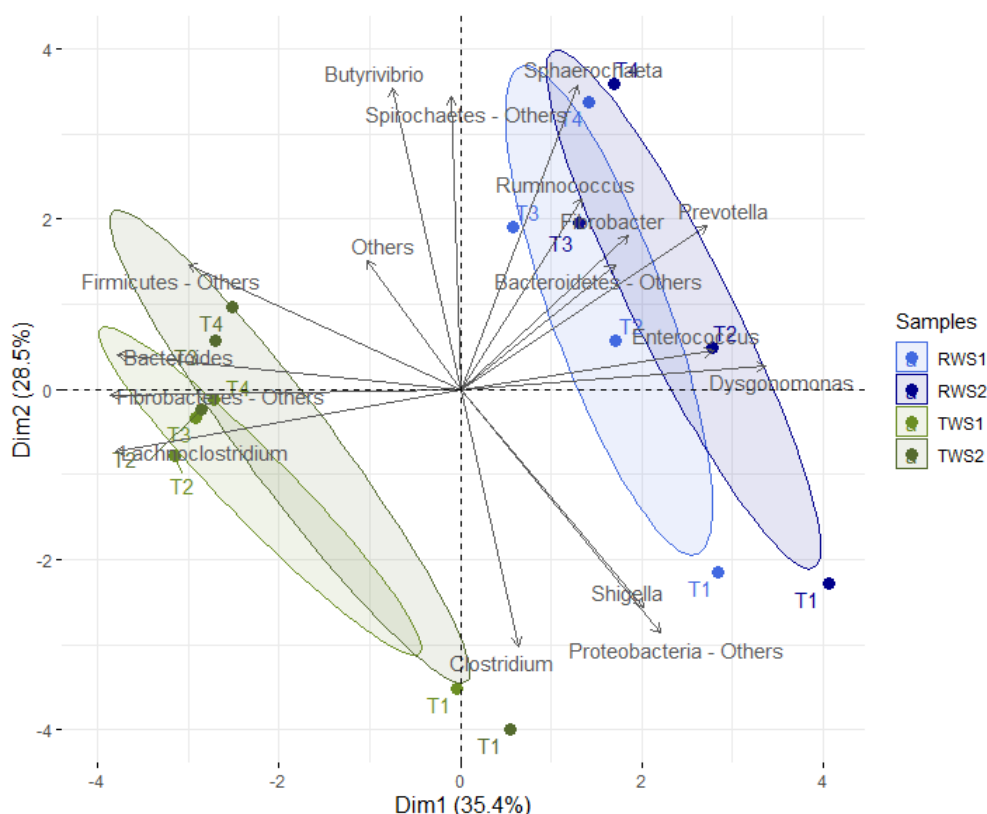
Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

Fibrobacteres-Others. RWS displayed a higher abundance of proteins belonging to *Prevotella*, *Dysgonomomas*, *Enterococcus*, *Sphaerochaeta*, *Fibrobacter* and minor genera gathered as Bacteroidetes-Others. In addition to inoculum separation, the y-axis of the PCA analysis isolated the T1 samples of both consortia on the basis of a higher abundance of proteins affiliated to *Clostridium*, *Shigella* and Proteobacteria-Others, while the samples related to the other time points (T2-T3-T4) were clustered on the basis of higher abundance of proteins belonging to *Ruminococcus*, *Butyrivibrio*, Firmicutes-Others and Spirochaetes-Others.

A.



**B.**



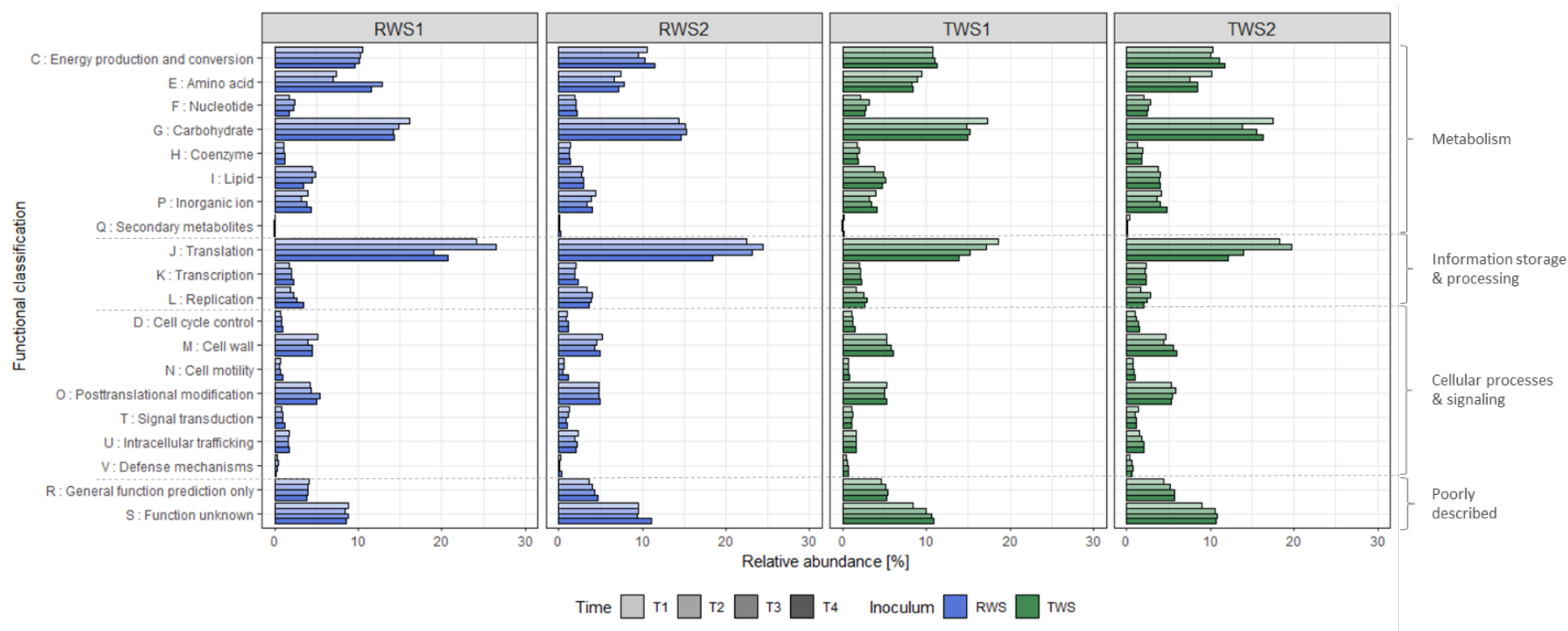
**Figure III-1 : A. Taxonomic community composition at genus level deduced from the identified proteins (metaproteomics data) for RWS (top graphs) and TWS (bottom graphs) (1 and 2 indicate the biological duplicates). Relative abundance of proteins based on NSAFs (normalized spectral abundance factors) was aggregated at the genus level for stacked bar plot representation. The group “Others” gather phyla with relative abundance less than 1% in the dataset. Within major phyla, the group “Others” gather genera with relative abundance less than 2% in the dataset or unclassified genera. Proteins belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red); Proteobacteria (green), Spirochaetes (purple), Fibrobacteres (orange). B. Principal components analysis (PCA) ordination plots with 95% confidence ellipses at the major genus level based on protein abundance (CLR-transformed data) at different sampling times of RWS (blue) and TWS (green) microbial consortia. Principal component 1 and 2 explained respectively 35.4% and 28.5% of the total variation.**

To better understand the functional dynamics of RWS and TWS communities, functional prediction of COGs based on RPS-BLAST was performed on the protein dataset. For both consortia, a relatively low fraction of proteins corresponded to unknown or poorly characterized functions (13.4±1.1% and 15.3±1.3% for RWS and TWS, respectively). Most of the proteins were assigned to

### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

---

main cellular functions, such as “metabolism” (RWS: 44.6±2.3% and TWS: 47.9±1.9%), “information storage and processing” (RWS: 27.5±2.6% and TWS: 20.6±2.6%) and “cellular processes and signaling” (RWS: 14.5±0.9% and TWS: 16.1±1%). Three subrole functions related to “translation, ribosomal structure and biogenesis” (RWS: 22.3±2.8% and TWS: 16.1±2.7%), “carbohydrate transport and metabolism” (RWS: 14.9±0.6% and TWS: 15.7±1.3%) and “energy production and conversion” (RWS: 10.3±0.6% and TWS: 10.9±0.6%) were highly abundant irrespective of the sampling time and consortium (Figure III-2).



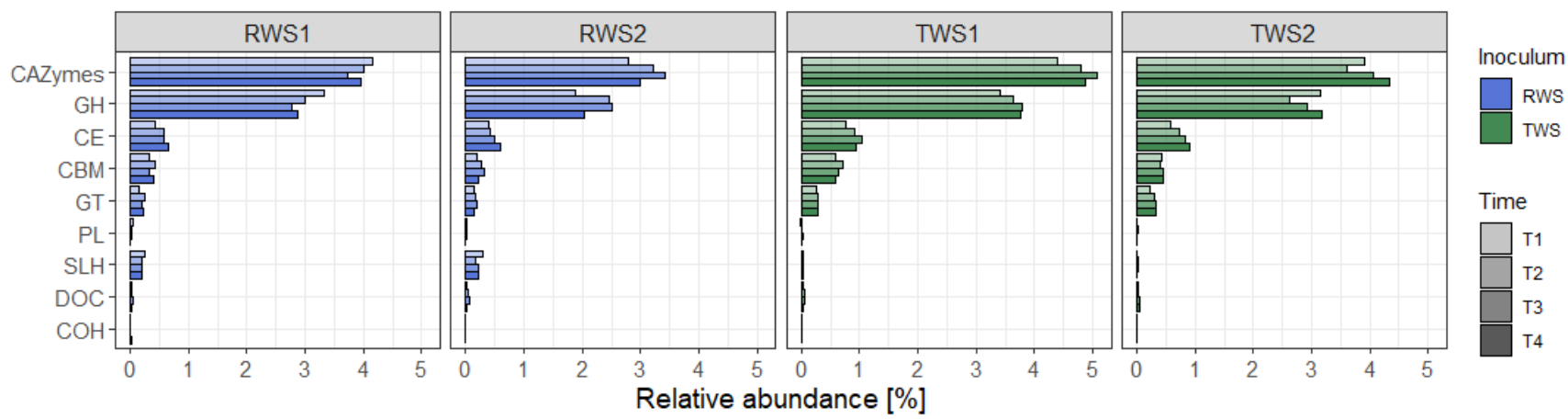
**Figure III-2 : Clusters of orthologous groups (COGs) in the metaproteomes of RWS and TWS replicates (1 and 2 indicate the biological duplicates).** The predicted proteins identified in the metaproteomes were mapped to different COGs using PROPHANE pipeline and RPS- BLAST program. Relative abundance of proteins based on NSAFs (normalized spectral abundance factors) was aggregated at the COG level for stacked bar plot representation for each time point. COGs affiliated to “Metabolism” functions: [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. COGs affiliated to “Information storage and processing” functions: [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair. COGs affiliated to “Cellular processes and signaling” functions: [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. COGs poorly described were: [R] General function prediction only, [S] Function unknown.

The functional COG profiles determined for the major phyla (Additional file 4: Figure S5A) highlighted Bacteroidetes as the most important contributor to proteins involved on “translation, ribosomal structure and biogenesis” (RWS:  $11.3 \pm 1.4\%$  and TWS:  $9.0 \pm 1.2\%$ ), “carbohydrate transport and metabolism” (RWS:  $7.9 \pm 0.4\%$  and TWS:  $8.9 \pm 1.2\%$ ) and “energy production and conversion” (RWS:  $6.0 \pm 0.5\%$  and TWS:  $6.9 \pm 1.1\%$ ). Firmicutes- and Proteobacteria- affiliated proteins also displayed these three main functions, but with lower abundance, suggesting that these phyla play a secondary role in RWS and TWS. As previously mentioned, the functional preponderance of Bacteroidetes- affiliated proteins reflects its high relative abundance, reflected in both 16S rRNA and metaproteomic data. Nevertheless, normalization of protein abundance at the phylum level for each sampling time revealed the specific COG profile of each phyla (Additional file 4: Figure S5B). It suggests that the metabolism of Firmicutes and Proteobacteria members was particularly focused on the “carbohydrate transport and metabolism” function (accounting for Firmicutes, RWS:  $18.2 \pm 3.0\%$  and TWS:  $19.5 \pm 3.1\%$  and for Proteobacteria, RWS:  $14.6 \pm 1.4\%$  and TWS:  $19.6 \pm 1.8\%$  of protein abundance) while Bacteroidetes species (RWS:  $13.0 \pm 0.7\%$  and TWS:  $13.4 \pm 0.6\%$ ) were involved in a wider range of functions.

### ***c) Carbohydrate-active enzymes in RWS and TWS consortia***

Protein annotation using the CAZy database (accessed March 30, 2018) revealed the carbohydrate-degrading potential of RWS and TWS consortia (Additional file 3: Supplementary Data 2: CAZy classification of proteins).

Among the 10,342 identified proteins, 423 proteins contain at least one CAZyme domain, they represented an average abundance of  $3.5 \pm 0.5\%$  in RWS and  $4.4 \pm 0.5\%$  in TWS metaproteome (Figure III-3 and Additional file 4: Figure S6A). Most of the proteins were classified as GH followed by those classified as CE, CBM, and GT, with least abundant proteins belonging to the PL family (Figure III-3).



**Figure III-3 : Relative abundance of CAZyme classes (glycoside hydrolases (GH), carbohydrate esterases (CE), carbohydrate binding modules (CBM), glycosyl transferases (GT), polysaccharide lyases (PL), S-layer homology (SLH), dockerins (DOC), cohesins (COH)) across sampling time of RWS and TWS consortia replicates.** Stacked bar chart show the distribution of identified protein groups assigned to each CAZyme domain. Relative abundance of proteins based on NSAFs (normalized spectral abundance factors) were aggregated at the CAZyme domain level for stacked barplot representation. The total abundance of CAZymes category is less than the sum of total abundance of each enzyme class (CBMs + CEs + COHs + DOCs + GHs + GTs + PLs + SLHs) because some multimodular proteins were detected.

### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

Among the CAZyme proteins identified, 174 proteins were common to both consortia, corresponding to  $83.3 \pm 4.8\%$  and  $88.7 \pm 5.2\%$  of the relative CAZymes abundance in the RWS and TWS metaproteomes, respectively. However, 127 and 122 proteins were unique to the RWS and TWS communities, respectively (Additional file 4: Figure S6B). RWS and TWS shared 69 CAZyme families while 15 and 10 were exclusively found in RWS or TWS, respectively. The richness of CAZyme family diversity is relatively high ( $n=94$ ) when compared to the whole of the CAZy database ( $n=353$ ). We found a large diversity of GH, CE and PL, representing 32%, 50% and 14% of the overall diversity found in the CAZy database (Table III-3). In addition, 11 proteins harboring SLH domains and 13 proteins containing COH or DOC domains were identified. Metaproteomics data revealed Bacteroidetes as the main CAZymes purveyor, producing about 80% of the total abundance of CAZymes (236 proteins) for both consortia (Additional file 4: Figure S6C). Firmicutes was the second contributor of CAZymes, with 138 proteins ( $16.2 \pm 4.4\%$  for RWS and  $13.3 \pm 2.7\%$  for TWS of the CAZyme families) while Proteobacteria expressed only 24 proteins, representing less than 1.5% of the CAZyme annotated proteins.

**Table III-3 : Diversity of CAZymes found in the metaproteomics dataset**

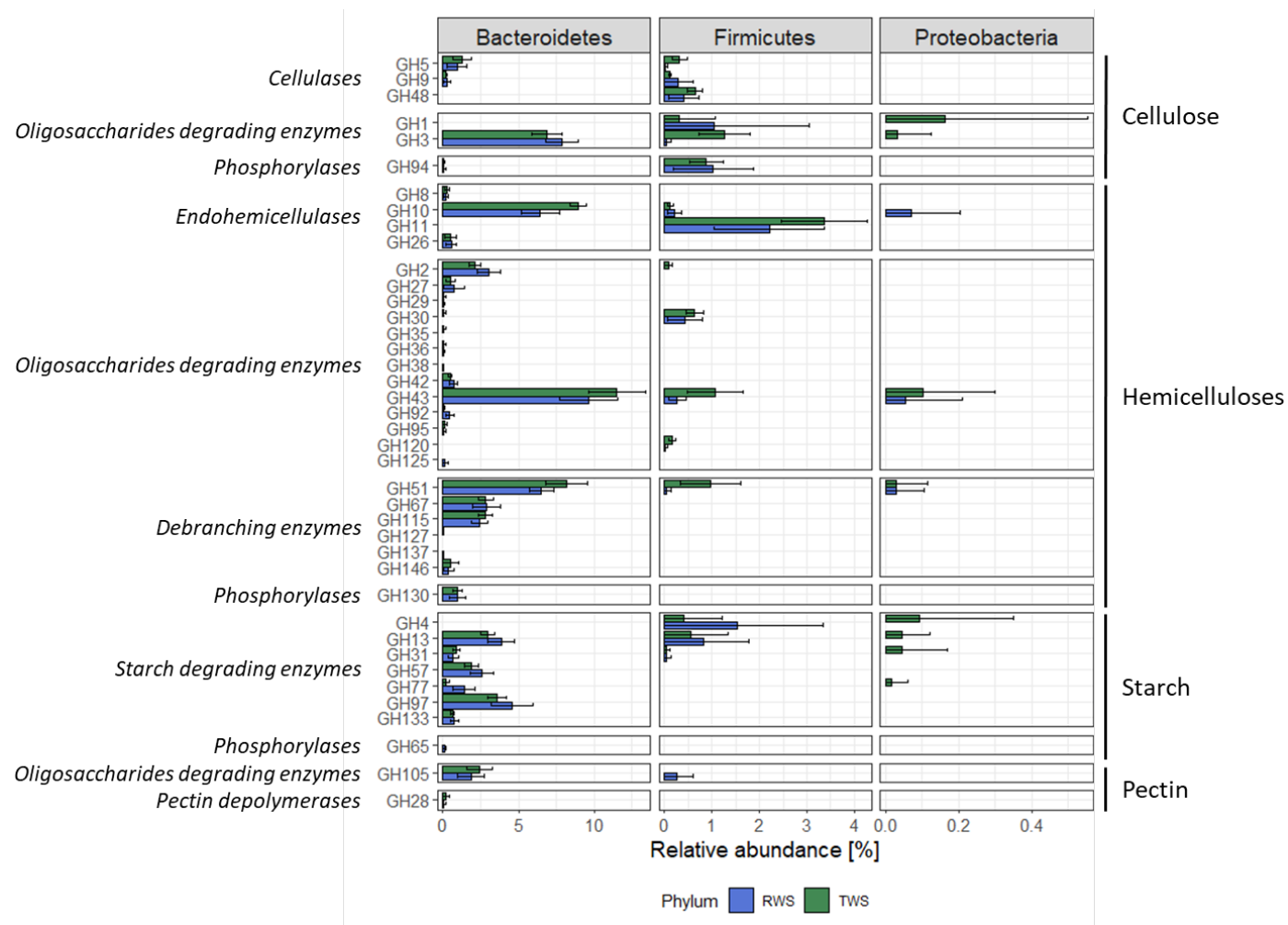
CAZy families	CAZy families in the dataset	CAZy families in the database	Number of proteins in the dataset
GH	49	153	330
GT	10	105	43
CE	8	16	31
PL	4	28	4
CBM	18	83	62



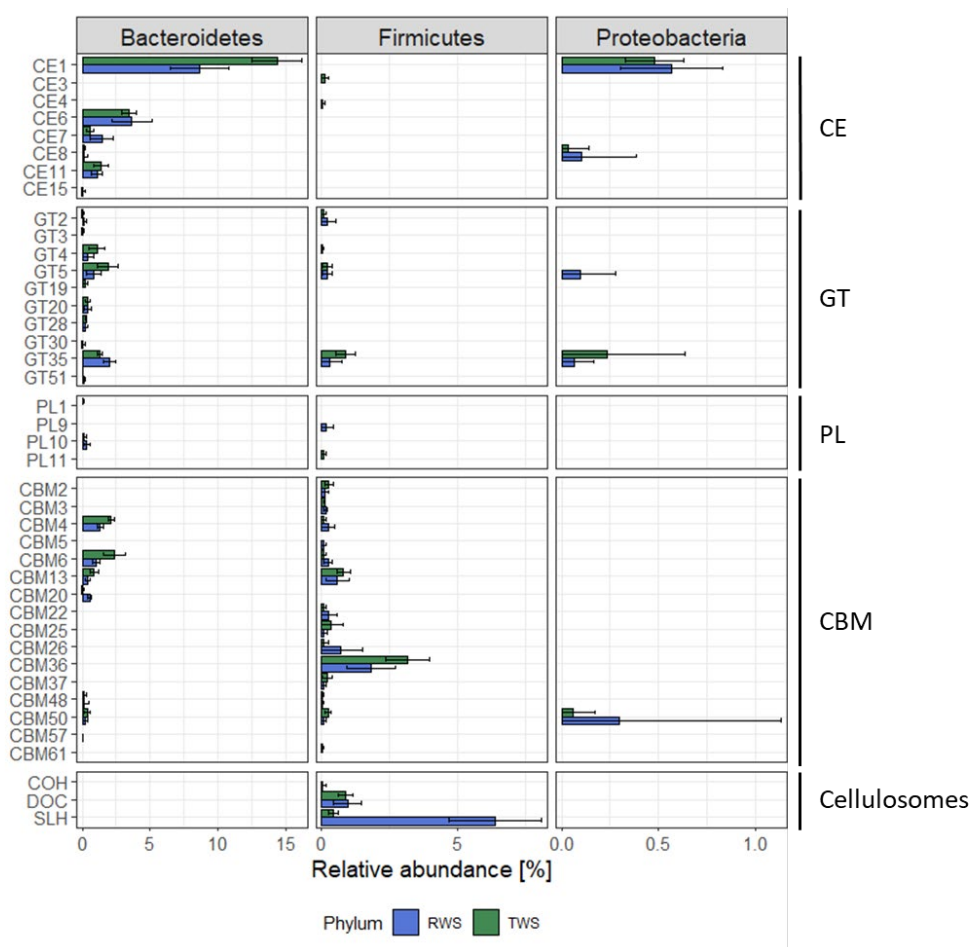
(1) Glycoside hydrolases (GH)

The most abundant and diverse CAZyme class was glycoside hydrolases (GH) (Figure III-3), representing more than 70% of the CAZymes detected in both RWS and TWS. The most highly expressed CAZymes were assigned hemicellulase and cellulase functions. A less abundant group of CAZymes was associated with starch and pectin hydrolysis (Figure III-4A).

A.



B.



**Figure III-4 : Comparison of abundance of CAZyme families classified by their function and taxonomy in the metaproteomes of RWS and TWS consortia.** Relative abundance of proteins (based on CAZymes-normalized NSAFs values) were aggregated according to the CAZyme family level. Error bars represent the standard deviation. **A.** Glycoside hydrolases (GH) expression for Bacteroidetes, Firmicutes and Proteobacteria phyla of RWS and TWS metaproteomes. Only GH targeting cellulose, hemicelluloses, starch and pectin fractions are showed. **B.** Carbohydrate esterases (CE), glycosyl transferases (GT), polysaccharide lyases (PL), carbohydrate-binding modules (CBM) and cellulosomes domains (S-layer homology (SLH), dockerins (DOC), cohesins (COH)) expressed by Bacteroidetes, Firmicutes and Proteobacteria phyla in RWS and TWS metaproteomes.

(a) *Hemicellulolytic activities*

Hemicellulolytic functions detected in both consortia (Figure III-4A, Additional file 3: Supplementary Data 2: CAZy classification of proteins) was provided by 76 proteins identified as hemicellulose-oligosaccharide-degrading enzymes displaying  $\alpha$ -galactosidase (GH27, GH36),  $\beta$ -galactosidase (GH2, GH35, GH42),  $\alpha$ -L-fucosidase (GH29, GH95),  $\alpha$ -mannosidase (GH38, GH92, GH125) and  $\beta$ -xylosidase/ $\alpha$ -L-arabinofuranosidase (GH43, GH120) activities. Most enzymes in this group were of Bacteroidetes (60 proteins) and Firmicutes (14 proteins) origin. In the case of both RWS and TWS, family GH43 (associated with either  $\beta$ -xylosidase or  $\alpha$ -L-arabinofuranosidase functions) was the most abundantly expressed. Moreover, 26% of GH43-proteins were modular enzymes, associated to a single, or multiple tandem repeats, of CBM6. This CBM family is known to bind amorphous cellulose and  $\beta$ -1,4-xylan and has been previously reported in association with GH43, although GH43-CBM6 represents a fairly small proportion of all GH43 in the CAZy database (Mewis et al., 2016). Likewise, GH43 were also associated with CBM13 or cellulosomal DOC domains, which enable the incorporation of the cognate GH43 domains into cellulosomes.

Hemicellulose-debranching enzymes produced by the two consortia (30 proteins) displayed  $\alpha$ -L-arabinofuranosidase (GH51),  $\beta$ -L-arabinofuranosidase (GH127, GH137, GH146) and xylan  $\alpha$ -1,2-glucuronidase (GH67, GH115) activities (Figure III-4A, Additional file 3: Supplementary Data 2: CAZy classification of proteins). GH51, GH115 and GH67 were the most abundant families of this group, mainly composed of single domain proteins. Most of the hemicellulose-debranching proteins were affiliated to Bacteroidetes except GH51 proteins that were expressed by all three main phyla.

Twenty-five endohemicellulases were detected in both consortia, displaying endo- $\beta$ -1,4-xylanase (GH8, GH10, GH11) and  $\beta$ -mannanase (GH26) activities (Figure III-4A, Additional file 3: Supplementary Data 2: CAZy classification of proteins). They were mostly expressed by Bacteroidetes (11 proteins) and Firmicutes members (10 proteins). In both consortia, the most abundant families were GH10 and GH11, the first exclusively of Bacteroidetes origin while all GH11 were of Firmicutes origin. GH11 was abundant in TWS; this is consistent with previous observation of this CAZy family in termite gut (Tokuda et al., 2018; Marynowska et al., 2020). GH10 and GH11 proteins were all multimodular proteins. Among them, some appeared to be putative components of cellulosomes originating from fibrolytic *Ruminococcus flavefaciens*, *Ruminococcus* sp. and *Eubacterium*, all active on hemicellulose.

(b) *Cellulolytic activities*

Regarding the cellulolytic and accessory enzymes for cellulose degradation, RWS and TWS expressed similar levels of proteins belonging to six GH families (GH1, 3, 5, 9, 48 and 94) (Figure III-4A, Additional file 3: Supplementary Data 2: CAZy classification of proteins).

Enzymes annotated as endoglucanases belonged to GH5 (11 proteins), GH9 (9 proteins) and GH48 families (5 proteins). It is known that GH5 and GH9 are active on a range of substrates, but GH9 family generally displays preference for soluble or insoluble (amorphous) cellulose (Ravachol et al., 2014). In this study, these GH families were affiliated to both Firmicutes and Bacteroidetes. In contrast, family GH48, known for their activity on crystalline cellulose (Chu et al., 2019), was only expressed by Firmicutes members affiliated to *Ruminococcaceae* and *Eubacteriaceae* families. Proteins assigned cellulolytic functions in this study are appended to CBM or cellulosome (COH and DOC) domains. For example, analysis of the data revealed that GH9-domains were commonly associated with CBM4, known to bind amorphous cellulose, while GH48 proteins, were commonly associated to DOC domains typical of cellulosomes.

The set of accessory enzymes targeting cellulose degradation was completed by proteins annotated as cellulose oligosaccharide-degrading enzymes mostly assigned to Bacteroidetes (26 proteins) and Firmicutes species (21 proteins). Interestingly, GH3 proteins annotated as  $\beta$ -glucosidases were mainly of Bacteroidetes origin, while proteins assigned to GH1 were exclusively affiliated to Firmicutes and Proteobacteria.

Additionally, 22 proteins belonging to family GH94, mostly expressed by Firmicutes, were annotated as cellobiose phosphorylases, suggesting phosphorylation for cellobiose uptake and metabolism during wheat straw conversion in both consortia.

(c) *Starch and pectin hydrolases*

Sixty-three starch-specific proteins were identified (GH families 4, 13, 31, 57, 77, 97 and 133). According to known functions for these families (and subfamilies thereof), activities include  $\alpha$ -galactosidase, 1,4- $\alpha$ -glucan branching enzyme,  $\alpha$ -glucosidase,  $\alpha$ -amylase, 4- $\alpha$ -glucanotransferase and amylo- $\alpha$ -1,6-glucosidase (Figure III-4A, Additional file 3: Supplementary Data 2: CAZy classification of proteins). These proteins mostly belonged to Bacteroidetes (36 proteins) and

Firmicutes species (19 proteins). Proteins belonging to GH13 (also known as the  $\alpha$ -amylase family) were linked to multiple starch-binding modules (CBM25, 26 and 48).

Pectin depolymerizing proteins were also detected in both consortia (GH105 and GH28), being mainly associated with Bacteroidetes species. In this regard, it is noteworthy that the previously mentioned GH43 family can also be included in pectin-degrading protein category, because some GH43 members hydrolyse arabinan, which is a pectin component.

## (2) Other carbohydrates-actives enzymes and binding modules

CE was the second most abundant CAZyme class with 31 proteins belonging to 8 families (CE1, 3, 4, 6, 7, 8, 11 and 15) (Figure III-3). These families are assigned functions related to ester bond cleavage, particularly removing acetyl moieties from acetylated xylan. This class of CAZymes was mainly of Bacteroidetes origin, although some CE1 members were also attributed to Proteobacteria (Figure III-4B). In this group, CE1 and CE6 (acetyl xylan esterases) were the most abundant, irrespective of the consortium, and were detected occasionally in tandem in multimodular proteins.

Proteins from the GT class (43 proteins) were assigned glycogen or starch phosphorylase, or glucosyltransferase functions and were mostly of Bacteroidetes origin (Figure III-4B) with low abundance in both metaproteomes (Figure III-3). The most abundant GT families were GT35 ( $\alpha$ -glucan phosphorylase), GT5 (glycogen synthase) and GT4 (glycosyltransferase).

The less abundant CAZyme proteins detected were polysaccharide lyases (PL) (Figure III-3 and Figure III-4B), consistent with the general rarity of PL in the CAZyme classification system. Four PL families were detected (PL1, 9, 10 and 11), being related to uronic acid cleavage. All of them belonged to either Bacteroidetes or Firmicutes species.

Concerning non-catalytic carbohydrates binding modules (CBM), 62 proteins containing CBM modules were detected in both metaproteomes, representing  $8.7 \pm 1.2\%$  and  $12.1 \pm 1.3\%$  of total CAZymes in RWS and TWS, respectively (Figure III-3 and Figure III-4B). According to current knowledge, when appended to catalytic domains, these proteins target enzymes to their substrates, increase the local concentration of bound enzyme and, in some cases, enhance the enzyme activity (Gilbert et al., 2013). The CBM detected here were mainly appended to catalytic

domains belonging to GH10, GH11 and GH43 (in principle hemicellulases), GH13 ( $\alpha$ -glucan-active enzymes) and GH9 (cellulose hydrolysis). The detected CBM potentially bind xylan (CBM families 2, 4, 6, 13, 22, 36 and 37), amorphous (CBM families 2, 4, 6, 11, 13, 30 and 37) and crystalline (CBM3 and 37) cellulose and some that potentially bind to both (CBM3). With respect to  $\alpha$ -glucan specific CBM, both consortia displayed potential starch binders (CBM20, 25, 26 and 48) and galactan binders (CBM61). Finally, CBM that present affinity for chitin and peptidoglycan (CBM families 5 and 50) were also observed. CBM originating from Firmicutes (33 proteins) displayed the largest diversity, followed by those from Bacteroidetes (15 proteins).

In anaerobic microbial consortia, the hydrolysis of insoluble carbohydrates is frequently mediated by cellulosomes that are extracellular multienzyme complexes. Cellulosomes attach to the cell surface through so-called S-layer homology (SLH) domains, while cohesin (COH) and dockerin (DOC) are complementary modules that form part of the cellulosomal building blocks. Our study revealed the presence of SLH, COH and DOC domains, confirming that cellulosome-bearing microorganisms form part of both consortia. Analysis revealed that all thirteen COH and DOC proteins are affiliated to fibrolytic *Ruminococcaceae* species and that these domains represent 1% of the overall CAZyme fraction (Figure III-4B). In cellulosomes, DOC link enzymes to COH which in turn are linked to the main scaffold of the complex. Accordingly, the DOC identified in this work are linked to GH families 5, 9, 10, 11, 16, 30, 43 and 48. Additionally, seven CBM (belonging to CBM families 4, 6 and 22) and one CE (CE4) are also appended to DOC. It is noteworthy that analysis revealed that one protein sequence (W7ULA4 attributed to *Ruminococcus flavefaciens*) displays the characteristic features of a scaffold protein bearing four tandemly arranged COH domains. Presumably, each of these forms a potential docking site for the attachment of catalytic CAZymes. Moreover, 11 SLH affiliated to Firmicutes species, were found in significantly higher proportion in the RWS ( $6.4 \pm 1.7\%$ ) CAZyme fraction compared to that of TWS ( $0.5 \pm 0.2\%$ ) (Figure III-4B).

#### **d) CAZyme activities in the inocula and during lignocellulose degradation**

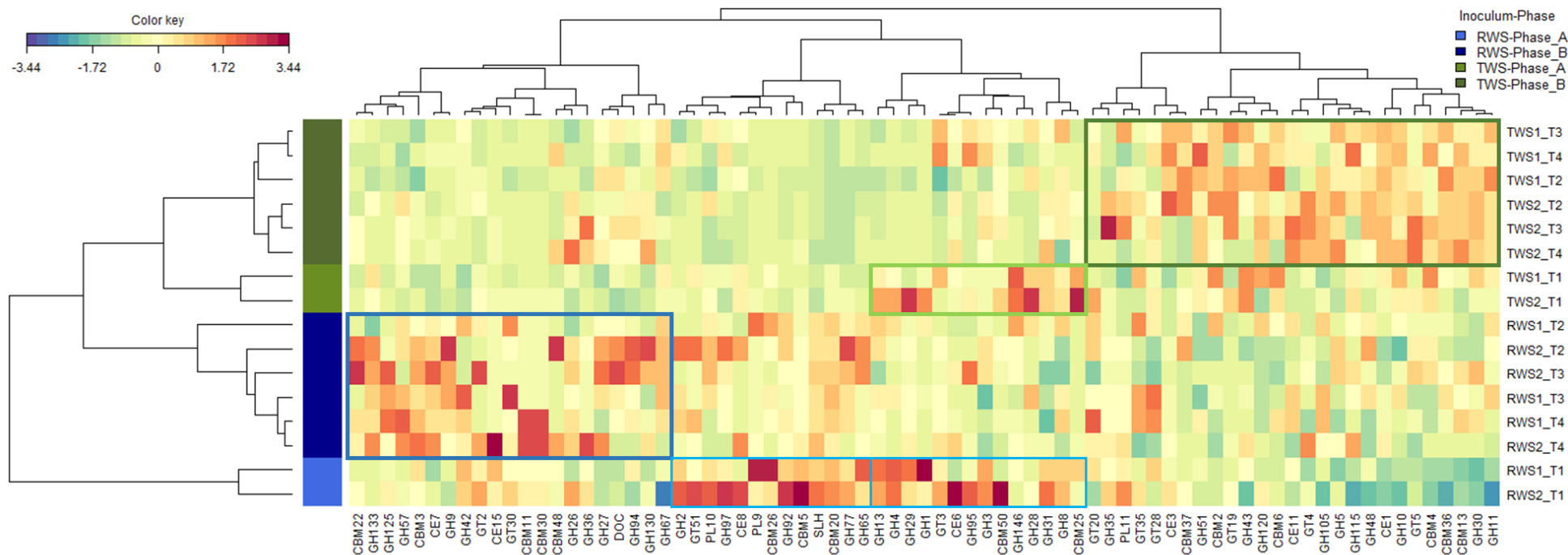
Multivariate statistical analyses were used to investigate how the inocula and the progression of lignocellulose degradation influence the expressed CAZyme profile. To achieve this, the analysis was conducted using the CAZyme CLR-transformed data along lignocellulose degradation. The PCA biplot clearly separated RWS and TWS samples (Additional file 4: Figure S7A), with RWS and TWS

co-localizing on the lefthand and righthand sides, respectively, of the first principal component (x-axis). Moreover, the positioning of the initial time points (T1) in the upper part of the y-axis of PCA biplot (second principal component), separate from the other time points T2, T3 and T4, indicates that the progression of substrate degradation influences the CAZyme expression profile (Additional file 4: Figure S7B). PERMANOVA analysis (Table III-4) supported the finding that inoculum source (R-squared value 0.453) and degradation phase (R-square value of 0.18) are the key determining factors of the CAZyme expression profiles, whereas the interaction term of these factors was statistically insignificant.

**Table III-4 : PERMANOVA analysis based on Euclidean distance, using inoculum source (“inoculum”: RWS or TWS), degradation phase (“phase” : phase A (T1 points) or phase B (T2-T3-T4 points)) and interaction between them (Inoculum:Phase”) as factors. The degrees of freedom (Df), sequential sum of squares (SumOfSqs), R-squared (R2) and p-values based in 9999 permutations Pr(>F).**

<i>Factors</i>	<i>Df</i>	<i>SumsOfSqs</i>	<i>R2</i>	<i>Pr(&gt;F)</i>	
<b>Inoculum</b>	1	0.035561	0.453	0.0001	***
<b>Phase</b>	1	0.014139	0.180	0.0021	**
<b>Inoculum:Phase</b>	1	0.004393	0.056	0.0920	
<b>Residual</b>	12	0.024368	0.311		
<b>Total</b>	15	0.078461	1.000		

Further confirmation of the importance of inoculum source and degradation phase was obtained using supervised partial least squares discriminant analysis (PLS-DA), using the discriminant factors identified by PCA and PERMANOVA analysis. In this analysis, Phase-A, corresponding to the incubation start point (T1), was distinct from Phase-B that groups all of the subsequent sampling points (Additional file 4: Figure S8). The clustered image map built from the PLS-DA analysis revealed that RWS and TWS samples clearly formed separated clusters (Figure III-5).



**Figure III-5 : Clustered Image Maps (CIM) represented the most discriminant CAZymes for RWS and TWS consortia at the different sampling times.** CIM was built using the extracted CAZymes from the first two PLS-DA dimensions considering the CLR-transformed CAZymes profile. Hierarchical clustering was derived using the Euclidean distance and Ward methodology. CAZymes are represented in columns and samples in rows. Axis colors on the left indicate the consortium and the degradation phase: green for TWS and blue for RWS; light colors correspond to phase A, and dark colors to phase B. The boxes highlight the clusters of CAZymes discriminating phase A of: TWS (light green) or RWS (light blue) and, phase B of: TWS (dark green) or RWS (dark blue).



### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

---

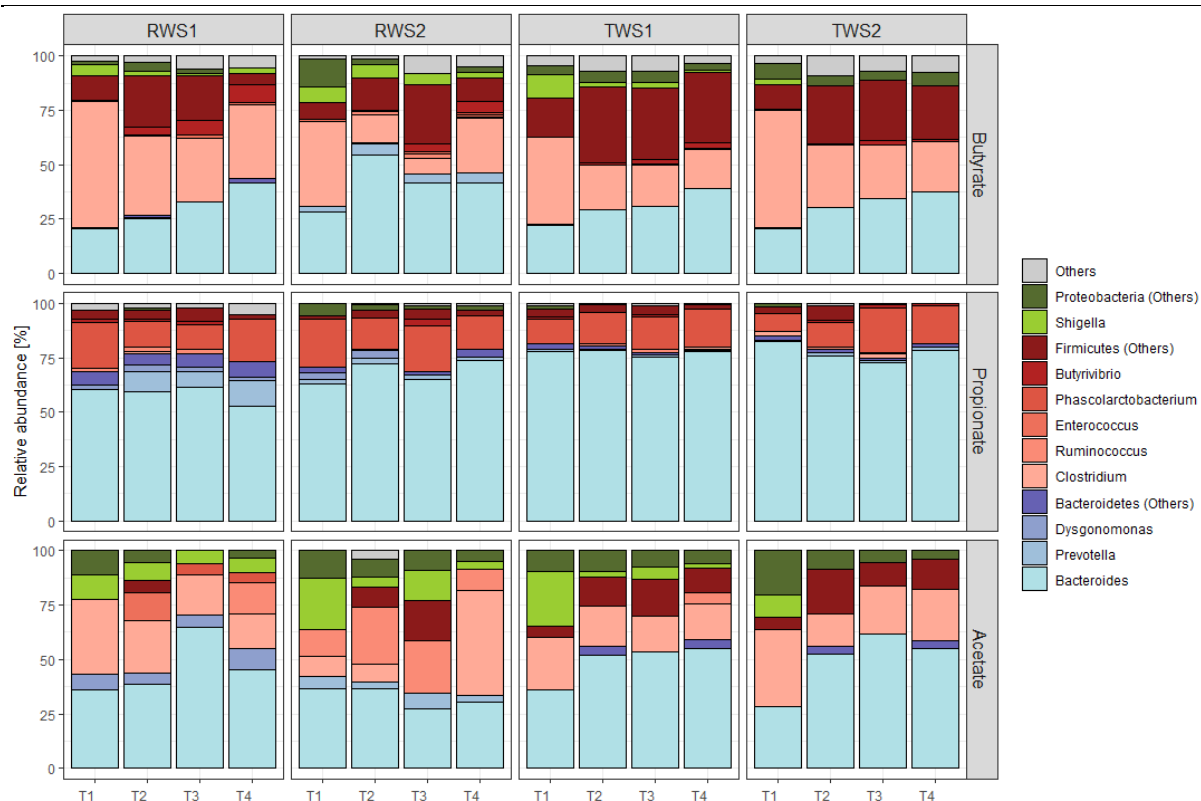
In each one of these major clusters, the specific-CAZymes distinctive of Phase-A were clearly identified (light-bleu and light-green boxes in Figure III-5). Interestingly, the Phase-A specific CAZymes for TWS were nevertheless common to those of RWS. This group is composed of CAZymes related to xylan and pectin depolymerization (GH8 and 28, respectively), phosphorylation and  $\alpha$ -glucan degradation (GH families 4, 13, 31), and oligosaccharide-degrading enzymes linked to cellulose (GH1, GH3) and hemicellulose breakdown (GH families 29, 95 and 146). This cluster is also characterized by the abundance of domains binding starch (CBM 25), and peptidoglycans (CBM50) binding modules. CE6 and GT3 were also present in this cluster. In addition to these CAZymes shared by both consortia, Phase-A of RWS also contains hemicellulose-oligosaccharide-degrading enzymes (GH2, 92, 95), enzymes linked to phosphorylation and  $\alpha$ -glucan degradation (GH families 65, 77 and 97) and domains binding starch (CBM20 and 26), cellulose and chitin (CBM5). PL9 and 10 and CE8 and GT51 were also present in this cluster. The presence of cellulosomal SLH is also distinctive of Phase-A of RWS samples.

The cluster of CAZymes determining Phase-B of TWS (dark-green box in Figure III-5) is composed of endoglucanases (GH5 and 48), hemicellulases (GH10 and 11), hemicellulose-debranching (GH51 and 115) and xylan deacetylating enzymes (CE1, 3 and 11), as well as GH families that degrade oligosaccharides specific to hemicellulose (GH30, GH35, 43 and 120) and pectin (GH105). It also contains a group of CBM that are either xylan-specific or bind to amorphous cellulose (CBM2, 4, 6, 13, 36 and 37). Finally, glucosyl/galactosyltransferases (GT families 4, 5, 19, 20, 28 and 35) and rhamnogalacturonan lyase (PL11) also belong to this phase.

The last cluster corresponds to CAZymes expressed in Phase-B of RWS (dark-blue box in Figure III-5). These include cellulose- (GH9 and 94) and starch- (GH57 and 133) active enzymes and those that are assigned polysaccharide deacetylation (CE families 7 and 15) and hemicellulose hydrolysis (GH26) functions. Additionally, the cluster contains exo-acting GH that probably degrade oligosaccharides (GH27, 36, 42, 67, 125 and 130). Like TWS, the Phase-B of RWS also contained glucosyl and galactosyltransferases (GT2 and 30) and CBM that are specific for  $\beta$ -glucans (CBM11), amorphous cellulose (CBM3 and 30), xylan (CBM22), starch (CBM48) and a dockerin domain typical of cellulosome structures.

**e) Enzymes related to volatile fatty acid production in RWS and TWS consortia**

Considering that lignocellulose degradation is usually associated with microbial VFA (or carboxylates) production and consumption by the host (Dijkstra, 1994; Brune, 2014) and that RWS and TWS consortia were selected for their ability to produce carboxylates from wheat straw, we investigated the occurrence of key enzymes related to acetate, butyrate and propionate biosynthesis in RWS and TWS metaproteomes. Based on previous COG assignments (Polansky et al., 2016; Deusch et al., 2017; Tilocca et al., 2017), the analysis revealed that VFA biosynthetic enzymes constituted  $2.8 \pm 0.4\%$  and  $3.5 \pm 0.2\%$  of total protein abundance in the metaproteomes of RWS and TWS, respectively. These values are of the same order of magnitude as those reported for CAZymes, indicating that VFA production is a major function in these lignocellulolytic consortia. Both consortia produced mainly acetate, followed by propionate and butyrate, but the proportion of the latter varied between RWS and TWS, displaying an average carbon molar ratio of acetate:propionate:butyrate of 59:23:18 and 47:20:33, respectively (Table III-1). Proteins related to butyrate biosynthesis (219 proteins) were more abundant in TWS ( $2.3 \pm 0.2\%$ ) than in RWS ( $1.7 \pm 0.4\%$ ) (Wilcoxon test – Additional file 4: Figure S9). Regarding this function, a large variety of enzymes were detected (numbers of each type indicated in parenthesis): acetyl/propionyl CoA carboxylase (3), butyrate kinase (7), 3-hydroxyacyl-CoA dehydrogenase (26), enoyl-CoA hydratase (17), acetyl-CoA acetyltransferase (52), alcohol dehydrogenases YqhD (8) and class IV (67), short-chain alcohol dehydrogenase (38) and Zn-dependent alcohol dehydrogenase (1). Most of them were affiliated to Firmicutes (138 proteins) and Bacteroidetes (23 proteins), with expression levels being similar for both phyla (Figure III-6). It is notable that at T1 *Clostridium* species were the major contributors of butyrate biosynthesis enzymes in both consortia (67 proteins, about 30%). However, subsequently the abundance of proteins affiliated to *Bacteroides* (16 proteins) increased, with higher levels of expression being reached in the latter phase of biomass bioconversion.



**Figure III-6 : Relative abundance and phylogenetic origin of bacterial enzymes involved in butyrate, propionate and acetate production in the metaproteomes of RWS and TWS (1 and 2 indicate the biological duplicates).** Proteins belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red) and Proteobacteria (green). For butyrate biosynthesis: COG4770 (acetyl/propionyl-CoA carboxylase), COG3426 (butyrate kinase), COG1250 (3-hydroxyacyl-CoA dehydrogenase), COG1024 (enoyl-CoA hydratase), COG0183 (acetyl/butyryl-CoA acetyltransferase), COG1979 (alcohol dehydrogenase YqhD), COG1454 (alcohol dehydrogenase, class IV), COG1028 (short-chain alcohol dehydrogenase), COG1064 (Zn-dependent alcohol dehydrogenase). For propionate biosynthesis: COG4799/0777 (acetyl/propionyl-CoA carboxylase), COG2185/COG1884 (methylmalonyl-CoA mutase), COG0346 (methylmalonyl-CoA epimerase). For acetate production: COG1012 (NAD-dependent aldehyde dehydrogenase), COG0282 (acetate kinase), COG0280 (phosphotransacetylase).

Propionate production was similar in both consortia (RWS:  $0.8 \pm 0.1\%$ , TWS:  $0.9 \pm 0.2\%$  - Additional file 4: Figure S9) and could be associated with the presence of 66 propionate biosynthesis-related proteins that represent a total of three functions: acetyl/propionyl-CoA carboxylase (34 proteins), methylmalonyl-CoA mutase (23 proteins) and epimerase (9 proteins). Their presence suggests that propionate was formed via the succinate pathway with methylmalonyl-coenzyme A mutases and epimerases. These proteins were mainly affiliated to Bacteroidetes, particularly to *Bacteroides* (37 from 66 proteins) (Figure III-6), while in Firmicutes, *Phascolarctobacterium* was the main contributor to this activity, providing about 20% of the relevant enzymes.

Surprisingly, for both consortia and irrespective of the sampling time, proteins involved in acetate biosynthesis were least abundant among the VFA biosynthesis enzymes (61 proteins accounting for  $0.3 \pm 0.1\%$  in both consortia – Additional file 4: Figure S9). The presence of acetate kinases (26 proteins), phosphate acetyltransferases (22 proteins), and aldehyde dehydrogenases (13 proteins), expressed by all phyla (Figure III-6), suggest that acetate production occurred via the Wood-Ljungdahl pathway. Seven of these proteins affiliated to *Bacteroides* formed the most abundant group, representing  $39.3 \pm 11.6\%$  and  $49.3 \pm 11.0\%$  of acetate biosynthesis enzymes in RWS and TWS metaproteomes, respectively, while those belonging to *Clostridium* represented about 20% (Figure III-6). Carboxylate biosynthesis-related proteins affiliated to Proteobacteria were also detected, but their abundance was low in comparison to other taxa. The contribution of Proteobacteria was slightly more important for acetate production, particularly at the initial incubation time.

## 5. Discussion

In Nature, lignocellulosic biomass degradation results from synergistic interactions between a variety of microorganisms occupying different ecological niches and operating under different environmental constraints. Understanding the interplay in lignocellulose degrading microbiota is thus challenging. A pragmatic strategy to tackle this is to reduce complexity using enrichment protocols that reduce microbial diversity and amplify the target functions. It is widely acknowledged that biomass degradation processes are largely determined by the genetic potential of the initial microbiome. Similarly, it is generally assumed that microbial diversity and enzyme expression continuously evolve in response to changing substrate structure and composition during degradation. Our study is thus based on these premises and aimed to

investigate what occurs when different enriched lignocellulolytic microbiota are confronted with identical bioprocess constraints.

This metaproteomic analysis provides a complete repertoire of enzymes involved in lignocellulose degradation and VFA production by two anaerobic lignocellulolytic microbial communities enriched from cow rumen and termite gut microbiomes. The number of non-redundant proteins observed here (10,342) was high compared to previous metaproteomic investigations of (D'haeseleer et al., 2013; Zhu et al., 2016; Alessi et al., 2018; Tomazetto et al., 2020), being more comparable to that observed in metaproteomics studies of much complex natural rumen ecosystems (Deusch et al., 2017). Therefore, this study allows more detailed and deeper taxonomical and functional investigation of the action of anaerobic microbiota on lignocellulose. It revealed also that despite RWS and TWS consortia were confronted to the same culture constraint, only a third of proteins were common to both consortia.

Taxonomic analysis of metaproteomic and 16S rRNA gene sequencing data showed that, despite the two initial microbial communities are drawn from quite different ecosystems, the active RWS and TWS communities are rather similar at the phylum level, displaying a predominance of Bacteroidetes and Firmicutes. Fungi proteins were not expected in these consortia because the gut of superior termites consists only of bacteria (Brune, 2014) and the anaerobic fungi present in the initial rumen inocula, were lost during the enrichment process and were absent from the RWS consortium. The similarity between RWS and TWS communities is probably the result of enrichment on wheat straw, meaning that certain highly specific features of the cow rumen and termite gut microbiota have been lost, particularly in the case of TWS. This convergence is largely driven by the identical conditions applied during the enrichment process (Auer et al., 2017). In particular, the substrate was almost certainly the major factor, consistent with previous observations of microbial communities growing on the same substrate (Wongwilaiwalin et al., 2013; Eichorst et al., 2014; Simmons et al., 2014). The predominance of Bacteroidetes and Firmicutes phyla, particularly *Bacteroides* and *Clostridium* genera in both consortia, suggests that these genera are the main wheat straw decomposers. This is consistent with the observation that herbivorous animals harbor many species belonging to the genera *Bacteroides* and *Clostridium* that are involved in the decomposition of complex carbohydrates including cellulose, xylan, and starch and represent hydrolysis key players (Nishiyama et al., 2009; Munir et al., 2014). Moreover, our 16S rRNA and metaproteomics data confirm that the most abundant species in both consortia

were also the major sources of plant cell wall degrading proteins in agreement with previous studies (Hart et al., 2018). This nevertheless contrasts with previous metatranscriptomics studies of the termite gut microbiome (Marynowska et al., 2017) and metaproteomics studies performed using an enriched microbial consortium to degrade rice and wheat straw (Alessi et al., 2017), sugar cane bagasse (Tomazetto et al., 2020) and a corn-stover degrading consortium (Zhu et al., 2016). Overall, these findings reveal that the relationship between relative abundance of microbial phyla in microbiota and their functional impact on the community activity is not a simple one. This fact highlights the usefulness of metaproteomics to elucidate the functional impacts of different phyla in specific environmental contexts.

The metabolic functions of the two consortia, assessed by COG analysis, revealed that the “translation, ribosomal structure and biogenesis”, “carbohydrate transport and metabolism” and “energy production and conversion” functional groups were predominant in both communities. This finding suggests that both RWS and TWS maintained similar enhanced capabilities for polysaccharide hydrolysis and sugar uptake throughout the wheat straw degradation process, consistent with results of previous metaproteomic studies (Zhu et al., 2016; Deusch et al., 2017; Alessi et al., 2018). However, previous studies also revealed that the ratio of the three functional groups is not always the same and depends on culture parameters. For example, independent studies performed on the same microbial consortium growing on wheat straw, revealed different abundances (5 and 16%, respectively) of proteins belonging to the “carbohydrate metabolism and transport” group (Alessi et al., 2017, 2018). Regarding the main COG functions expressed by the principal phyla in RWS and TWS, phyla-normalized data suggest that unlike Bacteroidetes (the major phylum), protein expression by Firmicutes and Proteobacteria was strongly directed towards carbohydrate and protein biosynthesis-related functions.

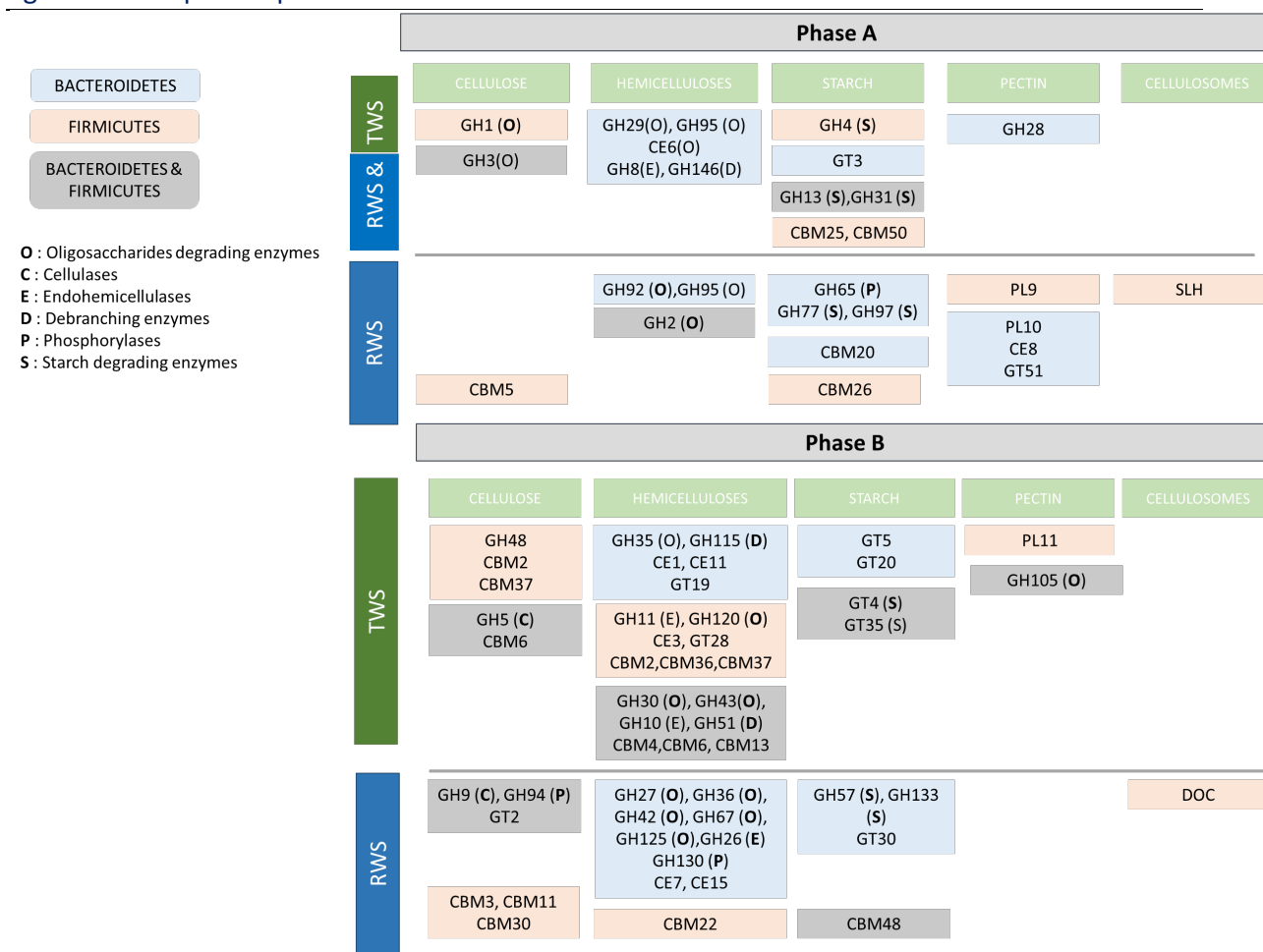
A large diversity of CAZymes related to lignocellulose degradation were detected in RWS and TWS metaproteomes. Among these, we identified a wide panel of biomass-degrading enzymes belonging to 94 families, including cellulases, hemicellulases as well as CBM. CAZymes identified in RWS and TWS accounted for around 4% of total proteins expressed by either consortium. Interpreting this result in the light of previous work is subject to precaution, because most previous studies reported the number of non-redundant proteins in the CAZyme fraction rather than their abundance (Zhu et al., 2016; Deusch et al., 2017; Hart et al., 2018). Nevertheless, it is possible to confirm that the value reported here is comparable to those measured (3.5% and 5.9%)

for other enriched cereal straw-degrading microbial communities (Alessi et al., 2017). This implies that microbial communities need to dedicate approximately 5% of their expressed proteins to lignocellulose degradation to achieve high (>45% in our study) biomass degradation levels.

Taxonomic affiliation of CAZymes revealed that lignocellulose degradation by RWS and TWS resulted from the complementary action of proteins affiliated to Bacteroidetes, which were the main purveyors of CAZymes, followed by Firmicutes. Furthermore, closer inspection reveals that the Bacteroidetes-affiliated species appear to be particularly focused on the production of hemicellulose-debranching and oligosaccharide-degrading enzymes as well as starch-degrading enzymes, while proportionally Firmicutes appear to produce more enzymes acting on  $\beta$ -glucan and  $\beta$ -xylan (cellulases and hemicellulases). Clearly this differential expression of CAZymes offers interesting prospects for synergy and distinguishes these two phyla as the key players in wheat straw degradation. Regarding CAZymes produced by Bacteroidetes species, it is known that their coding sequences often form part of so-called polysaccharide utilization loci (PULs) that are strictly regulated gene clusters that target the hydrolysis of specific carbohydrates (Grondin et al., 2017). In parallel to previous studies on lignocellulolytic communities, Bacteroidetes and Firmicutes were the dominant purveyors of CAZymes in RWS and TWS (Jose et al., 2017). Conversely, others studies have not always reported these phyla as the dominant purveyors for CAZymes, particularly those related to the termite gut microbiome (Calusinska et al., 2020), suggesting that this is strongly linked to the original microbial composition of the community under study and the process conditions employed (Jiménez et al., 2015; Zhu et al., 2016; Alessi et al., 2017; Tomazetto et al., 2020). Notably, the substrate, the enrichment process and the extent of oxygen availability will determine the relative development of obligate aerobes, facultative aerobes and strict anaerobes.

Although RWS and TWS consortia displayed similar community composition and taxonomic affiliation of CAZymes, multivariate analysis revealed that the CAZyme profiles were quite distinct, being linked to the original inoculum source. The difference in CAZyme profiles between RWS and TWS was particularly evident when examining each profile (Figure III-7).

### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV



**Figure III-7 : Main CAZymes and enzyme activities discriminating RWS and TWS consortia and the wheat straw degradation phases.**

The initial phase of lignocellulose degradation (Phase-A), although differences were observed between RWS and TWS consortia, both marshalled Bacteroidetes- and Firmicutes-affiliated enzymes that target the minor starch fraction and holocellulose oligosaccharides, facilitating the further breakdown of holocellulose polymers. Interestingly, CAZy families active on cellulose or hemicellulose derived-oligosaccharides were mostly affiliated to Firmicutes and Bacteroidetes, respectively. This observation highlights the complementarity between these taxa. The second phase (Phase-B) of biomass deconstruction mediated by both consortia was characterized by an increase in CAZy families related to cellulose hydrolysis, hemicellulose deconstruction and pectin depolymerization. It is noteworthy that comparison of the enzyme profiles of RWS and TWS in Phase-B reveals that these are quite distinct. During this phase, RWS expressed starch-degrading enzymes. Importantly, although both consortia produced hemicellulose oligosaccharide-



degrading enzymes, the diversity of these enzymes was much higher in RWS than in TWS. On the other hand, TWS produces a lower diversity of CAZymes in this phase but produces more hemicellulose-specific enzymes. In Phase-B, the members of the distinctive set of hemicellulolytic enzymes produced by TWS were present in higher abundance than those that characterize RWS (Figure III-5). This observation is consistent with higher hemicellulose degradation and higher xylanase activity achieved by TWS compared to RWS (Table III-1; (Lazuka et al., 2015, 2018)). Therefore, the two profiles clearly reflect different metafunctions whose origins possibly correlate with the differential adaptation of the two native microbiomes. The higher endoglucanase and xylanase activities of TWS were probably the result of the termite microbiome's ability to degrade hardwood core fibers (Warnecke et al., 2007; Marynowska et al., 2020). The greater CAZyme diversity in RWS and the high frequency of enzymes active on soluble carbohydrates (e.g those belonging to GH13, GH77, GH97 families) is consistent with the specialization of the original rumen microbiome towards less recalcitrant forage-based substrates (Brulc et al., 2009; Li et al., 2020a). Overall, this implies a clear influence of the original inoculum source, consistent with the conclusions of a previous study (Cortes-Tolalpa et al., 2016). This means that although identical substrate and culture conditions drive the microbial communities toward convergency, some specific inoculum-related features remain in RWS and TWS. Among the markers of this heritage, it is possible to underline the presence of clear evidence for ruminococcal and clostridial cellulosome components in RWS, which are specific to the cow microbiome (Lamed et al., 1983; Dai et al., 2015) and, accordingly with previous termite gut studies (Tokuda et al., 2018; Marynowska et al., 2020), xylanase proteins belonging to GH11 family were present in high proportion in TWS. A remarkable finding in this work is the fact that no lignin-specific enzymes (CAZyme AA class) were found, suggesting that ligninolysis did not occur under the anaerobic culture conditions used. This concurs with previous lignin measurements performed on RWS and TWS (Lazuka et al., 2015, 2018). Nevertheless, the fact that no lignin-degrading enzymes were evidenced does not imply that these are absent, because shotgun metaproteomics procures an incomplete image of expressed proteins. Ultimately, to assert that no lignin degradation had occurred in our experiments, it would be necessary to perform a thorough physicochemical and structural analysis of the substrate before and after microbial treatment.

The CAZy classification database has been growing at a fast rate in recent years, with new sequences being added daily and new families being regularly defined. Comparing the collection of GHs detected in our study with those detected in previous omics studies (metatranscriptomics,

metaproteomics, metagenomics) performed on bovine rumen and termite-gut microbiomes (Tableau III-5) revealed that the present study unmasked greater GH diversity. Moreover, the comparison revealed that our study captured most bacterial plant cell wall degrading GH families, whereas the previous studies were less successful in this regard. To a large extent, the greater coverage of GH families in our study can be correlated with the growth of the CAZy database (Lombard et al., 2014) and ongoing improvements to the experimental techniques and bioinformatics pipelines used. However, we believe it is also attributable to the fact that RWS and TWS are enriched consortia whose functions are highly adapted for the degradation of raw lignocellulosic biomass. Undoubtedly, future studies benefitting from further progress in metaproteomics (better protein separation and identification) and the further expansion of the CAZy database will surpass our study. Hopefully, these will provide an even deeper understanding of lignocellulolytic functional in microbial ecosystems and provide the means to identify proteins that are currently unclassified.



### Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

Study*	Ecosystem	Cellulase								Endohemicellulase						Debranching enzyme							Oligosaccharides degrading enzyme												Starch degrading enzyme			Reference												
		GH5	GH6	GH7	GH9	GH16	GH17	GH44	GH45	GH48	GH74	GH8	GH10	GH11	GH12	GH26	GH28	GH30	GH53	GH51	GH54	GH62	GH67	GH78	GH105	GH106	GH115	GH127	GH1	GH2	GH3	GH4	GH27	GH29	GH32	GH35	GH38		GH39	GH42	GH43	GH52	GH92	GH94	GH120	GH13	GH31	GH77		
TMG1	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X	X		X	X	X	X	X	X	X	X	X	X	X	X									X	X	X	X	X													X	X		Warnecke et al., 2007		
TMG2	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X	X		X		X	X	X	X	X	X	X	X	X	X			X	X						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		He et al., 2013	
TMG3	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X									X																		X	X															X			Grieco et al., 2019		
TMG4	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X	X		X	X	X	X	X	X	X	X	X	X	X	X			X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Romero-Victorica et al., 2020	
TMT1	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X	X		X	X	X	X	X	X	X	X	X	X	X	X			X	X						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Marynowska et al., 2017; 2020
TMT2	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X			X		X	X	X					X	X											X	X															X			He et al., 2013		
TMT3	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X								X	X	X				X													X																X			Tokuda et al., 2018		
RMP1	Holstein dairy cow rumen-derived consortium - unpretreated wheat straw	X			X	X		X		X	X	X	X	X	X	X			X		X							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	This study
TMP1	Termite hindgut ( <i>Nasutitermes</i> sp.)-derived consortium - unpretreated wheat straw	X			X	X		X		X	X	X	X	X	X	X			X		X							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
RMP2	Jersey cow rumen microbiome fed with corn silage, grass silage or grass hay	X						X											X									X	X	X														X		X	X	X	Deusch et al., 2017	
TMP2	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X			X					X	X																		X	X	X														X			Burnum et al., 2011		
TMP4	Termite hindgut microbiome ( <i>Nasutitermes</i> sp.)	X								X																			X	X	X																		Warnecke et al., 2007	

\* RMG: Rumen metagenome; RMT: Rumen metatranscriptome; TMG: Termite gut metagenome; TMT: Termite gut metatranscriptome; RMP: Rumen metaproteome; TMP: Termite metaproteome

Concerning VFA biosynthesis, a correlation between the abundance of oligosaccharides-degrading enzymes and higher VFA production levels has been established (Singh et al., 2014). In the present study, this category of enzymes was particularly abundant in Phase-A of both consortia. This implies that readily solubilized carbohydrates were rapidly hydrolyzed to simple fermentable sugars, consistent with the rather high levels of VFA detected, starting as early as the T2 sampling period. Unfortunately, because VFA production profiles levels were identical across this study, no new knowledge about the specific enzymes that underpin VFA production could be gleaned. Nevertheless, the data did show that acetate biosynthesis was the result of Wood-Ljungdahl pathway in both consortia. Moreover, our work highlighted the major role of acetogenic bacteria belonging to Bacteroidetes and Firmicutes phyla. Propionate production in RWS and TWS mostly resulted from the succinate pathway as evidenced by the detection of methylmalonyl CoA mutases and epimerases. This is consistent with previous data that shows that the succinate pathway is the main route reported for rumen (Allgaier et al., 2010). In both consortia, *Bacteroides* (Bacteroidetes) and *Phascolarctobacterium* species (Firmicutes) were the key propionate producers using the succinate pathway (Watanabe et al., 2012). Butyrate biosynthesis results from either from the conversion of butyryl-CoA into butyrate, using butyrate kinase (synthesizes butyryl-phosphate) and phosphotransbutyrylase, or the transfer of coenzyme A (catalyzed by butyryl-CoA:acetate-CoA transferase) between acetate and butyrate. Although both routes are exploited by Firmicutes species, the second one was strongly enhanced in both consortia. This observation is consistent with previous studies on the cow rumen and human gut microbiota that reveal that Bacteroidetes are responsible for the majority of acetate and propionate production, while butyrate biosynthesis is mainly handled by Firmicutes species (Abdallah Ismail et al., 2011; Deusch et al., 2017).

## 6. Conclusions

The work presented here combining 16S rRNA gene sequencing and metaproteomics allowed us to demonstrate that members of different taxa displayed complementary activities for anaerobic wheat straw bioconversion into VFAs. The study also highlighted several interesting features of TWS and RWS ecosystems. In particular, numerous lignocellulolytic enzymes were common to both consortia but they displayed different protein profiles; these specific features most likely resulted from the inoculum origin. Furthermore, our data showed correlations between the dynamics of putative (hemi)cellulolytic proteins and xylanase activity measurements. RWS and TWS consortia thus represent excellent simplified models to study the mechanisms governing the

Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

---

complex lignocellulose degradation process and to better understand and exploit multispecies lignocellulolytic enzyme systems for biotechnological applications.

## 7. Additional files

### a) Additional file 1: Supplementary Materials and Methods

#### (1) Lignocellulose substrate, rumen and termite gut inocula

Wheat straw from the winter wheat variety Koreli was collected at an experimental farm (INRAE, Boissy-le-Repos, France). After harvesting, the straw was milled to 2 mm and stored at room temperature (20–25°C). As described in our previous works *Nasutitermes ephratae* (Auer et al., 2017) and cow rumen (Lazuka et al., 2015) were used as inocula for a sequential batch bioreactor enrichments. The initial termite gut inoculum (500 dissected guts) were provided by IRD (Institute for Research and Development, Bondy, France). The initial rumen inoculum (2% w/v) was provided by the INRAE (French National Institute for Agriculture, Food and Environment) Center of Theix (St-Gènes-Champanelle, France).

#### (2) Anaerobic enrichment and lignocellulose degradation bioreactors of RWS and TWS

The lignocellulolytic communities from termite gut and cow rumen microbiomes were first enriched by successive cultivation using wheat straw as sole carbon source and mineral media (MM) (Lazuka et al., 2015), supplemented with 250 µL of V7 vitamin solution (Pfennig and Trüper, 1992) and 1 mL of sterilized (0.2 µm filtration) trace elements SL12 solution (Overmann et al., 1992). Each successive bioreactor was inoculated with 10% v/v of the previous bioreactor. All bioreactor experiments were carried out in 2 L anaerobic batch bioreactors (BIOSTAT® A+, Sartorius, Germany) except the first termite gut bioreactor that was realized in 400mL bioreactor (Applikon MiniBio 500). Bioreactors were operated under strict anaerobic conditions with agitation (400 rpm) and temperature control at 35°C; pH was maintained at 6.15 by adding a 2 M NaOH solution. The absence of dissolved oxygen was ensured by nitrogen flushing after inoculation and continuous monitoring with a polarographic dissolved oxygen probe (AppliSens). After several transfers, this enrichment procedure enabled to enrich the lignocellulolytic communities derived from cow rumen (RWS) and termite gut (TWS) (Lazuka et al., 2015, 2018).

Lignocellulose degradation capacity of RWS and TWS was assessed in two replicate anaerobic bioreactors under the same conditions described above. During the incubation time, methane

production was monitored and, if necessary, inhibited by the addition of 2-bromoethanesulfonate (BES), a methanogenesis inhibitor, until a maximum concentration of 10 mM. Along the incubation, samples were regularly taken to analyze the substrate composition, volatile fatty acid (VFA) and gas production, and cellulolytic and hemicellulolytic enzyme activities. Samples were also taken for biodiversity and metaproteomics analysis.

### (3) Dry matter degradation

Wheat straw concentration was determined along the incubation by measuring the dry matter. The total solids was determined using 10 mL samples that were first centrifuged ( $7,197 \times g$ , 10 min), rinsed twice with distilled water and dried for 24 h at 105°C. The mineral fraction (MF) was estimated by mineralization of the samples at 500°C for 2 h. The dry matter was determined by subtracting MF from total solids. Wheat straw degradation was reported as percentage of dry matter (% w/w) related to the initial dry matter.

### (4) Volatile fatty acid production

VFA production was determined by gas chromatography (GC), using a Varian 3900 chromatograph equipped with a flame ionization detector and CP-Wax 58 (FFAP) CB column (length: 25m, inside diameter: 0.53 mm) (Cavaillé et al., 2013).

### (5) Xylanase activity

For enzyme activity measurement, bioreactor samples (5 mL) were removed at regular intervals and kept at -20°C until further analysis. Xylanase activity was determined by end-point measurements performed after appropriate dilution on reactions containing a large excess of substrate. For each bioreactor and each sampling time, enzyme activity was measured in technical duplicates and expressed as average values; a blank without substrate was also included (Lazuka et al., 2015).

Xylanase activity was measured using 1% w/v xylan beechwood (Sigma) dissolved in 50 mM acetate buffer solution (pH 6). Activity was estimated by measuring the release of reducing sugar equivalents using the dinitrosalicylic (DNS) method and appropriate standard xylose solutions. Briefly, 200  $\mu$ L sample (or an appropriate dilution thereof) were added to 400  $\mu$ L of substrate solution and incubated at 35°C for 1 h (xylanase). Reactions were stopped by adding 600  $\mu$ L of DNS solution and heating for 5 minutes at 100°C. The amount of reducing sugars released was



Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

---

determined by measuring absorbance of the sample at 570 nm using a UV/VIS spectrophotometer (Multiskan Ascent, Thermo Scientific) and then comparing the value to a standard curve of xylose. One unit of xylanase activity (UA, unit of activity) was defined as the amount of enzyme that produces 1  $\mu\text{mol}$  of reducing sugars per minute.

***b) Additional file 2: Supplementary Data 1: Metaproteomics data***

Available data upon request.

***c) Additional file 3: Supplementary Data 2: CAZy classification of proteins***

Available data upon request.

d) *Additional file 4: Supplementary Figures S1 to S9*

(1) Figure S1

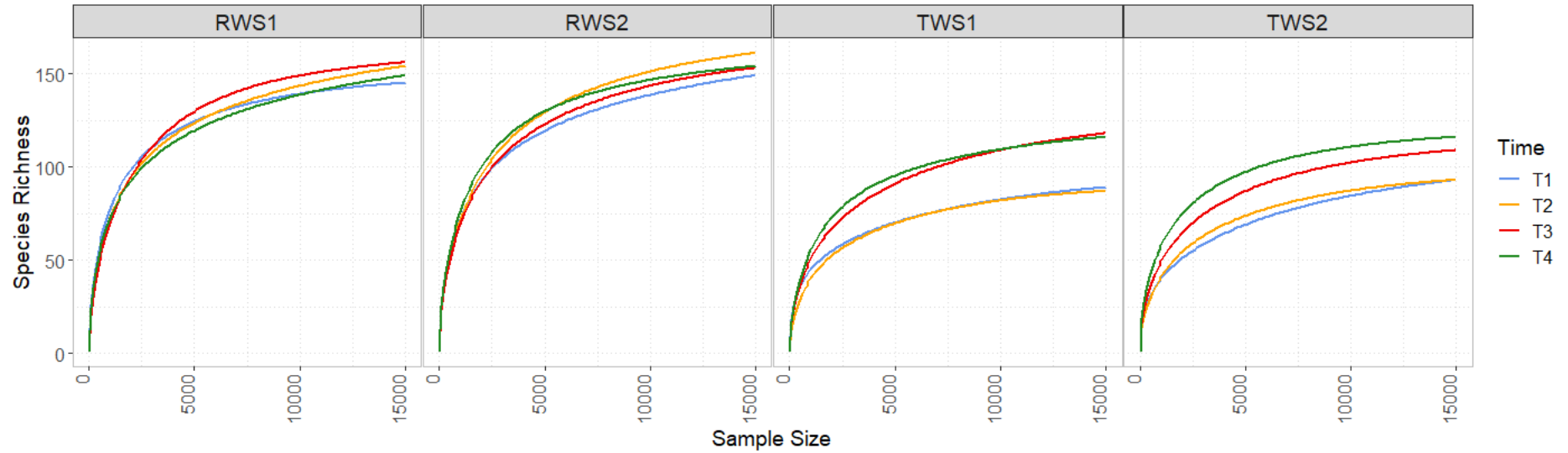


Figure III-S1 : Rarefaction curves (15,000 sequences) of each sampling time in RWS and TWS consortia replicates.

(2) Figure S2

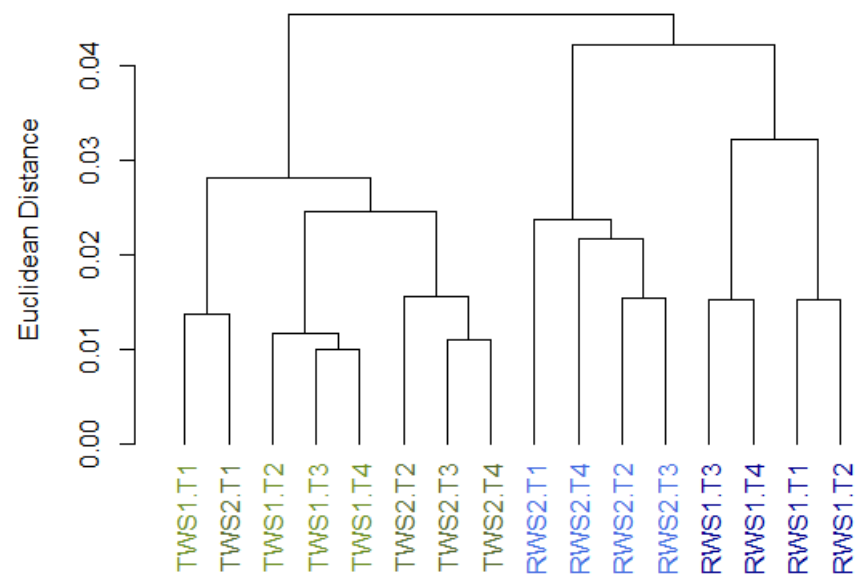
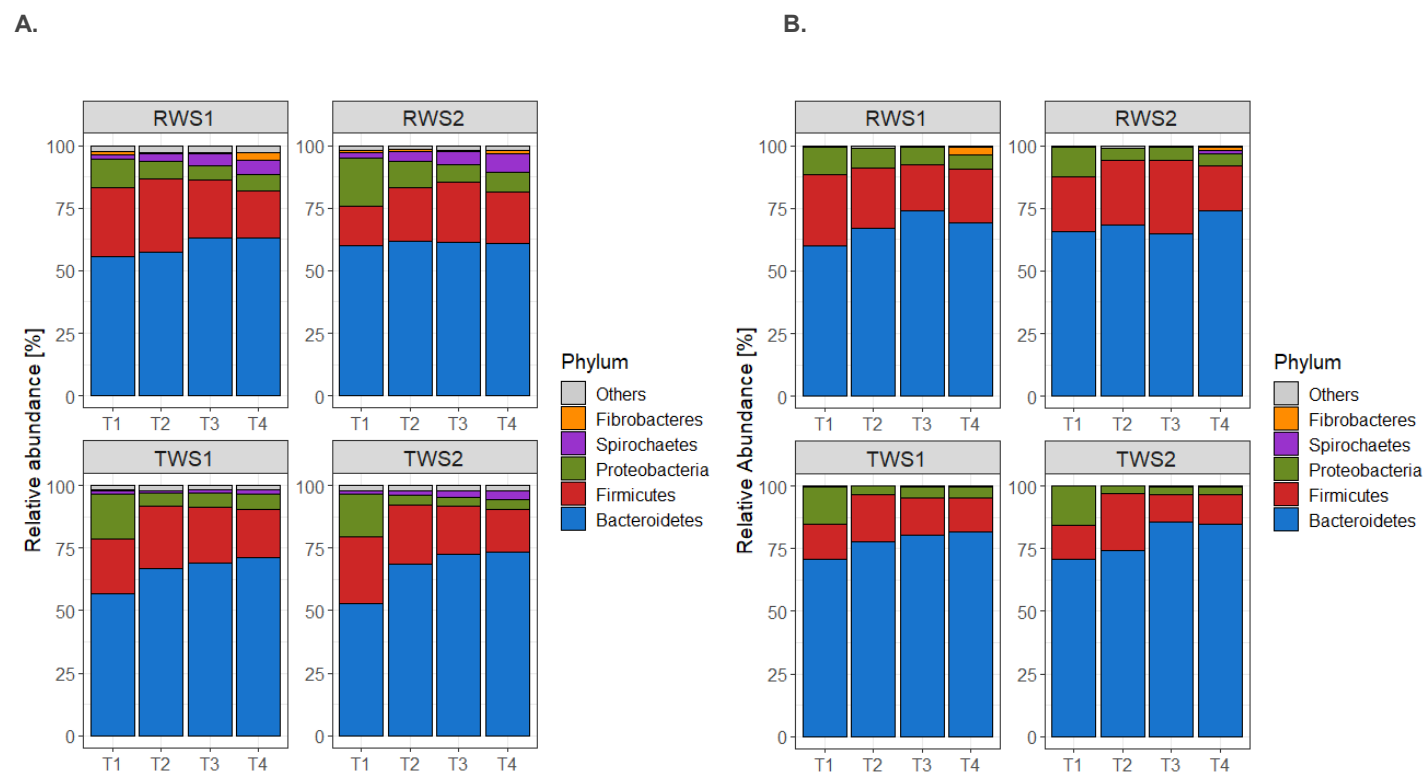


Figure III-S2 : Hierarchical clustering of metaproteomics data (CLR transformed), based in Euclidean distance and Ward D2 clustering method, showing the trend exhibited by the protein profile of the different samples of RWS (blue) and TWS (green) microbial consortia.

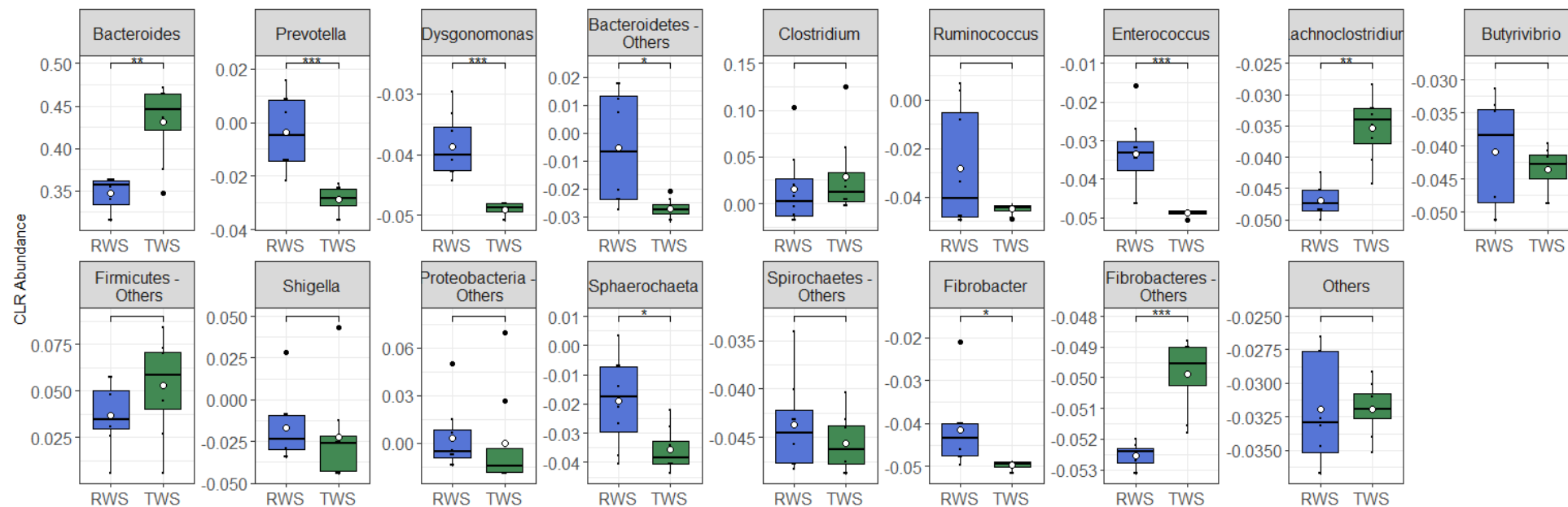
(3) Figure S3



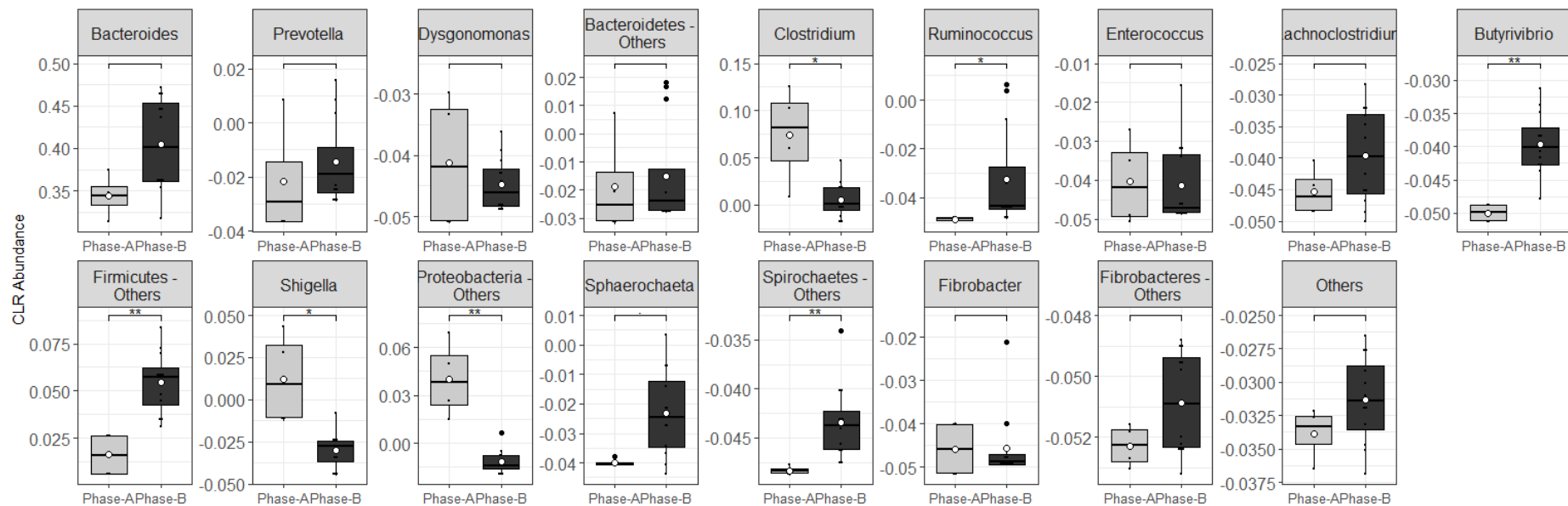
**Figure III-S3 : Relative abundance of bacterial phyla identified in metaproteomics (A) and 16S rRNA gene sequencing data (B).** Relative abundance of proteins based on NSAFs (normalized spectral abundance factors) or OTUs was aggregated at the phylum level for stacked bar plot representation. The group “Others” gather phyla with relative abundance less than 1% in the dataset. Proteins belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Spirochaetes (purple), Fibrobacteres (orange).

(4) Figure S4

A.



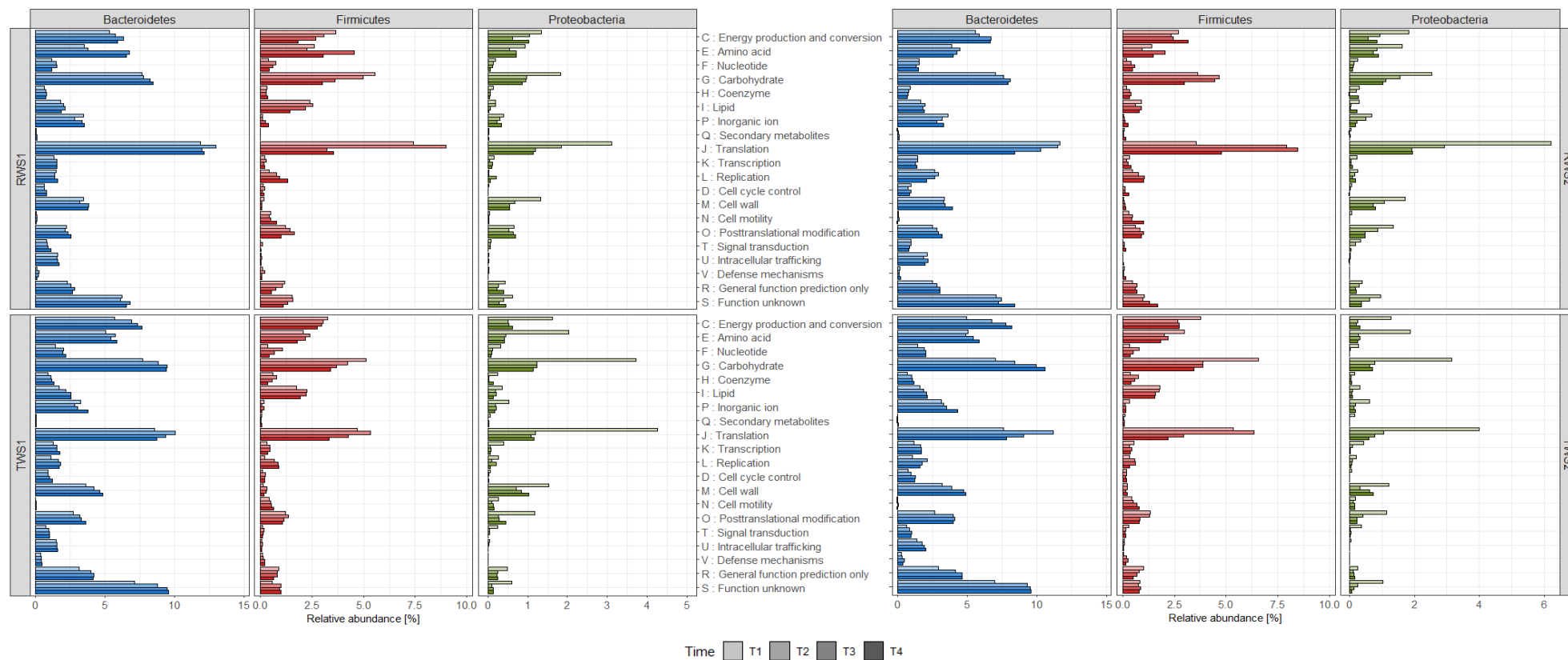
**B.**



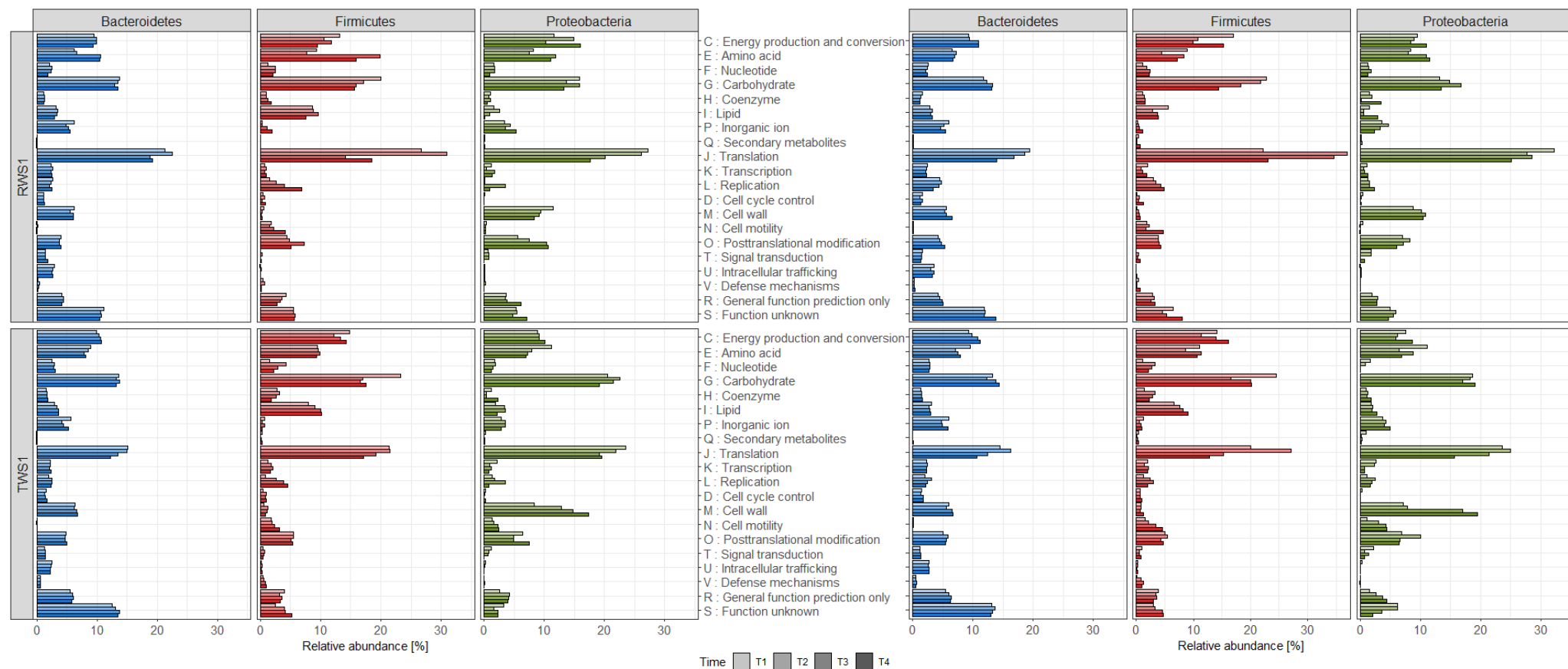
**Figure III-S4 : A.** Comparison of the protein abundance (CLR-transformed data) affiliated to the major genus present in RWS and TWS metaproteomes. **B.** Comparison of the proteins abundance (CLR-transformed data) affiliated to the major genus in phase A (T1 point) and phase B (T2-T3-T4 points). For the boxplot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  level of Wilcoxon rank sum test.

(5) Figure S5

A.



B.



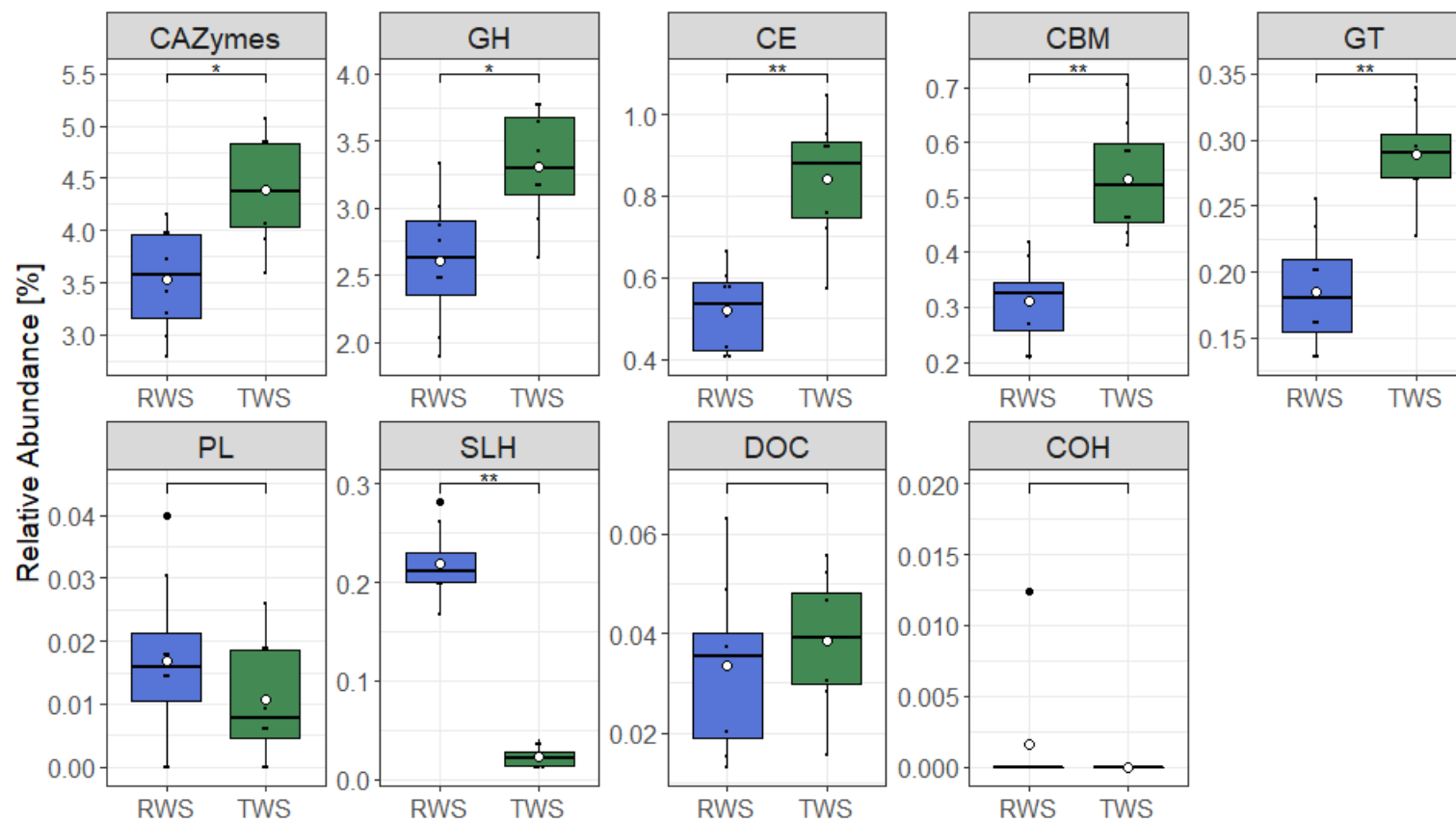
**Figure III-S5 : A.** Comparison of the expression level of clusters of orthologous groups (COGs) linked to three major phyla Bacteroidetes, Firmicutes and Proteobacteria found in the metaproteomes of RWS and TWS consortia replicates at different sampling times. Relative abundance of proteins is based on NSAF (Normalized spectral abundance factors) values. The predicted proteins identified were mapped to different COGs using PROPHANE pipeline and RPS-BLAST program. COGs affiliated to “Metabolism” functions: [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. COGs affiliated to “Information storage and processing” functions: [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair. COGs affiliated to “Cellular processes and signaling” functions: [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. COGs poorly described were: [R] General function prediction only, [S] Function unknown. **B. Comparison of the specific level of expression of clusters of orthologous groups (COGs) of three major phyla Bacteroidetes, Firmicutes and Proteobacteria found in the metaproteomes of RWS and TWS replicates at different sampling times.** Relative abundance of proteins is based on phyla-normalized NSAF (Normalized spectral abundance factors) values. The predicted proteins identified were mapped to different COGs using PROPHANE pipeline and RPS-BLAST program. COGs categories are the same as in **A**.

(6) Figure S6

A.



**B.**



c.

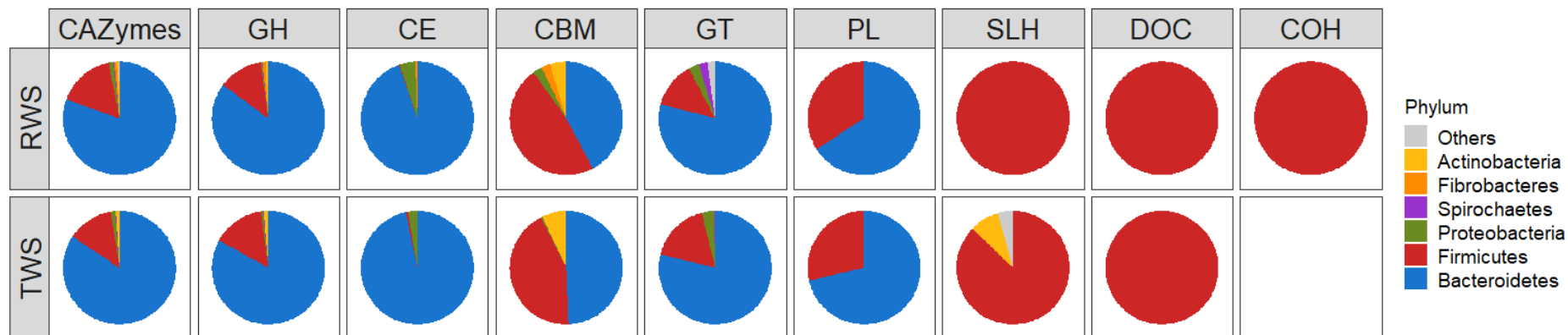
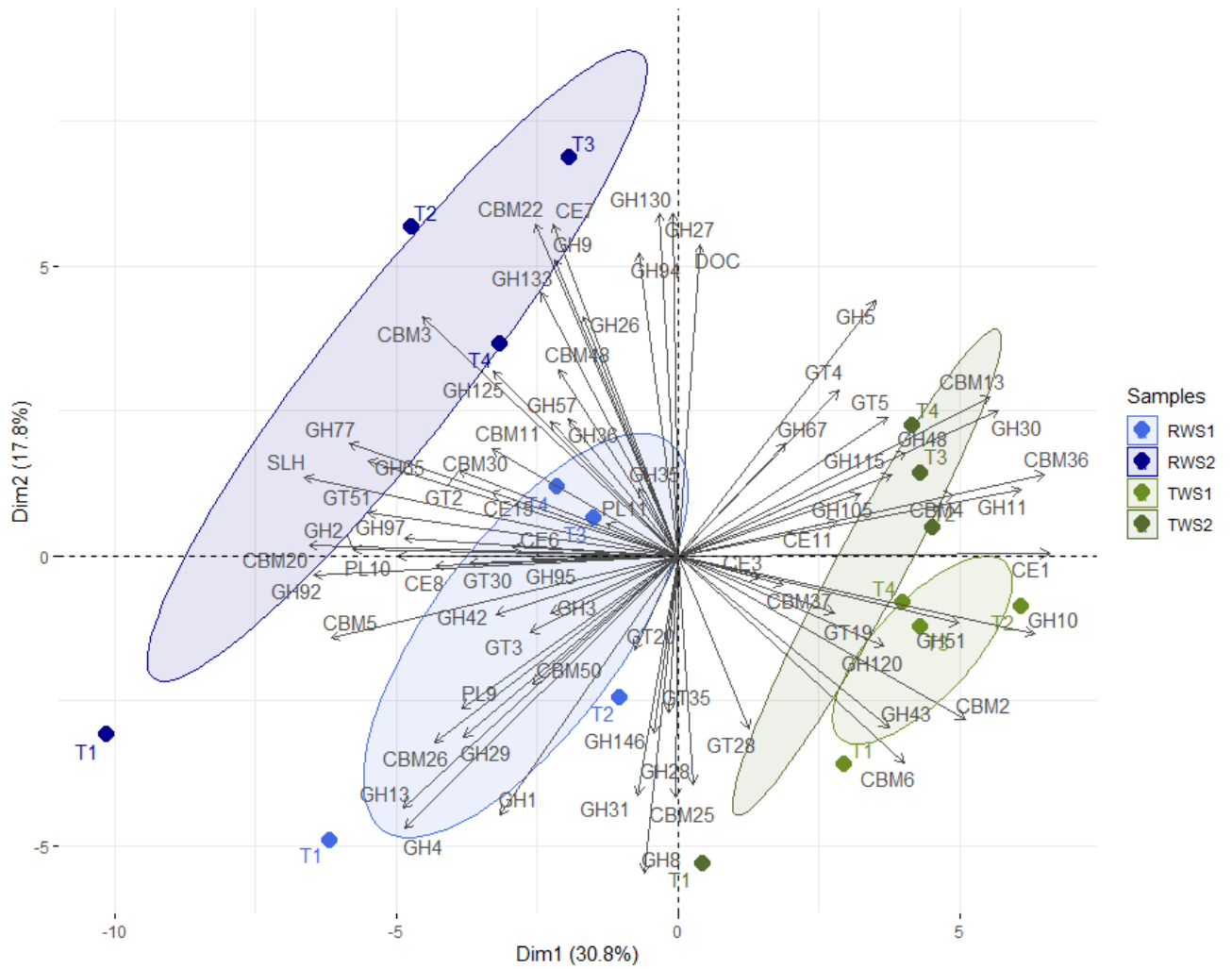


Figure III-S6 : **A.** Abundance of the CAZyme classes in RWS and TWS metaproteomes : glycoside hydrolases (GH), carbohydrate esterases (CE), carbohydrate binding modules (CBM), glycosyl tranferases (GT), polysaccharide lyases (PL), S-layer homology (SLH), dockerins (DOC), cohesins (COH). Relative abundance is based on average NSAF values. For the boxplot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. \*  $P < 0.05$ , \*\*  $P < 0.01$  level of Wilcoxon rank sum test. **B.** Venn diagram of shared CAZyme proteins (left) and families (right) between RWS and TWS metaproteomes. **C.** Phyla contribution to CAZyme domains (based on average NSAF values) in RWS and TWS. Proteins belonging to the same bacterial phylum are colored as follow: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Spirochaetes (purple), Fibrobacteres (orange), Actinobacteria (yellow), Others (gray).

(7) Figure S7

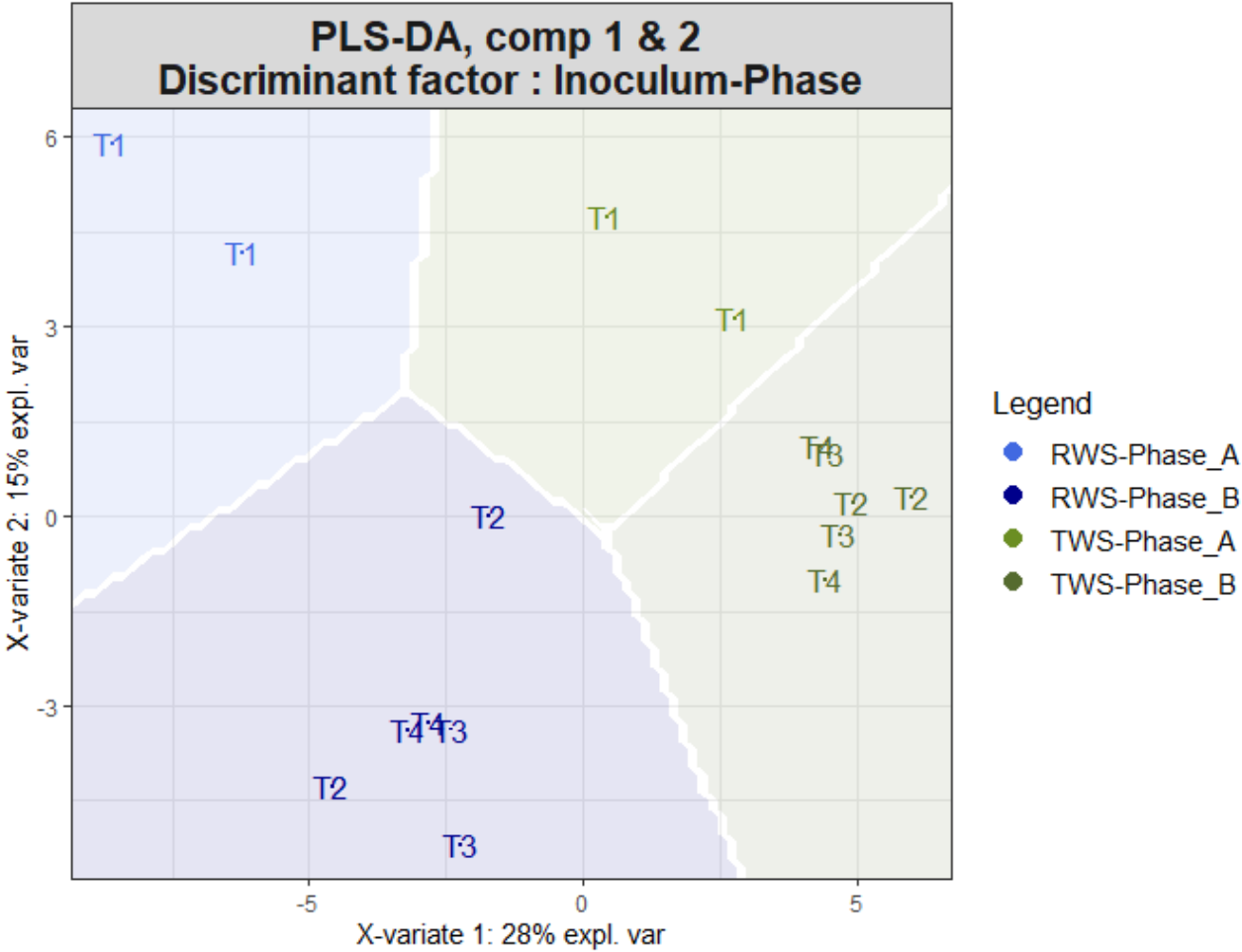
A.





(8) Figure S8

A.



B.

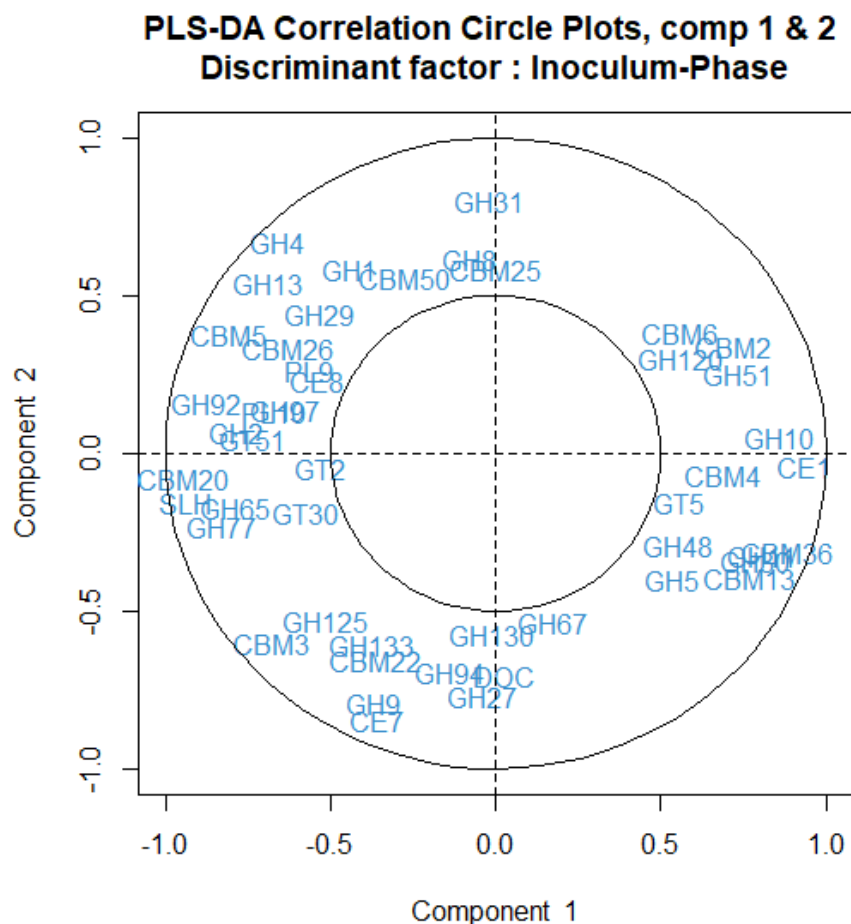
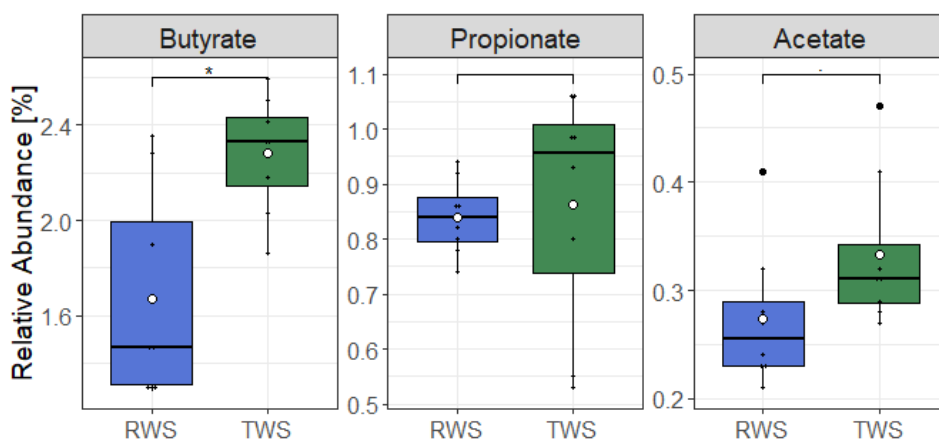


Figure III-S8 : A. Partial Least Square Discriminant Analysis (PLS-DA) of CAZyme expression in RWS and TWS consortia using CAZymes normalized abundance (CLR-transformed data). PLS-DA was built using the inoculum source (RWS or TWS) and the degradation phase (phase A corresponding to T1 samples and phase B corresponding to T2-T3-T4 points) as discriminant factors. Colors correspond to the consortium: blue for RWS and green for TWS consortium, and the phase: light colors for phase A and dark colors for phase B. B. The corresponding PLS-DA correlation circle plot.



(9) Figure S9



**Figure III-S9: Boxplots representing the aggregated abundance of proteins involved in volatile fatty acids production (acetate, propionate, butyrate) in RWS and TWS consortia.** For butyrate biosynthesis: COG4770 (acetyl/propionyl-CoA carboxylase), COG3426 (butyrate kinase), COG1250 (3-hydroxyacyl-CoA dehydrogenase), COG1024 (enoyl-CoA hydratase), COG0183 (acetyl/butyryl-CoA acetyltransferase), COG1979 (alcohol dehydrogenase YqhD), COG1454 (alcohol dehydrogenase, class IV), COG1028 (short-chain alcohol dehydrogenase), COG1064 (Zn-dependent alcohol dehydrogenase). For propionate biosynthesis: COG4799/0777 (acetyl/propionyl-CoA carboxylase), COG2185/COG1884 (methylmalonyl-CoA mutase), COG0346 (methylmalonyl-CoA epimerase). For acetate production: COG1012 (NAD-dependent aldehyde dehydrogenase), COG0282 (acetate kinase), COG0280 (phosphotransacetylase). For the box plot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. \*  $P < 0.05$  level of Wilcoxon rank sum test.

e) *Additional file 5: Supplementary Table S1*

**Supplementary Table S1. Reproducibility of label-free metaproteomics quantification of technical replicates estimated by Pearson correlation in RWS and TWS consortia replicates.**

<b>RWS 1</b>					
	<b>RWS1- T1_A</b>	<b>RWS1- T1_B</b>		<b>RWS2- T1_A</b>	<b>RWS2- T1_B</b>
<b>RWS1- T1_B</b>	0.909		<b>RWS2- T1_B</b>	<b>0.814</b>	
<b>RWS1- T1_C</b>	0.890	0.927	<b>RWS2- T1_C</b>	<b>0.447</b>	<b>0.462</b>
	<b>RWS1- T2_A</b>	<b>RWS1- T2_B</b>		<b>RWS2- T2_A</b>	<b>RWS2- T2_B</b>
<b>RWS1- T2_B</b>	0.930		<b>RWS2- T2_B</b>	<b>0.911</b>	
<b>RWS1- T2_C</b>	0.906	0.917	<b>RWS2- T2_C</b>	<b>0.901</b>	<b>0.897</b>
	<b>RWS1- T3_A</b>	<b>RWS1- T3_B</b>		<b>RWS2- T3_A</b>	<b>RWS2- T3_B</b>
<b>RWS1- T3_B</b>	<b>0.952</b>		<b>RWS2- T3_B</b>	<b>0.851</b>	
<b>RWS1- T3_C</b>	<b>0.944</b>	<b>0.953</b>	<b>RWS2- T3_C</b>	<b>0.746</b>	<b>0.714</b>
	<b>RWS1- T4_A</b>	<b>RWS1- T4_B</b>		<b>RWS2- T4_A</b>	<b>RWS2- T4_B</b>
<b>RWS1- T4_B</b>	<b>0.764</b>		<b>RWS2- T4_B</b>	<b>0.861</b>	
<b>RWS1- T4_C</b>	<b>0.768</b>	<b>0.935</b>	<b>RWS2- T4_C</b>	<b>0.850</b>	<b>0.840</b>

**RWS2**

Chapitre 3 – Analyse métabolomique des consortia microbiens lors de la dégradation de la lignocellulose pour la production des AGV

<b>TWS 1</b>			<b>TWS2</b>		
	<b>TWS1- T1_A</b>	<b>TWS1- T1_B</b>		<b>TWS2- T1_A</b>	<b>TWS2- T1_B</b>
<b>TWS1- T1_B</b>	0.825		<b>TWS2- T1_B</b>	0.924	
<b>TWS1- T1_C</b>	0.918	0.834	<b>TWS2- T1_C</b>	0.921	0.947
	<b>TWS1- T2_A</b>	<b>TWS1- T2_B</b>		<b>TWS2- T2_A</b>	<b>TWS2- T2_B</b>
<b>TWS1- T2_B</b>	0.934		<b>TWS2- T2_B</b>	0.868	
<b>TWS1- T2_C</b>	0.886	0.903	<b>TWS2- T2_C</b>	0.872	0.929
	<b>TWS1- T3_A</b>	<b>TWS1- T3_B</b>		<b>TWS2- T3_A</b>	<b>TWS2- T3_B</b>
<b>TWS1- T3_B</b>	0.900		<b>TWS2- T3_B</b>	0.928	
<b>TWS1- T3_C</b>	0.939	0.895	<b>TWS2- T3_C</b>	0.834	0.818
	<b>TWS1- T4_A</b>	<b>TWS1- T4_B</b>		<b>TWS2- T4_A</b>	<b>TWS2- T4_B</b>
<b>TWS1- T4_B</b>	0.902		<b>TWS2- T4_B</b>	0.923	
<b>TWS1- T4_C</b>	0.904	0.916	<b>TWS2- T4_C</b>	0.931	0.926

## **8. Acknowledgements**

We thank the GeT-PlaGe for performing sequencing. Authors would also like to Mrs Abadie for assistance with protein extraction as well as K. Eismann, S. Haange and V. Lünsmann for their help with metaproteomics protocols and data analysis. LC-MS/MS analysis was conducted using the facilities in the Proteomics Department of UFZ-Leipzig. The authors also thank M. Bounouba and E. Mangelle for their technical support with bioreactor experiments.

## **9. Funding**

This work was funded by the French National Research Agency (ANR) under the Project Hi-Solids ANR-14-CE19-0013-01, the French National Agency for Energy and the Environment (ADEME), the TRANSFORM department of INRAE, the Carnot Institute 3BCAR under the Insyme project. This study was also supported by the PHC Procope program from Campus France under the grant N° 3077WC.

## **10. Author contributions**

E.A carried out and analyzed metaproteomics and microbial diversity data under G.H-R supervision. A.L. realized the bioreactor experiments and prepared the samples for metaproteomics analysis under the supervision of G.H-R, M.O and N.J. N.J. performed and analyzed LC-MS/MS experiments. B.H and V.L participated on CAZy analysis. E.A performed the statistical data analysis under the supervision of S.D. As project coordinator, G.H-R designed the study, participated on experimental design and contributed at all stages. The manuscript was written by E.A. under G.H-R and M.O supervision with important intellectual contributions from all authors. All authors contributed and have given approval to the final version of the manuscript. G.H-R and N.J participated in funding this research.



## C. Conclusion

L'analyse métabolomique des deux consortia microbiens a démontré son potentiel pour l'étude des dynamiques taxonomiques et fonctionnelles de communautés lignocellulolytiques. Notre analyse métabolomique a permis d'identifier les similarités et les différences existantes entre les deux consortia à l'échelle protéique. Ceci a permis d'expliquer certains des résultats obtenus à d'autres niveaux d'analyses de ces consortia : composition microbienne, vitesse et profil de bioconversion de la biomasse, activités enzymatiques.

L'analyse métabolomique et les résultats de séquençage du gène codant pour l'ARNr 16S ont révélé que la composition taxonomique des deux communautés est très similaire et qu'une forte proportion des OTUs et des protéines sont affiliés au genre *Bacteroides* (Bacteroidetes), suivi du genre *Clostridium* (Firmicutes). Cette similarité compositionnelle est probablement due aux conditions identiques d'enrichissement sur le même substrat. L'annotation des protéines avec la base de données CAZy a permis de dévoiler la richesse des CAZymes présentes au sein des deux consortia microbiens, exprimées essentiellement par les phyla Bacteroidetes et Firmicutes, et plus spécifiquement par *Bacteroides* et *Clostridium*. Bien que des différences d'abondance aient été identifiées pour certaines familles CAZymes entre RWS et TWS, les familles exprimées uniquement par un seul des consortia sont assez rares. Les familles CAZymes présentant des différences significatives entre RWS et TWS sont exprimées surtout par des Bacteroidetes et des Firmicutes. Sur la base des résultats de cette analyse métabolomique et les propriétés fonctionnelles connues de ces genres, nous avons donc émis l'hypothèse d'une complémentarité fonctionnelle entre ces deux taxa pour la bioconversion de la lignocellulose. En effet, les Firmicutes sont connus pour être plutôt spécialisés dans la dégradation de la cellulose tandis que les Bacteroidetes dégradent préférentiellement les hémicelluloses. La grande diversité de familles CAZymes produites par les différents taxons montre la diversité métabolique de ces consortia. Ceci est un des avantages majeurs de l'utilisation des consortia microbiens par rapport à des souches pures, possédant des répertoires de CAZymes limités, qui leur permet d'hydrolyser de manière stable la biomasse lignocellulosique. Cette complémentarité fonctionnelle est également visible au niveau de l'expression des enzymes impliquées dans la biosynthèse des acides gras volatils, certains taxons contribuant de façon préférentielle à la synthèse de certains acides gras.

Cette première étude de métabolomique réalisée au laboratoire, appliquée à des communautés complexes enrichies en bioréacteur, n'a pas été sans difficultés et a démarré en 2014. Son coût

élevé, le temps consacré aux manipulations, et les limites des outils bioinformatiques ont conduit à l'analyse d'un nombre réduit d'échantillons et de réplicats, compromettant parfois la robustesse des tests statistiques. D'autre part, comme les écosystèmes étudiés ici comportent des bactéries encore peu ou pas décrites dans les bases des données, la réalisation d'un métagénome de référence permettrait d'améliorer l'identification et l'annotation des protéines. Enfin, les données obtenues en métagénomique shotgun sans marquage (dite « *label free* ») sont des données d'abondances relatives obtenues en calculant le NSAF (*Normalized Spectral Abundance Factor*) de chaque « *protein group* », et sont donc, au même titre que les données issues du séquençage de l'ARNr 16S, susceptibles d'être sujettes à des biais compositionnels. La correction de ce biais dans les données omiques est un domaine en pleine maturation parmi les biologistes et les statisticiens. La méthode retenue ici, la transformation CLR (*Centered Log Ratio*), est parfois difficile à utiliser et interpréter (voir la section « A propos de la transformation CLR des données compositionnelles»). Il est encore difficile de se passer de représentations graphiques et d'interprétations basées sur les abondances relatives pour la description de la composition taxonomique ou fonctionnelle des communautés. Néanmoins, la transformation CLR permet d'atténuer le caractère compositionnel intrinsèque aux données omiques, pour établir des conclusions valables à partir des analyses multivariées et des tests statistiques appliqués à l'étude des consortia microbiens.







## IV. Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

### A. Introduction

Les prétraitements de la biomasse lignocellulosique sont variés mais essentiels pour améliorer les rendements de bioconversion. Leurs effets ont été largement documentés et ont permis de caractériser certains paramètres du substrat déterminant une meilleure hydrolyse (et fermentation) tels que la surface du substrat accessible aux enzymes, la capacité de rétention d'eau et la cristallinité du substrat, la saccharification enzymatique, ou le taux de lignine et des liaisons ester. Bien que l'impact du type du substrat lignocellulosique (paille de blé, de maïs, de riz, bois...) sur le métabolisme chez des souches pures ou sur la sélection des communautés microbiennes a été largement étudié, peu d'études ont évalué l'effet du prétraitement d'un substrat sur la structure et le fonctionnement métabolique d'une communauté microbienne.

Dans ce chapitre, une analyse par métabolomique *shotgun* sans marquage du consortium microbien RWS (précédemment étudié dans le chapitre 3) est présentée. Ce consortium a été mis en culture avec le substrat avec lequel il a été enrichi, c'est-à-dire de la (i) paille de blé broyée à 2 mm, mais aussi avec ce même substrat ayant subi un autre prétraitement mécanique et/ou chimique : (ii) paille de blé broyée à 2 mm et imprégnée à la soude, (iii) paille de blé broyée à 0.1 mm uniquement et (iv) paille de blé broyée à 0.1 mm et imprégnée à la soude.

En s'appuyant sur les résultats précédemment publiés sur la caractérisation des cinétiques obtenues pour chacun de ces 4 substrats (taux et vitesse de dégradation, taux et profil de production des acides gras volatils et suivi de l'activité xylanase au cours du temps - Lazuka et al., (2017)), cette analyse métabolomique a pour objectif d'étudier finement la réponse d'un consortium microbien à de modifications physiques, structurelles et éventuellement biochimiques du substrat induites par des prétraitements, permettant de faciliter l'hydrolyse et la bioconversion

#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

---

de la biomasse. Il s'agit d'identifier les taxons et les fonctions impactés par les prétraitements et de les mettre en relation avec les variations de bioconversion observées selon les différents substrats prétraités utilisés. Pour cela, quatre prélèvements ont été sélectionnés à des niveaux similaires en termes de cinétique (même taux de dégradation de la biomasse). Des groupes bactériens et des fonctions enzymatiques spécifiquement stimulés par les modifications de substrat induites par les prétraitements peuvent fournir des outils de pilotage des consortia microbiens mis en œuvre dans des procédés.

## **B. Impact of substrate changes induced by dry-alkali pretreatment on the lignocellulolytic bioconversion of rumen-derived microbial consortium**

Emeline Auer<sup>1</sup>, Nico Jehmlich<sup>2</sup>, Vincent Lombard<sup>3,4</sup>, Bernard Henrissat<sup>3,4,5</sup>, Guillermina Hernandez-Raquet<sup>1\*</sup>

<sup>1</sup>) Toulouse Biotechnology Institute - TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, F-31077 Toulouse, France.

<sup>2</sup>) Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, Germany.

<sup>3</sup>) CNRS UMR 7257, Aix-Marseille University, F-13288, Marseille, France.

<sup>4</sup>) USC 1408 AFMB, INRAE, F-13288, Marseille, France.

<sup>5</sup>) Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

AUER Emeline emeline.auer@inrae.fr

JEHMLICH Nico nico.jehmlich@ufz.de

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](mailto:hernandg@insa-toulouse.fr)

---

## 1. Abstract

### Background

Understanding the enzymatic deconstruction of lignocellulose (LC) is of increasing industrial and social concern to develop the bioeconomy. Microbial consortia enriched on lignocellulosic feedstocks represent an interesting source of complex mixture of carbohydrate active enzymes (CAZymes) required for an efficient biomass bioconversion. The mixture of enzymes needed for LC deconstruction varies in function of the physico-chemical properties of the biomass substrate. In the current strategy, pretreatment, by modifying the substrate features, is an unavoidable step to facilitate the enzyme activity and reduce the biomass recalcitrance. While the response of single strains to substrate changes has been well characterized, little is known on the response of microbial consortia to substrate pretreatments. We elucidated how stable microbial consortium adapt to transform wheat straw, in different particle size (0.1 and 2 mm) and pretreated or not by dry-alkaline impregnation. This new insight could help to understand the mechanisms of LC deconstruction and propose optimal synthetic microbial consortia and possible enzymatic cocktails required for a more efficient biomass conversion.

### Methods

A lignocellulolytic microbial rumen-derived consortium, named RWS, was incubated over 15 days, with untreated or pretreated wheat straw under anaerobic conditions. We investigated the diversity of the active RWS community members, their carbohydrate-degrading potential and their metabolic capacities to convert LC into volatile fatty acids (VFA) by metaproteomics and multivariate analysis to study how substrate modifications affect the community structure and the functional expression.

### Results

Fermentations results demonstrated that dry-alkaline pretreatment, whatever the particle size of substrate, enhanced the deconstruction of the wheat straw. Metaproteomics revealed that although the taxonomic profile across these pretreatment conditions were relatively similar, the abundance of certain genera, such as *Bacteroides* and *Lachnospirillum*, increased at the expense of others, suggesting adaptations to the substrates properties. *Bacteroides* are particularly known

for their high metabolic capability to hydrolyze holocellulosic biomass without substrate attachment.

### **Conclusions**

In summary, metaproteomic analysis have enabled to identify the active populations impacted by substrates changes induced by chemical pretreatment. We also identified interesting CAZy families which could be exploited for the formulation of efficient enzyme cocktails, in combination to chemical pretreatment, to improve the industrial bioconversion of biomass.

### **Keywords**

Lignocellulose bioconversion; anaerobic consortium; CAZyme; metaproteomics; alkali pretreatment

## 2. Introduction

Lignocellulose is the most abundant renewable organic carbon source for the production of biofuel and commodity chemicals and consists mainly of cellulose, hemicelluloses and lignin (Chandel and Singh, 2011). However, biomass bioconversion is still challenging because of the complex structure of lignocellulose (Mosier et al., 2005). Its deconstruction in natural environments involves complex lignocellulolytic microbial communities able to produce a large panel of carbohydrate-active enzymes (CAZymes) which break down, modify or build glycosidic bonds (Lombard et al., 2014). CAZymes are classified in five main categories in the CAZy database, namely glycoside hydrolases (GH), glycosyltransferases (GT), carbohydrate esterases (CE), polysaccharide lyases (PL), auxiliary activities (AA), and their appended non-catalytic carbohydrate-binding modules (CBM). Cellulosomes, which are complex assemblies of enzymes that allow bacteria to bind to the substrate, require the presence of cohesins (COH) and dockerins (DOC) and S-layer homology (SLH) domains at the cell surfaces.

Currently, lignocellulose bioconversion strategies integrate pretreatment methods to enhance the enzymatic saccharification of lignocellulose materials, to reduce biomass recalcitrance and improve bioconversion yields. Among them, mechanical and chemical pretreatment are two well-established pretreatment approaches. The impact of these pretreatments on substrate structure and composition and its consequences on biomass digestibility has been extensively studied (Mosier et al., 2005; Zhao et al., 2008; Hendriks and Zeeman, 2009; Guo et al., 2011; Monlau et al., 2012). It has been demonstrated that the combination of these two strategies strongly enhanced the biomass digestibility of pretreated wheat straw by an anaerobic rumen-derived microbial consortium (Lazuka et al., 2017) as the chemo-mechanical pretreatment could : (i) decrease particle size and increase total accessible specific surface area which gives more possibilities to enzymes to interact with the substrate, (ii) improve biomass porosity, (iii) induce cellulose modifications by reducing the crystallinity and swelling cellulose fibers, (iv) break lignin-carbohydrates linkages. Dry-alkali pretreatments are known to have various effects on hemicelluloses, on which they can promote depolymerization by cleaving hydrolysable linkages in lignin and glycosidic bonds of polysaccharides, swell of the fibers, saponification of acetyl and uronic esters ester intermolecular bonds, and “peel” off reducing end groups (Hendriks and Zeeman, 2009). Some studies suggest that the diversity and the carbohydrate metabolism of microbial communities varies in function of the feedstock's composition and physicochemical properties (Reddy et al., 2011; Eichorst et al., 2014; Jiménez et al., 2016). Nevertheless, little is

known on the response of microbial community to such substrate modifications induced by pretreatment, particularly at CAZymes level and enzymes involved in the VFA synthesis. A few studies have identified that the secreted profiles of fungal and bacterial lignocellulolytic strains, growing in pure culture, and even the composition of the cellulosome and other hydrolytic enzymes were dependant on substrate features and carbon sources (Raman et al., 2009; de Souza et al., 2011; Delmas et al., 2012; Lü et al., 2017; Ren et al., 2019). Also, previous studies demonstrated that the composition of biomass substrates influences the diversity and enzyme profile of microbial consortia enriched on specific substrate (Gladden et al., 2012; Jiménez et al., 2015, 2016; Zhu et al., 2016). Microbial consortium would also adapt its CAZymes expression to the substrate modifications induced by pretreatment. Elucidating the impact of substrate changes induced by pretreatments on the community structure of stable lignocellulolytic microbial consortium, called RWS, and its dynamic response of enzymes expression will provide insights to the lignocellulosic communities. Several bacterial groups and enzyme functions specifically stimulated by pretreatments were identified and provide tools for driving the microbial consortia to the improvement of lignocellulosic biomass bioconversion processes.

### **3. Materials and methods**

#### ***a) Lignocellulosic pretreated substrates***

Wheat straw of the Koreli variety was milled to 2 mm using a knife mill (Retsch SM 100, Germany). It was named substrate "A - 2 mm" and used for the enrichment of the RWS microbial consortium. For pretreatment, this substrate A was dry-impregnated or not with dry sodium hydroxide at 5% (w/w), combined or not to milling (0.1 mm) as previously described (Lazuka et al., 2017). Pretreated wheat straws were called "B - 0.1 mm", "C - 2mm NaOH" and "D - 0.1 mm NaOH". The physico-chemical characterization of raw and pretreated substrates was previously published (Lazuka et al., 2017).

#### ***b) Microbial fermentation conditions***

For each pretreated substrate, lignocellulose degradation was assessed in duplicate anaerobic batch bioreactors (2L BIOSTAT® A+, Sartorius, Germany) as previously described (Lazuka et al., 2015, 2017). Briefly, the raw wheat straw and pretreated substrates (20 g.L<sup>-1</sup>) were used as the sole carbon source in a mineral medium supplemented with vitamins. Duplicate biological bioreactors were inoculated (10% v/v) with the cow-rumen derived consortium (RWS), previously enriched on wheat straw substrate "A - 2 mm" (Lazuka et al., 2015). Bioreactors were operated at

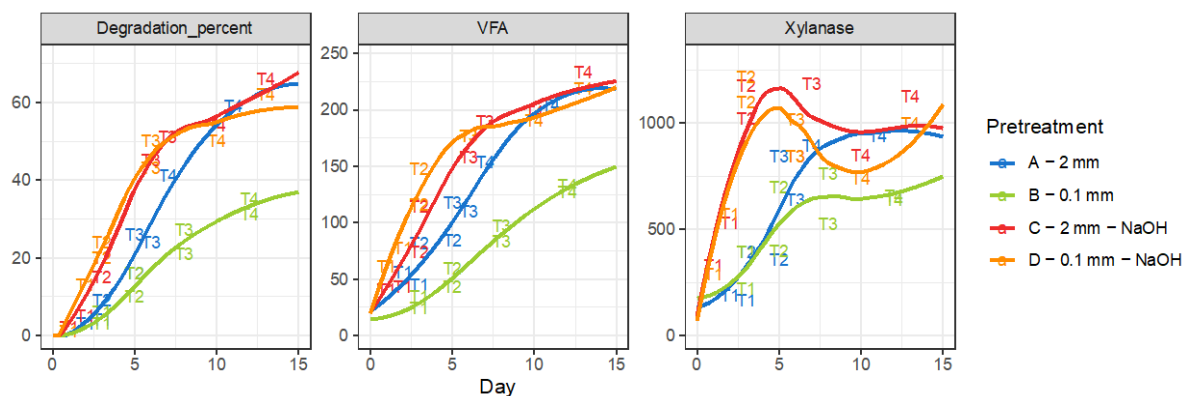


## Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

35°C under anaerobic conditions at controlled pH (6.15), in order to favor carboxylates production, as detailed by Lazuka et al. (2015). If methane was detected during the experiment, a solution of 2-bromoethanesulfonate (BES), a methanogenesis inhibitor, was injected (maximum concentration of 10 mM) (Webster et al., 2016).

### c) Metaproteomics analysis

To assess the temporal dynamics of microorganisms and expressed proteins with the different substrates along their deconstruction, shotgun metaproteomics analysis were performed on four sampling points for each biological replicate bioreactor. To integrate the differences in kinetic parameters induced by pretreatments, time point selection was based on similar level of wheat straw degradation, VFA production rate, and xylanase activity profiles (Figure IV-1). The four sampling points were named T1 to T4, with T1 corresponded to the maximum of degradation rate and xylanase activity, T2 preceded the inflection point, T3 was at the inflection point and T4 at the plateau.



**Figure IV-1 : Wheat straw degradation percentage (%), VFA production (mCmol/L) and xylanase levels (UA/L) over time for each pretreatment (mean of two replicates per pretreatment, coloured as detailed in legend). For each replicate and pretreatment, the sampling points were indicated as T1 to T4 labels.**

For each duplicate bioreactor, at least three technical replicate samples (3 mL) were collected at the selected time points, snap-frozen in liquid nitrogen and stored at -80°C. Protein extraction was then performed according to the procedure dedicated to complex sediment samples (Lin et al., 2016) and liquid chromatography-tandem mass spectrometry analysis and spectra identification

were realized with Proteome Discoverer according to the same conditions described previously for RWS samples (Auer et al., submitted – Chapter 3). Taxonomic and functional predictions were achieved by the bioinformatics pipeline PROPHANE ([www.prophane.de](http://www.prophane.de), Schneider et al., 2011) as previously mentioned (Auer et al., submitted – Chapter 3) and the normalized spectral abundance factor (NSAF) (Zybailov et al., 2006) was calculated for each protein. CAZymes annotation was assessed using the annotation pipeline used for the day-to-day updates of the CAZy database as indicated in Svartström et al., (2017).

#### **d) Data analysis and visualization**

Replicate-to-replicate variation was assessed by Pearson correlation analysis using the *cor* R function. Technical replicates with Pearson correlation value below than 0.7 were discarded from metaproteomics data. Protein groups present in only one of the technical replicates were considered as false positives and suppressed, whereas remaining proteins were expressed as mean values after normalization by scaling between 0 and 1. Annotated abundance tables were analyzed using the R packages *dplyr* and *tidyr*. Taxonomy, CAZy families and VFA-related proteins were grouped based on their annotation, and their relative abundances were summed and normalized by Centered Log Ratio (CLR) transformation (Calle, 2019) using an offset of 1. “Others” category gathered phyla with relative abundances lower than 2% in every samples, and genera lower than 2% were either regrouped in “Others” or in “Others” of each specific phylum. Barplots and boxplots were plotted using R’s *ggplot2* package (v3.3.5). Differences between conditions were tested using non-parametric Wilcoxon rank sum tests (also called Mann-Whitney test) (Xia and Sun, 2017) and p-value resulting from pairwise comparisons were adjusted using False Discovery Rate (FDR) adjustment method using *ggpubr* package (v0.4.0). Non-Metric Multidimensional Scaling (NMDS) ordinations and Permutational Multivariate Analysis of Variance (PERMANOVA) analyses (Anderson, 2001) were performed using R package *vegan* (v2.5-7) and Euclidean distance based on CLR-transformed data with 9999 permutations. In functional analyses using CAZy families and VFA-related proteins, the allocation index (AI) was defined as the ratio between the genus-specific relative abundance of a given function (e.g. CAZy families or VFA-related proteins) to the total relative abundances of all proteins for that genus. When compared between pretreatments, the AI ratio reflects differences in the proportion of proteins in a genus dedicated to the function under study. A significant variation of AI ratio therefore reflects a change in the proteins expression related to that function. Principal Component Analyses (PCA) were

Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

---

performed using *mixOmics* package (v6.12.2) (Rohart et al., 2017), based on protein abundances after CLR transformation, without genus or functional aggregation step.

#### 4. Results and discussion

##### *a) Positive impact of dry-alkali pretreatment on microbial digestion of wheat straw*

To assess the lignocellulose degradation efficiency of a microbial consortium derived from cow rumen (RWS) growing on each substrate, duplicate bioreactors were conducted with wheat straw pretreated only by comminution (A - 2 mm or B - 0.1 mm; thereafter referred as non-alkali pretreated substrates) or dry-alkali pretreated (C - 2 mm NaOH or D - 0.1 mm NaOH) as sole carbon source. The impact of these four pretreatments on the carboxylate production and substrate degradation rate by RWS were previously described (Lazuka et al., 2017); the main fermentation results are briefly summarized (Figure V-1). The present study extends our previous results by investigating the impact of substrate pretreatment on the RWS profile of CAZymes and proteins related to VFA synthesis throughout the LC degradation process.

Biochemical analysis of pretreated substrates revealed no significant compositional changes between raw (substrate A) and pretreated substrates (substrates B to D), with an average composition of  $43.6\pm 3.4\%$  (w/w) cellulose,  $23.2\pm 1.4\%$  hemicelluloses and  $19.7\pm 1.7\%$  lignin. However, dry-alkali pretreated substrates C and D displayed higher amounts of soluble reducing sugars. Fermentation results underlined a positive impact of the dry-alkali pretreatment on LC bioconversion in terms of lignocellulose degradation and VFA production rate. Indeed, considering the same particle size, LC degradation after 15 days of incubation was significantly higher for alkali-pretreated substrates compared to the non-alkali pretreated ones. Maximum degradation rates were observed earlier for dry-alkali pretreated substrates (C and D), compared to substrates only pretreated by comminution (A and B). Measurement of VFA production showed a similar trend. The highest VFA production rates were measured for dry-alkali pretreated substrates C and D at day 3 and day 2 respectively, while substrates A and B reached peak VFA production on day 5. However, the final VFA production was similar at 15 days for substrates A, C and D, while pretreatment B had a negative impact on the final VFA production. Concerning enzyme activity, a positive impact of the dry-alkali pretreatment on the xylanase activity of RWS was also observed. RWS fermentation of the milled substrates (A and B) resulted in a maximum xylanase activity of  $0.61\pm 0.14$  UA.mL<sup>-1</sup> and  $0.81\pm 0.11$  UA.mL<sup>-1</sup>, respectively, after 5 and 7 days of incubation, followed by a stable activity. In contrast, when RWS was grown on dry-alkali pretreated straws (C and D), it reached higher maximum xylanase activity earlier ( $1.2$  UA.mL<sup>-1</sup> after just 3 days of incubation).

**b) Metaproteomic samples description**

To decipher the impact of dry-alkali pretreatment on the RWS functional community structure, and particularly on its enzymatic profile relevant for its lignocellulolytic activity, we investigated the taxonomic composition of the functional community, deduced from the proteins affiliation, as well as the profile of CAZymes and proteins involved in the VFA synthesis by RWS growing in each substrate. We performed a metaproteomic analysis of four sampling points per biological duplicate bioreactors, as detailed in the Elucidating the impact of substrate changes induced by pretreatments on the community structure of stable lignocellulolytic microbial consortium, called RWS, and its dynamic response of enzymes expression will provide insights to the lignocellulosic communities. Several bacterial groups and enzyme functions specifically stimulated by pretreatments were identified and provide tools for driving the microbial consortia to the improvement of lignocellulosic biomass bioconversion processes.

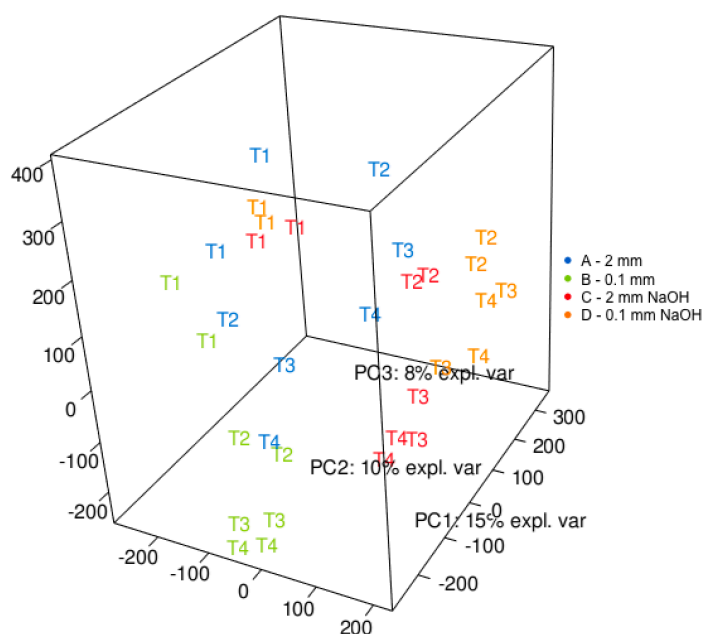
Materials and methods section and in Figure IV-1. For each time point analyzed, their reproducibility was assessed using Pearson's correlation (Additional file 1: Supplementary Table S1). Two technical replicates were removed due to their low correlation coefficient ( $< 0.7$ ).

A similar number of proteins was found in all samples (from 2,200 to 4,363 proteins), regardless of substrate or sampling time, suggesting similar coverage for all samples. However, only 22.6% of the total proteins (2,918 from 12,911 proteins) were common to all substrates (Additional file 3: Figure S1). Nevertheless, these shared proteins contributed substantially to the total protein abundance in the dataset ( $76.7 \pm 1.1\%$ ). Considering only whether the substrates were pretreated with dry-alkali or no, it could be noted that they shared 5,869 proteins, accounting  $89.9 \pm 3.4\%$  of the total protein abundance, while 3,506 and 3,536 proteins were specific to non-alkali pretreated ( $10.2 \pm 3.2\%$ ) and dry-alkali pretreated samples ( $9.9 \pm 3.8\%$ ), respectively. Thus, the metaproteomic profile of RWS growing on the different substrates shared the most abundant proteins, constituting the core proteome. This suggests that differences in the abundance of certain proteins or functions expressed by RWS on specific substrates could be responsible for the differences observed in fermentation results.

Principal Component Analysis (PCA) of the protein abundance normalized by Centered Log Ratio (CLR)-transformation (Figure IV-2) discriminated the RWS protein profiles produced in each substrate according to different factors for each of the first three principal components (PC). PC1, which explains 15% of the variance, separated the metaproteomic profiles of RWS growing on

## Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

non-alkaline pretreated substrates (A and B) from those of RWS growing on dry-alkaline pretreated substrates (C and D). PC2 (10%) and PC3 (8%) separated samples over time, with initial time points (T1) being well discriminated from the others. This temporal dynamics of the protein profile has already been noticed for RWS incubated on raw substrate A (Auer et al., submitted – Chapter 2) and appears to be preserved in the pretreated substrates. Based on this observation, two degradation phases were defined: Phase A with T1 points, and Phase B with T2, T3 and T4 points. PCA did not separate the protein profiles of RWS found on alkali pretreated substrates (C and D), suggesting that these substrate pretreatments have similar effects on the proteomic profile of RWS.

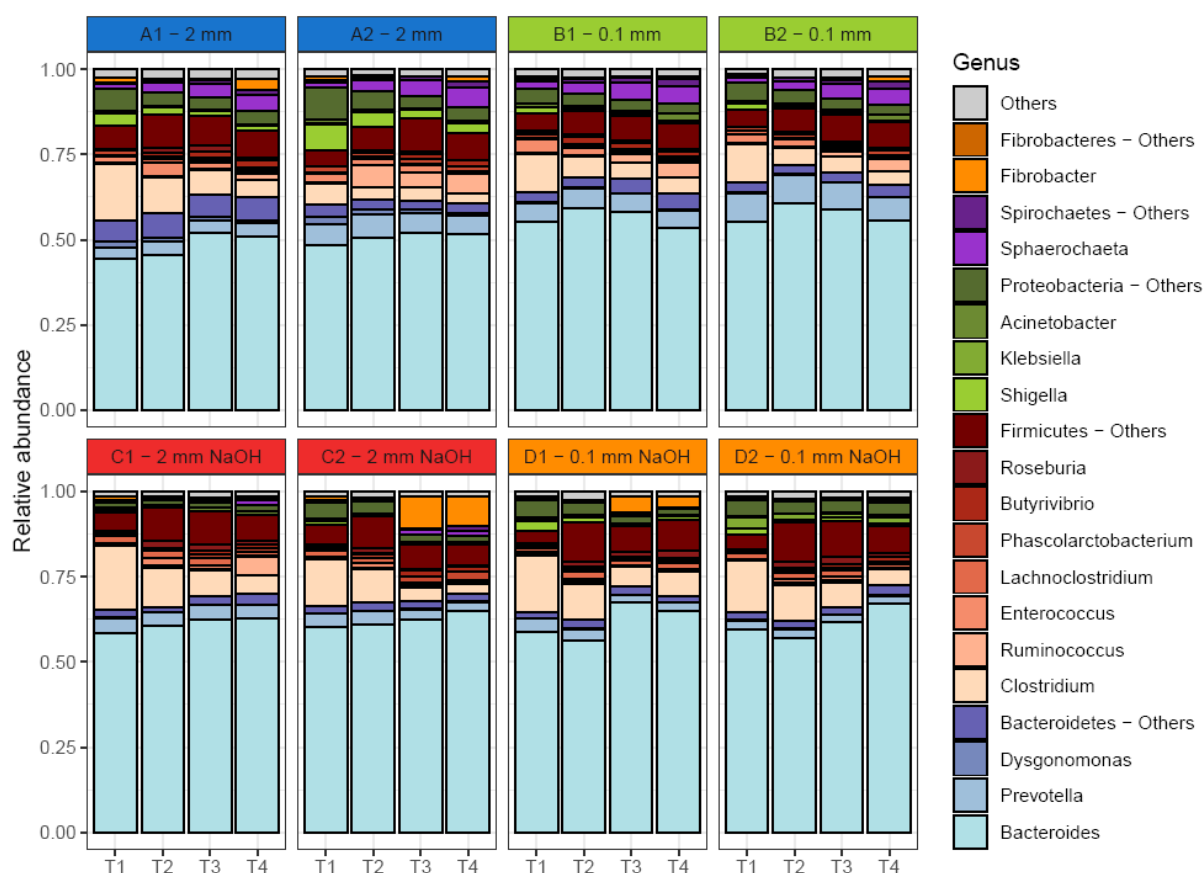


**Figure IV-2 : Three-dimensional Principal Component Analysis (PCA) of Centered Log Ratio (CLR)-transformed metaproteomic data (PC1 to PC3).** Colors reflect the different substrates as indicated in legend (blue, A - 2 mm; green, B - 0.1 mm; red, C - 2 mm NaOH; orange, D -0.1 mm NaOH). T1 to T4 labels indicate the sampling time as presented in the Elucidating the impact of substrate changes induced by pretreatments on the community structure of stable lignocellulolytic microbial consortium, called RWS, and its dynamic response of enzymes expression will provide insights to the lignocellulosic communities. Several bacterial groups and enzyme functions specifically stimulated by pretreatments were identified and provide tools for driving the microbial consortia to the improvement of lignocellulosic biomass bioconversion processes.

Materials and methods section.

**c) Impact of substrate modifications induced by dry-alkali pretreatment on active populations throughout lignocellulose degradation**

Taxonomic affiliation of the proteins revealed that Bacteroidetes (65.6±4.2%), followed by Firmicutes (22.1±4.3%), were the dominant protein-producing taxa in RWS, regardless of the substrate condition studied (Figure IV-3). At the genus level, *Bacteroides* and *Clostridium* provided most of the proteins identified in all substrate conditions (57.4±5.8% and 8.1±4.5%, respectively). Minor proportions of proteins belonged to Proteobacteria (6.1±3.2%), Spirochaetes (2.8±2.3%) and Fibrobacteres (1.4±1.4%), showing some notable differences between biological replicates.



**Figure IV-3 : Taxonomic community composition of RWS deduced from identified proteins for each time point and substrate condition in the biological replicate bioreactors.**

Biological replicates are identified with numbers 1 and 2 associated with pretreatments labels. Relative abundance of proteins based on NSAFs values (Normalized Spectral Abundance Factors) were aggregated at the genus level for stacked bar plot representation. “Others” group gather phyla with relative abundance less than 2% in the dataset. Within major phyla,

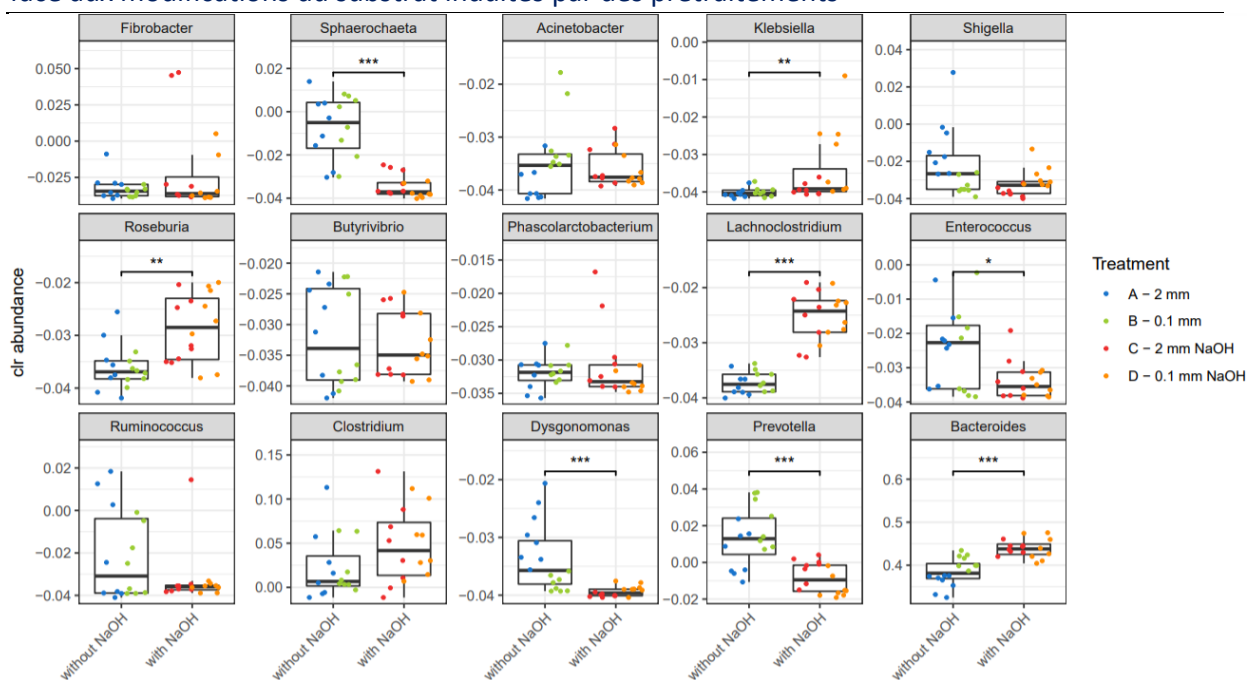
the group "Others" gather genera with relative abundance less than 2% in the dataset or unclassified genera. Proteins belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Spirochaetes (purple), Fibrobacteres (orange).

NMDS ordination confirmed that dry-alkali pretreated substrates resulted in similar taxonomic proteins profile for RWS, clustering all corresponding points (Additional file 3: Figure S2). The temporal separation of protein profiles in two degradation Phases A (T1) and B (T2, T3 and T4) was also confirmed, with T1 points being well separated from the others. Using "Degradation\_phase" (Phase A versus Phase B), "Particle\_size" (2 mm versus 0.1 mm) and "NaOH" (alkaline versus non-alkali pretreatment) as explanatory variables, PERMANOVA analysis allowed to quantify their contribution to the dispersion observed in NMDS analysis. "NaOH" showed the strongest effect, with an explanatory power (R<sup>2</sup>) of 29% ( $p < 0.001$ ), followed by the "Degradation\_phase" (R<sup>2</sup> =22%,  $p < 0.001$ ) while "Particle\_size" was barely significant (R<sup>2</sup>=5%,  $p < 0.05$ ). Therefore, the main effect of substrate pretreatment on the taxonomic profile of RWS proteins was induced by dry-alkali pretreated substrates, while the impact of substrate particle size was lesser.

Since the "NaOH" pretreated substrates had a major effect on the proteins profile of RWS, the protein abundance affiliated to each major genus found in RWS was therefore compared between the fermentation of substrates pretreated or not by dry-alkali (Figure IV-4).



## Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements



**Figure IV-4 : Comparison of the protein abundance affiliated to major genera found in RWS during fermentation of non-alkali (A and B) and dry-alkali pretreated (C and D) substrates (CLR-transformed data).** Colored dots represent the values of each sample (8 samples for each substrate). The median, first and third quartiles are represented. Asterisks indicate the statistically significant differences between NaOH treated and non-treated substrates (\*\*\*:  $< 0.001$  ; \*\*:  $< 0.01$  ; \*:  $< 0.05$ ). Wilcoxon adjusted p-value according to the FDR method.

The CLR abundances of proteins affiliated to *Prevotella*, *Dysgonomonas*, *Enterococcus* and *Sphaerochaeta* were negatively impacted by alkali pretreated substrates; proteins from these genera were more abundant in RWS community growing on substrates without alkali pretreatment. At the opposite, proteins from *Bacteroides*, *Lachnospirillum*, *Klebsiella* and *Roseburia* were significantly more abundant when RWS consortium was incubated with dry-alkali pretreated substrates. *Klebsiella* was reported as purveyor of extracellular cellulases to catalyze hydrolysis of pretreated lignocellulosic substrate (Jones et al., 2018) and hemicelluloses debranching enzymes in corn stover-adapted microbial consortium (Zhu et al., 2016). *Bacteroides* and *Roseburia* are the major xylanolytic taxa for dietary fibre xylan degradation in the human gut community (Chassard et al., 2007; Mirande et al., 2010; Leth et al., 2018), while *Lachnospirillum* has been previously described as a producer of extracellular cellulases (Ravachol et al., 2015) and hemicellulases (Zhu et al., 2016). In co-cultures or in human intestine, these two taxa have shown a symbiotic relationship for cellulose fermentation (Murray, 1986). Indeed, Murray, (1986)

demonstrated that two species of *Bacteroides* and *Lachnospirillum* (formerly classified as *Clostridium* species (Yutin and Galperin, 2013)) were in symbiotic mutualism relationship in co-culture, in which *B. cellulosolvens* provides sugars and growth factor to *C. saccharolyticum* while *C. saccharolyticum* removes metabolites toxic to *B. cellulosolvens*. This symbiotic relationship improves total cellulose degradation by 33% than did *B. cellulosolvens* alone. Interestingly, *Bacteroides* and *Lachnospirillum* have been reported in the termite-derived TWS consortium displaying higher xylanase activity than in the rumen-derived consortium RWS (Auer et al., submitted – Chapter 2). Moreover, a previous study showed that the hydrolysis of biomass was positively correlated with the abundance of Bacteroidetes and their abundant presence accelerated the hydrolysis of lignocellulose. Here, the alkaline pretreatment fostered the prevalence of xylanolytic bacteria, which could explain the results observed in terms of increasing the rate of substrate hydrolysis and the xylanase activity level.

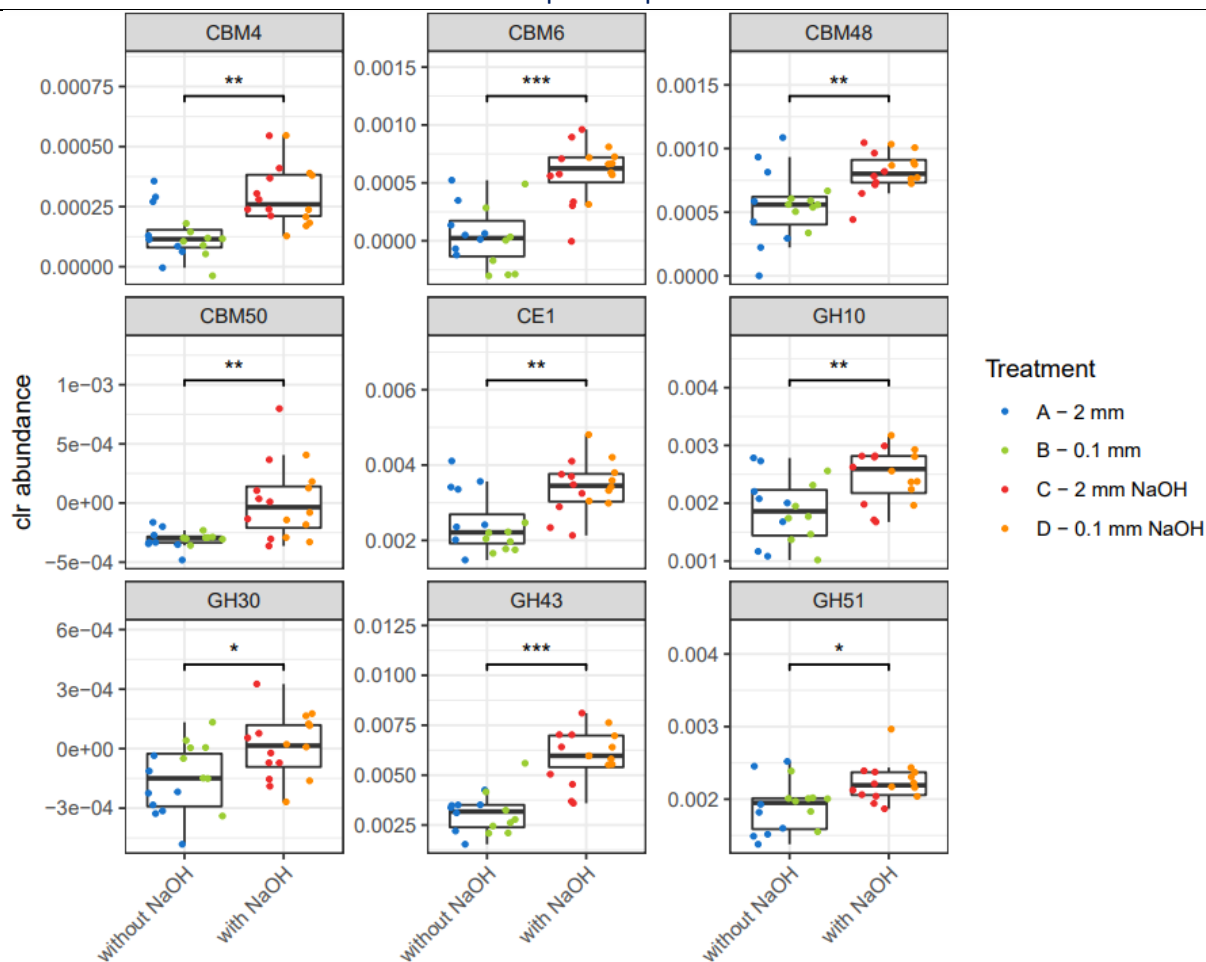
**d) Impact of dry-alkali pretreatment on the expression of carbohydrate-active enzymes throughout lignocellulose degradation**

Protein annotation using the CAZy database (accessed July 2<sup>nd</sup>, 2018) revealed changes in the carbohydrate degradation potential of the RWS consortium when incubated with the four different pretreated substrates. Of the 12,911 proteins identified in RWS, 490 harbored at least one CAZyme domain. The protein profile of RWS growing in the four substrates shared a core of 134 CAZy proteins, constituting 85.1±4.8% of the total CAZyme abundance. 245 proteins, representing 94.3±2.7% of the total CAZyme abundance, were found in at least two substrates and 183 CAZy proteins corresponding to a small fraction of the total CAZyme abundance (1-4%) were only observed in one specific substrate. Metaproteomic analysis revealed the two major contributors to CAZymes in the RWS consortium. For all substrates, *Bacteroides* played the most active role in carbohydrate hydrolysis, contributing 78.6±4.3% of CAZyme abundance (206 proteins) (Additional file 3: Figure S3). *Clostridium* was the second major contributor to CAZymes, expressing 70 proteins that accounted for 6.5±4.6% of the CAZyme abundance. *Bacteroides* and *Clostridium* are both reported as well-known holocellulose degraders in microbial communities due to their high CAZymes diversity and unique strategies for producing them to hydrolyze plant polymers such as cellulose, xylan, pectin, and starch (Blouzard et al., 2010; Munir et al., 2014; Naas et al., 2014; Mackenzie et al., 2015).

#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

NMDS and PERMANOVA analysis of the abundance of the 99 CAZy identified families (CLR-transformed data) confirmed that CAZymes followed similar patterns to those described using the CLR-abundance data of the taxonomic affiliation of proteins. Dry-alkali pretreated substrates showed similar CAZyme profiles and initial sampling points (T1) were well-separated from the others (T2 to T4) (Additional file 3: Figure S4). A further PERMANOVA analysis was performed on these data, considering “NaOH” (substrates pretreated with vs without NaOH), “Particle\_size” (2 mm vs 0.1 mm) and “Degradation\_phase” (Phase A with T1, Phase B with T2, T3 and T4) as explanatory variables. This analysis confirmed that the use of NaOH for substrate pretreatment ( $R^2 = 35.0\%$ ,  $p < 0.001$ ) was the main factor explaining the variations in CAZyme profile between samples, followed by the “Degradation\_phase” ( $R^2 = 14.5\%$ ,  $p < 0.001$ ). In contrast, the effect of “Particle\_size” ( $R^2 = 3\%$ ,  $p > 0.05$ ) was not significant. The interaction term between factors “NaOH” and “Particle size” was minor and barely significant ( $R^2 = 5\%$ ,  $p < 0.05$ ), and probably reflected the small differences between substrates A - 2 mm and B - 0.1 mm.

The effect of dry-alkali pretreatment on the abundance of CAZy families was therefore tested for each CAZy family. After FDR adjustment of the Wilcoxon tests p-values and removal of scarce families (which relative abundances were null in more than 70% of the samples), nine CAZy families showed significant CLR abundances differences between non-alkali pretreated and dry-alkali pretreated substrates (Figure IV-5).



**Figure IV-5 : Comparison of the abundances of the nine CAZy families (CLR-transformed data).** Only CAZy families with significant differences between dry-alkali untreated (A and B) and treated (C and D) substrates are shown. Colored dots represent the values of each sample (8 samples for each substrate). The median, first and third quartiles are represented. Asterisks indicate statistical significance of differences between NaOH pretreated and non-treated substrates (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \*: < 0.05). Wilcoxon adjusted p-value according to the FDR method.

Among them, glycoside hydrolases classified in the GH10 (endo- $\beta$ -1,4-xylanase), GH30 (glucuronarabinoxylan endo- $\beta$ -1,4-xylanase), GH43 (bifunctional  $\beta$ -xylosidase/ $\alpha$ -arabinosidase) and GH51 ( $\alpha$ -L-arabinofuranosidase) families are related to hemicelluloses hydrolysis. This hemicelluloses-targeted arsenal was also completed by the CE1 family which is a very diverse family but mainly composed of acetyl xylan and feruloyl esterases. All these families, related to the degradation of hemicelluloses and particularly arabinoxylan, were overexpressed in response to dry-alkali pretreatment. This is in agreement with the high xylanase activity measured in dry-

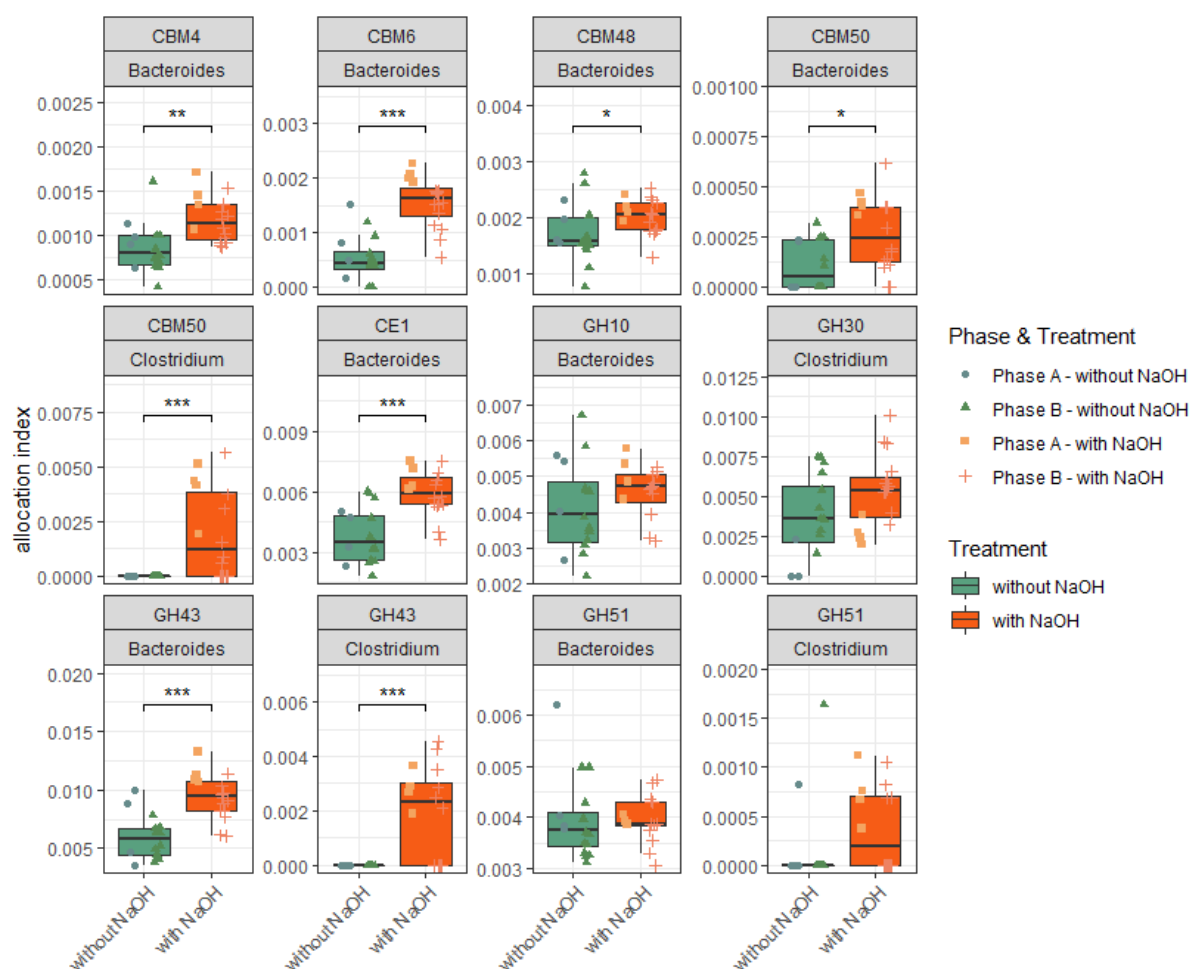
#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

alkali pretreated substrates C and D (Lazuka et al., 2017 and presented in Figure IV-1). Four CBM families (CBM4, 6, 48, 50) were also more abundant in the alkali-pretreated samples and certain were associated to overexpressed GH families. By improving the contact between the enzyme and the substrate, CBMs are known to potentiate the activity of associated catalytic enzymes (Munir and Levin, 2016). CBM4 and CBM6 are known to bind amorphous cellulose and xylan (McCartney et al., 2006) and they are contained in multimodular proteins in this dataset with respectively GH10 (95.4±10.2% of CBM4 relative abundance in the dataset) and GH43 (83.0±8.8 of CBM6 relative abundance in the dataset) domains. CBM48 is usually associated with starch binding (Janeček et al., 2011). In this study, CBM48 is frequently cognate to GH13 domains (starch debranching enzymes; 53.3±13.9% of CBM48 relative abundance in the dataset) but also to CE1 domains (46.7±13.9 of CBM48 relative abundance in the dataset). A recent study postulated that CE1 feruloyl esterases with appended CBM48 domains would unexpectedly act on arabinoxylans (Holck et al., 2019). CBM50 was found in a wide diversity of enzymes, with mainly peptidoglycanase, chitinase, peptidase and amidase activities. The overexpression of CBM50 in the alkali-pretreated substrates remains to be explained.

The taxonomic affiliation of these nine CAZy families enabled to identify the genera contributing the most to their expression (Additional file 3: Figure S5). For each treatment, the average abundance of the genus-specific CAZy families showed that all of them except GH30 and CBM50, were affiliated to *Bacteroides*, which was the major and sometimes the unique contributor. CBM50 showed more variation, being related to *Bacteroides*, *Clostridium* and *Lachnoclostridium*, while GH30 was expressed almost exclusively by *Clostridium*. Since most of the proteins of the dataset were affiliated to *Bacteroides*, the major affiliation of CAZy families to this genus was not surprising. However, although this genus accounted for 60-70% of total protein abundance in the dataset, more than 95% of the identified CAZy families were affiliated to *Bacteroides*. This reflects the specialization and contribution of this genus to the expression of CAZymes. Nevertheless, biological reality is certainly more contrasted as the dominance of *Bacteroides* is probably accentuated due to metaproteomics analysis depth effects. Indeed, several technical and methodological aspects of metaproteomics involve an overrepresentation of abundant taxa and a lower detectability of minor ones (Kleiner et al., 2017), especially with cellular functions counting for low amounts of total proteins. Because the main contributors could also themselves differentially abundant between dry-alkali untreated and treated substrates, the increase of these CAZy families could be due to a) the increase in the abundance of the contributing genus without

Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d’un consortium microbien face aux modifications du substrat induites par des prétraitements

any changes in their expression, or b) an overexpression of the corresponding genes in response to dry-alkali pretreatment, with or without variation of the contributing genus. To answer this question, an “allocation index” (AI) was calculated by normalizing the abundance of CAZy families affiliated to a genus by the total protein abundance of that genus. As *Bacteroides* and *Clostridium* were the main contributors to the CAZy families differentially expressed between substrates, the AI of these genera were compared between alkali-treated and non-treated substrates (Figure IV-6).



**Figure IV-6 : Comparison of the allocation index (AI) of *Bacteroides* and *Clostridium* between alkali-treated and non-treated substrates for the nine CAZy families with significant abundance differences.** Samples from the degradation phases A and B are represented with different shapes (A : T1 and B : T2 to T4 points). Wilcoxon tests and FDR adjustment were performed to compare all samples from NaOH treated and non-treated substrates (without considering the degradation phase). Statistical differences between substrates pretreated with NaOH or not are indicated with asterisks (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \* : < 0.05).

Several *Bacteroides* CAZy proteins displayed significantly different allocation indexes between dry-alkali substrates conditions (CBM4, CBM6, CBM48, CBM50, CE1 and GH43). *Clostridium* displayed a significantly different allocations indexes for CBM50 and GH43 only. Remarkably, *Clostridium* GH30 reached almost 1% AI, meaning that almost 1% of total *Clostridium* related proteins were GH30. In the same way, *Bacteroides* GH43 reached 1.3% AI and CE1 reached 0.8% on alkali treated substrates, so at least 2% of *Bacteroides* related proteins were directed toward arabinoxylans, regulated in response to dry-alkali pretreatment (without counting the other *Bacteroides* CAZymes families which described hemicelluloses-degrading activities), which could be linked to the higher measured xylanase activity (Figure IV-1).

All of these elements suggest that one of the main effect of dry-alkali wheat straw pretreatment was a metabolic shift lead to an increase of the expression of hemicelluloses-degrading enzymes of the dominant genus *Bacteroides*. Specific regulation of PULs that encode esterases and hemicellulolytic enzymes for degradation of wheat arabinoxylans in *Bacteroides* species in response to the alkali pretreatment could be suggested since the main effect of this pretreatment is to promote depolymerization by cleaving hydrolysable linkages between lignin and holocellulose fibers. As polysaccharide and their breakdown products enables the regulation of PULs in *Bacteroides* species (Schwalm et al., 2016; Pereira et al., 2021), the genes expression in PULs is rapidly induced from the first days of kinetics, once the wheat arabinoxylan is more accessible due to the alkaline pretreatment. This hypothesis could also explain the increase of bioconversion rate, without an improvement of the yield. Unfortunately, the xylanase enzyme assays, performed to characterize RWS response to dry-alkali pretreatment, used xylan as substrate, and did not allow to identify precisely the enzyme activities regulated by the substrate changes induced by pretreatments. Nevertheless, our results and previous works collectively showed that the combination of dry alkali pretreatment and hemicellulolytic enzymes, particularly feruloyl esterases and xylanases (such as GH10, GH30, GH43, GH51, CE1) and their appended binding modules (CBM4, CBM6, CBM48) probably facilitated the access of endoglucanases, exoglucanases and  $\beta$ -glucosidases to cellulose fibers, and improved lignocellulosic biomass hydrolysis (Braga et al., 2014; Goldbeck et al., 2016; Mkabayi et al., 2020; Agrawal et al., 2021; Mafa et al., 2021).



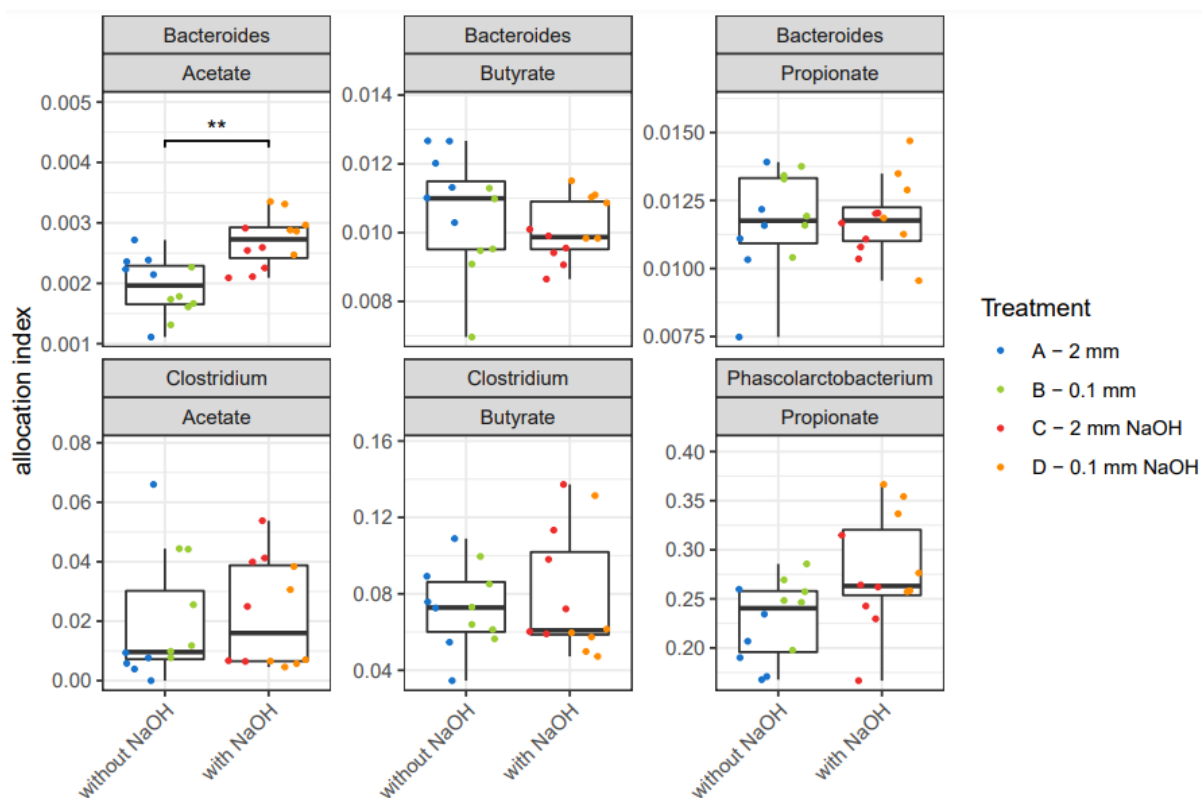
**e) Dry-alkali pretreatment influences the expression of enzymes related to VFA production by RWS**

Carbohydrates hydrolysis is one aspect of lignocellulose bioconversion in the carboxylate platform, the second aspect is the volatile fatty acids production. Previous study on the response of RWS to substrate pretreatment by dry-alkali showed an increase in the butyrate/acetate ratio. Authors suggested that the metabolism of *Clostridium* has probably shifted from acetate production using hexoses to butyrate production based on xyloses consumption (Lazuka et al., 2017).

The occurrence of key enzymes related to VFA biosynthesis (acetate, butyrate and propionate) was investigated on the basis of previous COG assignments (Polansky et al., 2016; Deusch et al., 2017; Tilocca et al., 2017). 440 proteins belonging to 130 genera and 19 COGs related to VFA biosynthesis were identified in the metaproteomic dataset, reflecting good detectability and high diversity of VFA-related proteins, even in low abundance taxa. T1 points samples of each treatment conditions were removed from the statistical analyses to better determine the effect of pretreatment on the proteins related to VFA production at later stage. Comparison of VFA-related protein abundance (CLR-transformed data) between dry-alkali treated and untreated samples, showed a significantly lower abundance of acetate-related proteins and a higher level of butyrate-related proteins in the alkali-treated samples (Additional file 3: Figure S6). This is in agreement with the changes of VFA production described for RWS (Lazuka et al., 2017). The main RWS contributors to VFA-related proteins expression were distributed between *Bacteroides*, *Clostridium*, *Ruminococcus* and *Phascolarctobacterium* (Additional file 3: Figure S6). Although *Bacteroides* was the main contributor to total protein abundance, only 30-40% of VFA-related proteins were produced by this genus, at a level equal to that of *Clostridium* which was much less abundant. *Ruminococcus* were quite rare, and *Phascolarctobacterium*, which had a low contribution to total protein abundance, showed a high contribution (>20%) to propionate biosynthesis-related proteins.

Allocation indexes for VFA production were thus investigated for *Bacteroides*, *Clostridium* and *Phascolarctobacterium* (Figure IV-7).





**Figure IV-7 : Comparison of Allocation Index (AI) between treatments abundances of the proteins related to VFA production.** Wilcoxon tests and FDR adjustment were performed at NaOH level without considering treatments, which were colored for visualization but not tested. Statistically significant differences between without NaOH and with NaOH treatments are indicated with asterisks (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \*: < 0.05). For butyrate biosynthesis: COG4770 (acetyl/propionyl-CoA carboxylase), COG3426 (butyrate kinase), COG1250 (3-hydroxyacyl-CoA dehydrogenase), COG1024 (enoyl-CoA hydratase), COG0183 (acetyl/butyryl-CoA acetyltransferase), COG1979 (alcohol dehydrogenase YqhD), COG1454 (alcohol dehydrogenase, class IV), COG1028 (short-chain alcohol dehydrogenase), COG1064 (Zn-dependent alcohol dehydrogenase). For propionate biosynthesis: COG4799/0777 (acetyl/propionyl-CoA carboxylase), COG2185/COG1884 (methylmalonyl-CoA mutase), COG0346 (methylmalonyl-CoA epimerase). For acetate production: COG1012 (NAD-dependent aldehyde dehydrogenase), COG0282 (acetate kinase), COG0280 (phosphotransacetylase).

*Bacteroides* was the major contributor of xylanase activity but to a lesser extent for VFA production, with a relatively low allocation index for VFA production (0.3% acetate AI, 1% butyrate AI) compared to *Clostridium* (1 to 2% acetate AI, 6 to 12% butyrate AI). *Clostridium* appeared as a very active producer of VFA-related proteins. *Phascolarctobacterium* is a well-known propionate producer (Watanabe et al., 2012) and displayed high propionate AI (from 20 to 35% of its identified

proteins were propionate-related proteins), but this level was probably overestimated because of the lower detectability of its proteins due to its low total abundance. No significant differences were observed except for acetate-related proteins expressed by *Bacteroides* (with a small increase from 0.2% to 0.28% AI), so the main differences are due to genera abundances effects and not to expression regulation, especially for *Clostridium* which did not show any significant AI modifications for acetate and butyrate.

Metabolic switch from acetate to butyrate-producers (such as *Clostridium*, *Bacteroides*, *Ruminococcus*, *Roseburia*, *Lachnoclostridium* and others) hypothesis still remained possible. Indeed, as hemicelluloses-degrading enzymes is stimulated by alkaline pretreatment and particularly those degrading arabinoxylan, it is known that this enzyme combination release significant amount of xylo-oligosaccharides and arabinoxylan-oligosaccharides end-products (pentoses) which can provide substrates for butyrate-producers and increase the butyrate levels (Ravn et al., 2018). Also, addition of the xylanolytic bacteria (*Roseburia sp.*) in the monoculture of the acetogenic bacteria (*Ruminococcus sp.*) culture led to reduction of acetate concentration while butyrate concentration increased (Chassard and Bernalier-Donadille, 2006). Further metatranscriptomics or metabolomics analysis on acetate and butyrate synthesis pathways could confirm the hypothesis that members of RWS community adjusted their carbohydrate metabolism and took the advantages of the set of arabinoxylan-degrading enzymes expressed by xylanolytic species and induced by alkaline pretreatment.

## 5. Conclusions

Biomass pretreatments benefits had been studied for years, but their effects on the lignocellulolytic microbial communities were rarely described, and were scarcely ever reported at metaproteomics scale. Dry-alkali pretreatment, hereby presented, was previously described to have positive effects on bioconversion rate of RWS microbial consortium. Metaproteomics analysis allowed to precise its effect on RWS bacteria: among the fifteen most abundant genera, four of them had their abundance positively impacted by alkali treatment, particularly the most dominant xylanolytic genus, *Bacteroides*. Dry-alkali pretreatment had a significant impact on the CAZymes expression, with nine families overexpressed, which most of them are directed towards arabinoxylan polymer and associated oligosaccharides. *Bacteroides* and *Clostridium* were the major producers of these CAZymes and showed evidences for an active regulation of their CAZyme repertory, some families being overexpressed with higher allocation indexes when cultivated on

#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

---

dry-alkali pretreated substrates. Proteins involved in VFA production were also impacted by substrate changes induced by dry-alkali pretreatment as the metabolic shift from acetate formation to butyrate is also observed at metaproteomic scale but probably result from a better availability of pentoses resulting from the degradation of hemicelluloses. To conclude, this metaproteomics analysis suggest interplays between RWS community members as alkaline pretreatment induced CAZymes expression of xylanolytic species which target arabinoxylan and their hydrolysis products were converted into butyrate by butyrate producers.

**6. Additional files**

***Additional file 1: Supplementary Table S1***

**Supplementary Table S1. Reproducibility of label-free metaproteomics quantification of technical replicates of the bioreactor samples estimated by Pearson correlation. A2-T1\_C and D1-T2\_B samples were removed from further analysis.**

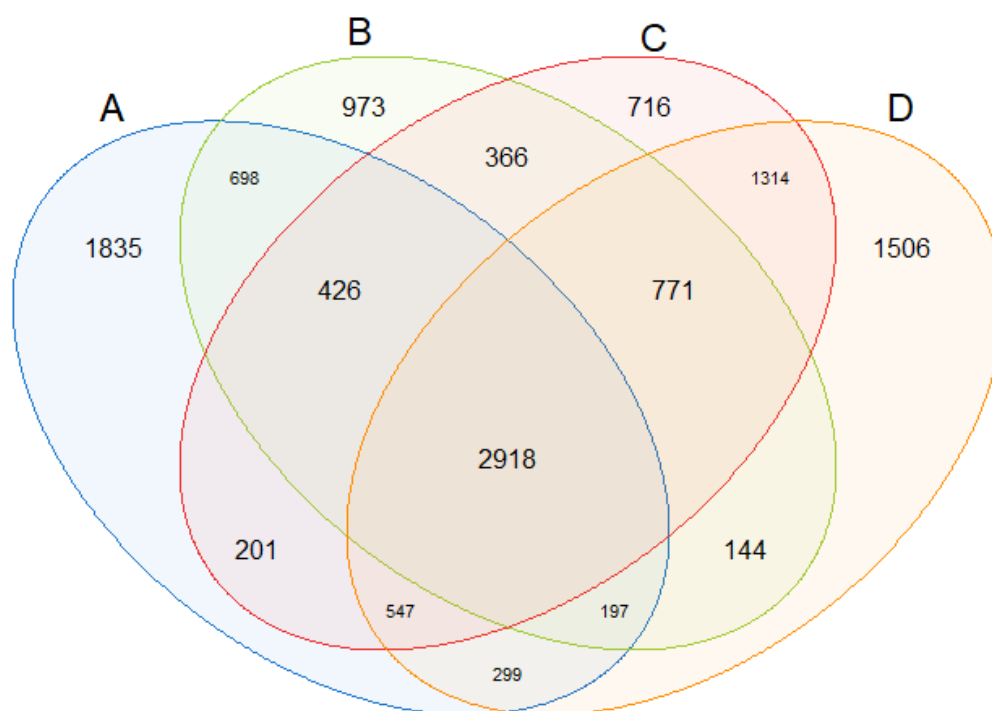
Available data upon request.

***Additional file 2: Supplementary Data 1: Metaproteomics data***

Available data upon request.

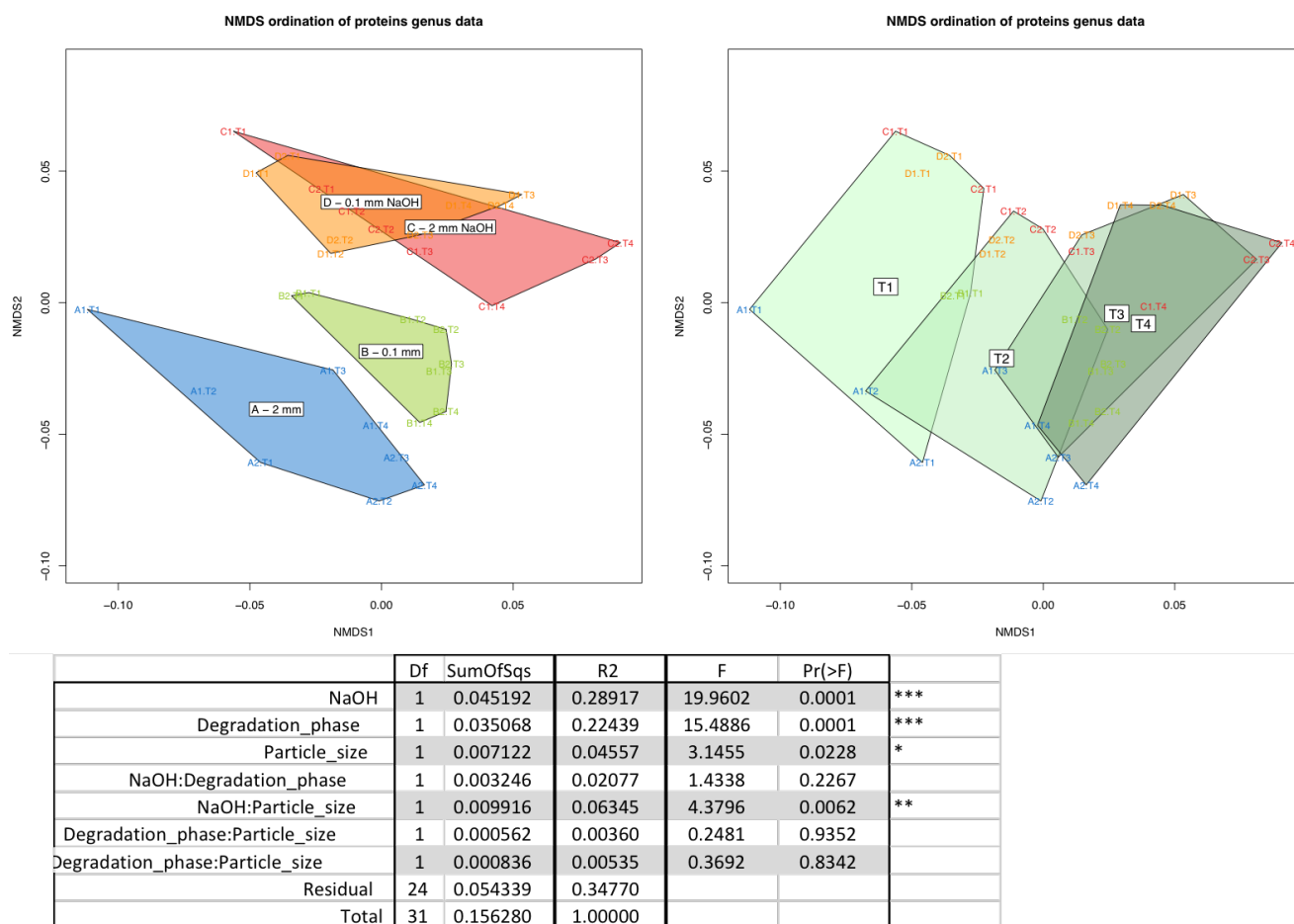
***Additional file 3: Supplementary Figures S1 to S7***

(1) Figure S1



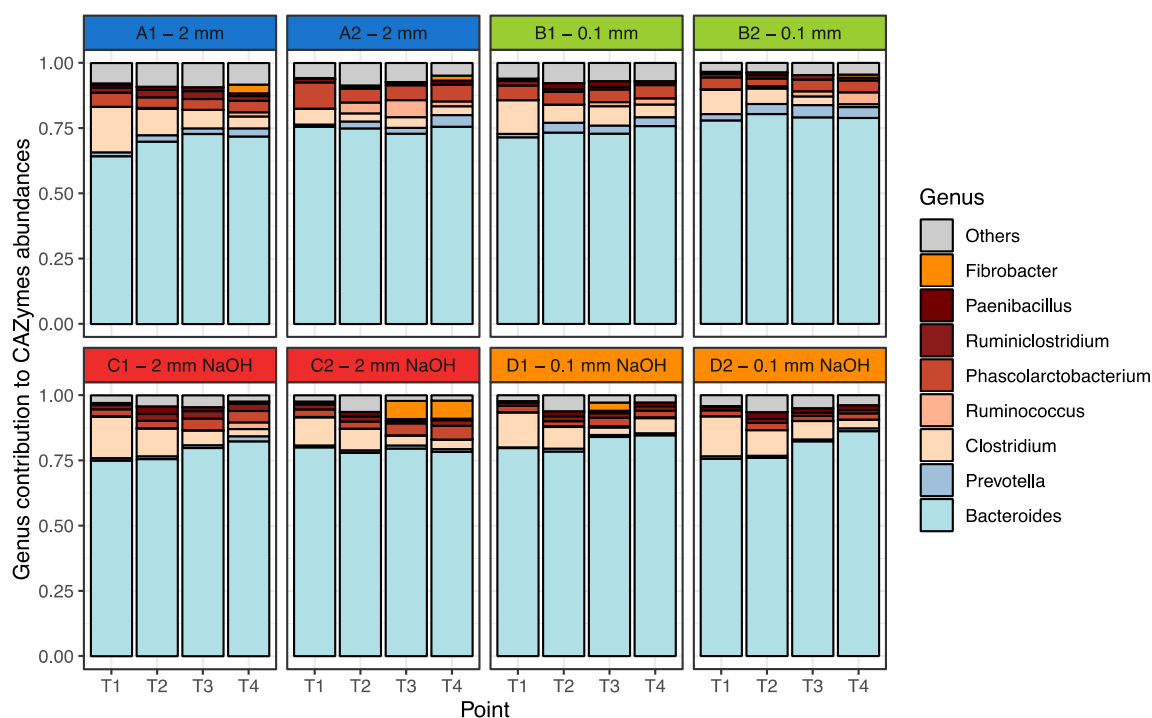
**Figure IV-S1 : Venn diagram showing the overlap in the RWS expressed proteins in the four substrate pretreatments (blue, A - 2 mm; green, B - 0.1 mm; red, C - 2 mm NaOH; orange, D -0.1 mm NaOH).**

(2) Figure S2



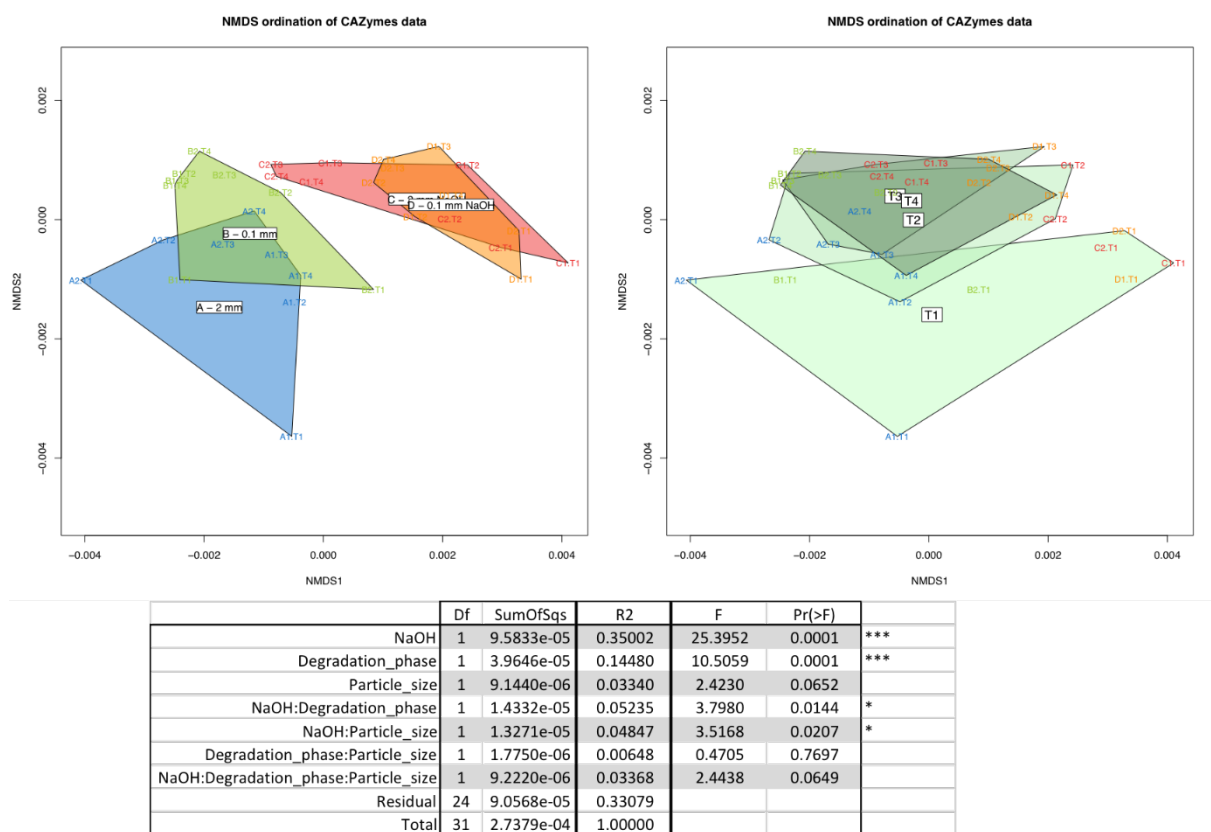
**Figure IV-S2 : Non-Metric Multi-Dimensional Scale (NMDS) ordination and Permutation Multivariate Analysis of Variance (PERMANOVA) analysis based on abundance of proteins affiliated at the genus level (CLR-transformed data).** NMDS ordination based on Euclidean distances. Sample codes indicate the substrate pretreatment (A to D), the biological replicate bioreactor (1 or 2), and the sampling time (T1 to T4). Polygons joining the samples were drawn and colored according to substrate pretreatment (left plot: A – 2 mm in blue, B – 0.1 mm in green, C – 2 mm NaOH in red and D – 0.1 mm NaOH in orange) and time points (right plot: with T1, T2, T3 and T4 in shades of green). PERMANOVA analysis was performed using “NaOH” (alkaline versus non-alkaline), “Particle\_size” (2 mm vs 0.1 mm) and “Degradation\_phase” (T1 vs T2-T3-T4). R<sup>2</sup> reflects the effect strength of each variable, and Pr, its significance (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \*: < 0.05). Df and SumOfSquares indicate the degrees of freedom and sum of squares, respectively.

(3) Figure S3



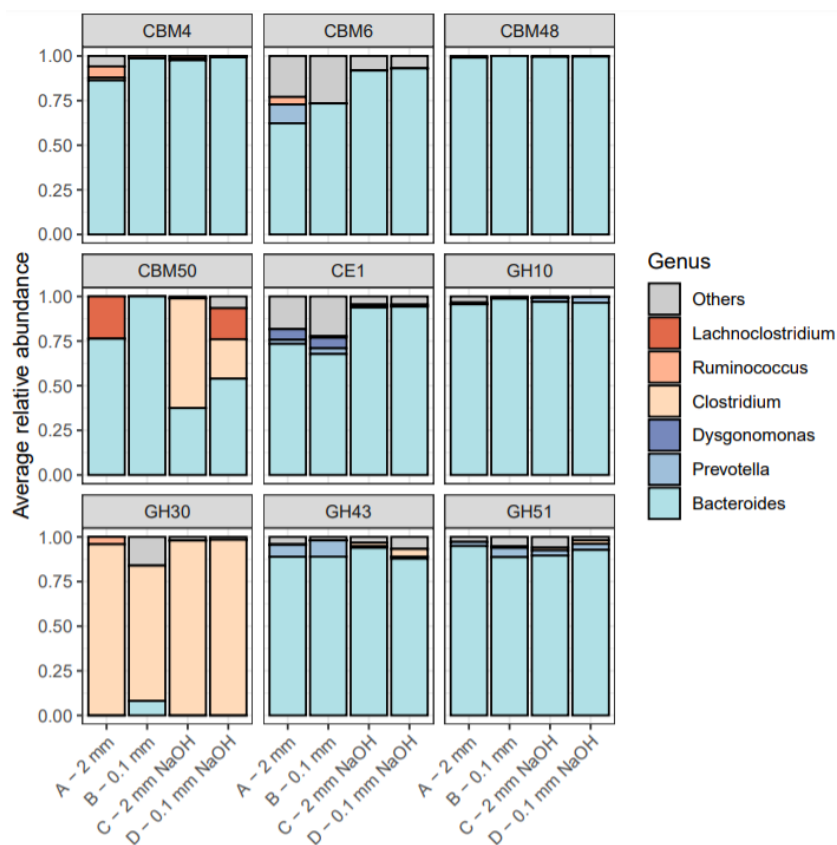
**Figure IV-S3 : Relative abundance of CAZymes according to their taxonomic affiliation.** Time points (T1 to T4) for each biological replicate bioreactor (numbers 1 and 2) and pretreated substrate (A to D) are represented. For stacked bar plot representation, relative abundance of CAZymes, based on NSAFs values (Normalized Spectral Abundance Factors), were TSS normalized and aggregated at the genus level. Genera representing less than 2% of the CAZyme abundances were regrouped in “Others” category. CAZymes belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Fibrobacteres (orange).

(4) Figure S4



**Figure IV-S4 : NMDS and PERMANOVA analysis based on the abundance of CAZymes aggregated according to their taxonomic affiliation (CLR-transformed data).** NMDS ordinations were based on Euclidean distances. Sample codes indicate the substrate pretreatment (A to D), the biological replicate bioreactor (1 or 2), and the sampling time (T1 to T4). Polygons joining the samples were drawn and colored according to substrate pretreatment (left plot: A – 2 mm in blue, B – 0.1 mm in green, C – 2 mm NaOH in red and D – 0.1 mm NaOH in orange) and time points (right plot: with T1, T2, T3 and T4 in shades of green). PERMANOVA analysis was performed using “NaOH” (alkaline versus non-alkaline), “Particle\_size” (2 mm vs 0.1 mm) and “Degradation\_phase” (T1 vs T2-T3-T4). R<sup>2</sup> reflects the effect strength of each variable, and Pr, its significance (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \*: < 0.05). Df and SumOfSquares indicate the degrees of freedom and sum of squares, respectively.

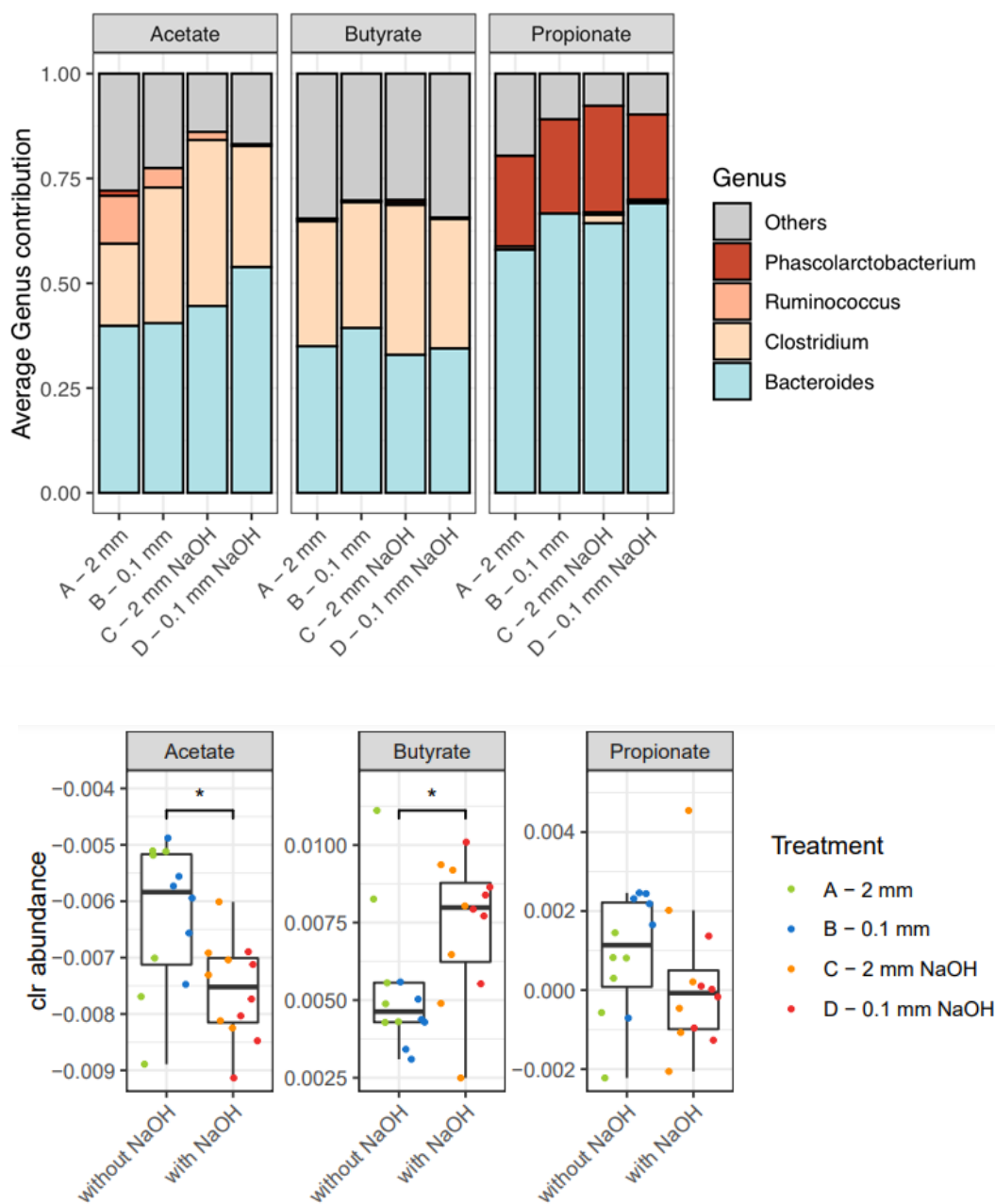
(5) Figure S5



**Figure IV-S5 : Taxonomic contribution to CAZy families at the genus level. CAZy families are the nine families overexpressed in dry-alkali pretreated substrates.** Relative abundance of CAZy families based on NSAFs values (Normalized Spectral Abundance Factors) were TSS normalized and aggregated at the genus level for stacked bar plot representation. Genera with less than 2% contribution (in average relative abundance) in all samples were grouped in “Others” category. CAZymes belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red).



(6) Figure S6



**Figure IV-S6 : Top graph : Genus contribution to VFA-related protein abundance per VFA molecule (acetate, butyrate, propionate) in the four pretreated substrates.** Genera contributing with less than 5% to any VFA molecule were grouped in “Others” categories. Taxa affiliation was colored according to the phylum and genus code (Bacteroidetes in blue, Firmicutes in red). **Bottom graph: Comparisons of VFA-related protein abundance (CLR-transformed data), for each VFA, between alkali-treated and untreated substrates.** Values of individual samples were plotted as treatment-colored dots. Statistically significant differences between NaOH-treated and untreated substrates are indicated with asterisks (\*\*\*: < 0.001 ; \*\*: < 0.01 ; \*: < 0.05). For butyrate biosynthesis: COG4770 (acetyl/propionyl-CoA carboxylase), COG3426 (butyrate kinase), COG1250 (3-hydroxyacyl-CoA dehydrogenase), COG1024 (enoyl-CoA hydratase), COG0183 (acetyl/butyryl-CoA acetyltransferase), COG1979

#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

---

(alcohol dehydrogenase YqhD), COG1454 (alcohol dehydrogenase, class IV), COG1028 (short-chain alcohol dehydrogenase), COG1064 (Zn-dependent alcohol dehydrogenase). For propionate biosynthesis: COG4799/0777 (acetyl/propionyl-CoA carboxylase), COG2185/COG1884 (methylmalonyl-CoA mutase), COG0346 (methylmalonyl-CoA epimerase). For acetate production: COG1012 (NAD-dependent aldehyde dehydrogenase), COG0282 (acetate kinase), COG0280 (phosphotransacetylase).

## **7. Funding**

This work was funded by the French National Research Agency (ANR) under the Project Hi-Solids ANR-14-CE19-0013-01, the French National Agency for Energy and the Environment (ADEME), the TRANSFORM department of INRAE. This study was also supported by the PHC Procope program from Campus France under the grant N° 3077WC.

## **8. Author contributions**

E.A carried out and analyzed metaproteomics data under G.H-R supervision. Adele Lazuka realized the bioreactor experiments and Maïder Abadie prepared the samples for metaproteomics analysis under the supervision of G.H-R and N.J. N.J. performed and analyzed LC-MS/MS experiments. B.H and V.L annotated the CAZymes proteins. E.A performed the statistical data analysis. As project coordinator, G.H-R designed the study, participated on experimental design and contributed at all stages. The manuscript was written by E.A. under G.H-R supervision. G.H-R and N.J participated in funding this research.

### C. Conclusion

En s'appuyant sur la méthodologie établie dans le chapitre précédent, l'analyse métabolomique comparative d'un consortium bactérien lignocellulolytique en bioréacteur a une nouvelle fois démontré son fort potentiel. Tout en fournissant des données sur l'abondance des protéines exprimées par les différents taxa (d'une qualité équivalente au séquençage de l'ADNr 16S), la métabolomique a permis de déterminer les familles CAZymes différentiellement exprimées dans chaque prétraitement, particulièrement par le prétraitement chimique, et d'identifier aussi les taxons producteurs des CAZymes impactées. Nous avons également observé l'impact du prétraitement alcalin sur les voies de production des acides gras volatils. L'utilisation d'un « indice d'allocation », normalisant pour un taxon l'abondance d'une fonction par le total de l'abondance des protéines du taxon, a permis d'étudier plus finement l'impact des prétraitements sur le fonctionnement d'un écosystème lignocellulolytique, et notamment d'étudier individuellement l'activité des taxons identifiés comme taxons d'intérêt. Cet indice, reflétant la part des protéines d'un taxon allouée à une fonction, permet de séparer les effets « de masse » (effet sur la croissance d'un taxon et donc effet indirect sur l'abondance de ses protéines, qui augmentent toutes) des effets de « régulation » (effet sur le niveau d'expression d'une fonction protéique indépendamment des autres protéines du taxon étudié). Ainsi, l'indice d'allocation a permis d'identifier plusieurs familles CAZymes, impliquées dans l'hydrolyse des arabinoxylans exprimées par des *Bacteroides*, comme spécifiquement régulées en réponse au prétraitement alcalin. Bien que le prétraitement alcalin en voie sèche ne semble pas avoir modifié la composition du substrat (% holocellulose et lignine similaires), il semble avoir significativement impacté la cristallinité de la cellulose ainsi que la porosité du substrat (Lazuka et al., 2017). Aussi, les effets majeurs du prétraitement alcalin, cités dans la littérature, sont la décristallisation de la cellulose mais aussi l'élimination de la lignine améliorant l'accès à l'holocellulose, et particulièrement aux hémicelluloses. En effet, les prétraitements alcalins éliminent les groupements acétyles et les diverses ramifications d'acides uroniques des hémicelluloses, qui réduisent l'accessibilité des enzymes à la surface des hémicelluloses et de la cellulose (Chang and Holtzapple, 2000). De plus, il est rapporté que le prétraitement alcalin saponifie les liaisons ester liant les hémicelluloses et la lignine, les rendant encore davantage accessibles aux enzymes (Sun and Cheng, 2002). Ainsi, au vue du mécanisme de régulation de l'expression des CAZymes, il apparaît évident qu'une meilleure accessibilité des hémicelluloses régule positivement l'expression des CAZymes impliquées dans l'hydrolyse des hémicelluloses (Martens et al., 2009), et particulièrement celles impliquées dans

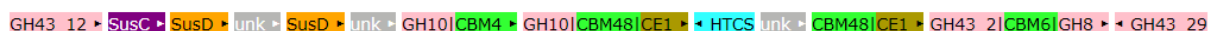
l'hydrolyse des arabinoxylans, fortement présents dans les hémicelluloses du blé. Des analyses chimiques plus détaillées telles que la mesure du taux d'acétylation et d'estérification de la biomasse, pourraient nous permettre d'expliquer les effets observés et valider nos hypothèses.

Concernant les voies de production des acides gras volatils, il semble que le prétraitement alcalin favorise les voies de production du butyrate au détriment de l'acétate. Néanmoins, l'indice d'allocation indique qu'il n'y a pas d'effet de régulation chez les *Clostridium*, genre producteur de butyrate majoritairement connu et présent dans le consortium RWS. Ce shift métabolique serait probablement une conséquence directe du prétraitement alcalin et de l'induction de la machinerie enzymatique dédiée à l'hydrolyse des arabinoxylanes chez les *Bacteroides* libérant une large quantité de pentoses (xylose, arabinose). Ces derniers sont fermentés en butyrate par les genres producteurs de butyrate connus et présents dans le consortium microbien RWS comme *Clostridium*, *Bacteroides*, *Ruminococcus*, *Lachnoclostridium*, *Roseburia* entre autres. L'ensemble des résultats concordent donc avec les observations macrocinétiques collectées au cours de la fermentation (*i.e* augmentations du ratio butyrate/acétate, de l'activité xylanase et de la vitesse de dégradation du substrat).

L'hydrolyse de la lignocellulose joue un rôle limitant dans la digestion anaérobie et peut avoir un fort impact sur les taux de production et le profil des acides gras volatils selon le type de monomères de sucres libérés (pentoses ou hexoses). Les prétraitements sont un des principaux leviers permettant de faciliter l'hydrolyse de la biomasse lignocellulosique et leurs effets sur le substrat peuvent avoir des conséquences directes sur l'induction des gènes responsables de la machinerie enzymatique dédiée à l'hydrolyse du substrat. Face aux hypothèses évoquées dans cette étude, il serait intéressant d'approfondir l'impact du prétraitement sur les CAZymes, et en particulier les PULs xylanolytiques des *Bacteroides*, et spécifiquement ceux de *Bacteroides graminisolvens*, espèce majoritaire dans le consortium RWS (30-60%). Récemment, le développement d'un outil de prédiction automatique des PULs a permis la création de la « Polysaccharide-Utilization Loci DataBase » par l'équipe du Dr Bernard Henrissat (PULDB, <http://www.cazy.org/PULDB/index.php>; (Terrapon et al., 2018)). Dans notre étude, *B. graminisolvens* est l'espèce retrouvée comme la plus abondante à la fois dans les analyses de diversité 16S et de métabolomique. Au sein de la PULDB, 34 PULs ont été prédits pour *B. graminisolvens*. Parmi eux, les PUL20 et PUL23 semblent intéressants puisqu'ils codent les familles CAZymes, que nous avons précédemment décrit comme régulées par le prétraitement alcalin. Le

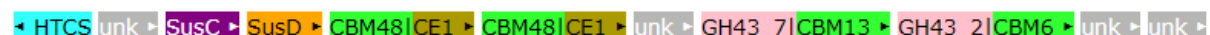
#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

PUL20 est un cluster de 13 gènes codant les enzymes impliquées dans la dégradation des arabinoxylyanes (les glycosides hydrolases des familles GH10, GH43, GH8 et la famille des carbohydres esterases CE1) et les gènes *susC/D* codant pour des protéines de membrane extracellulaires impliquées dans l'internalisation des produits d'hydrolyses des arabinoxylyanes dans la cellule. Des modules impliqués dans la liaison des carbohydres sont également présents : un domaine CBM4 associé au domaine GH10, deux domaines CBM48 associés à deux domaines CE1 et un domaine CBM6 associé à un domaine GH43 et à un domaine GH8. Un gène régulateur (*HTCS*) est présent dans ce PUL xylanolytique et trois gènes (*unk*) sont pour le moment inconnus.



**Figure IV-8 : Représentation du PUL20 de *Bacteroides graminisolvens* identifié et référencé dans la PULDB (<http://www.cazy.org/PULDB/index.php> , dernièrement consulté le 24 septembre 2021).**

De la même façon, le PUL23 est un cluster intéressant de 11 gènes qui code également certaines enzymes impliquées dans la dégradation des arabinoxylyanes (GH43, CE1) et les gènes *susC/D*. Des modules impliqués dans la liaison des carbohydres sont également présents : un domaine CBM6 associé à un domaine GH43, un domaine CBM13 associé à un domaine GH43 et deux domaines CBM48 associés à deux domaines CE1. Un gène régulateur (*HTCS*) est également présent dans ce PUL xylanolytique et quatre gènes (*unk*) restent inconnus.



**Figure IV-9 : Représentation du PUL23 de *Bacteroides graminisolvens* identifié et référencé dans la PULDB (<http://www.cazy.org/PULDB/index.php> , dernièrement consulté le 24 septembre 2021).**

Aucune étude n'a à ce jour permis de montrer à l'échelle d'un consortium microbien l'impact potentiel des changements de substrat induits par les prétraitements sur la dynamique de bioconversion de la lignocellulose (dégradation de la biomasse et production d'AGV), la régulation des CAZymes et particulièrement des PULs chez les *Bacteroides* ainsi que les conséquences pour les producteurs des acides gras volatils. Seules des études ont démontré de façon indépendante qu'un cocktail xylanolytique composé d'acétyl feruloyl esterases, de xylanases et d'enzymes

#### Chapitre 4 – Analyse métabolomique de la réponse fonctionnelle d'un consortium microbien face aux modifications du substrat induites par des prétraitements

---

accessoires (xylosidases ou arabinofuranosidases) améliorait efficacement l'hydrolyse des arabinoxylanes entraînant une libération conséquente de pentoses et une augmentation de la production de butyrate à partir de ces derniers. Cette étude de métabolomique montre ses limites pour déterminer les mécanismes sous-jacents mais des analyses complémentaires en métatranscriptomique et en métabolomique permettraient d'étudier finement les conséquences du prétraitement alcalin à l'échelle des voies métaboliques et valider ces hypothèses. Néanmoins, une nouvelle piste s'ouvre donc pour piloter les consortia microbiens et optimiser la bioconversion de la lignocellulose en produits d'intérêt à haute valeur ajoutée, en complément des prétraitements.

Bien que l'analyse métabolomique ait montré une nouvelle fois son intérêt pour l'étude et la compréhension des communautés microbiennes lignocellulolytiques, certains choix de méthodologie peuvent être discutés. La durée des analyses (temps d'extraction et analyse des résultats) nous ont conduits à nous limiter à l'analyse d'une centaine d'échantillons et le choix des points était donc crucial. Nous avons fait le choix d'analyser des échantillons selon le niveau de dégradation du substrat et d'activité xylanase, et le taux de production d'AGV. Ce choix permettait de comparer le métabolisme du consortium sans être biaisé par la vitesse globale de dégradation, qui est différente pour chaque substrat. Par contre, il nous a conduit à analyser un plus faible nombre d'échantillons et cela a limité la puissance des tests statistiques. Pour pallier à cette faiblesse, nous avons fait le choix de nous concentrer sur les différences entre traitement alcalin (C et D) et non-alcalin (A et B), en écartant les différences du prétraitement mécanique seul (broyage à 2 mm et 0.1 mm), qui montrait des effets plus faibles. Avec des analyses complémentaires, il serait néanmoins possible d'étudier plus en détail ces différences (A vs B et C vs D).







## V. Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

### A. Introduction

La suite de ces travaux de thèse s'intègre dans le projet ANR Hi-Solids (<https://anr.fr/Projet-ANR-14-CE19-0013>), projet proposant d'appliquer une approche intégrative à l'étude d'un procédé de bioraffinerie en voie sèche, en combinant le prétraitement de la biomasse lignocellulosique par extrusion-réactive et l'action de consortia microbiens afin de produire des carboxylates à fortes teneurs en solides (>15% TS), avec pour substrat modèle des résidus de maïs.

L'objectif premier de ce projet fut d'obtenir un consortium microbien stable, et présentant de bonnes capacités de bioconversion du substrat en acides gras volatils (AGV) à faible teneur en solides (2% TS) afin de déterminer les paramètres optimaux du procédé d'extrusion-réactive pour la fermentation. C'est dans ce contexte que dans ce cinquième chapitre, un nouvel inoculum issu du rumen bovin (différent des chapitres 3 et 4) a été enrichi à une faible teneur de résidus de maïs (2% TS). L'enrichissement a été réalisé en bioréacteurs séquentiels discontinus (*Sequencing Batch Reactor*, SBR), en dupliquat biologique, méthode déjà éprouvée au laboratoire pour l'obtention d'un consortium microbien stable à partir de rumen de vache et à faible teneur en résidus de paille de blé (2% TS) (Lazuka et al., 2015). En tenant compte des enseignements acquis lors des précédents enrichissements réalisés au laboratoire, et pour étudier la dynamique de la communauté microbienne au cours des cycles d'enrichissement (seul le dernier temps du cycle final avait été étudié auparavant), des prélèvements pour séquençage du gène codant pour l'ARNr 16S ont été planifiés. Un seul des deux consortia microbiens obtenus lors de cet enrichissement a été sélectionné pour sa meilleure production d'AGV et a été conservé à -80°C sans cryo-conservateur. Des essais de stabilité après cryo-conservation ont donc été réalisés afin de valider la robustesse du consortium obtenu pour son utilisation au long terme pour les expérimentations

Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

---

du projet. Un essai de mise à l'échelle (*scaling-up*) de culture a également été accompli avec succès en bioréacteur de 30 L.

Enfin, la dynamique de bioconversion du consortium obtenu a été caractérisée plus finement par des prélèvements quotidiens dans les phases « actives » de dégradation de la biomasse (9<sup>ers</sup> jours) puis tous les 3 jours jusqu'au 21<sup>ème</sup> jour. Nous avons procédé à l'intégration des données de diversité microbienne 16S et de production des AGV par analyse longitudinale pour détecter d'éventuels liens entre les taxons et la production des AGV. De plus, afin de comprendre le rôle de la biomasse microbienne attachée au substrat, nous avons entrepris l'analyse de la diversité 16S et de l'activité xylanase sur des échantillons fractionnés (culot / surnageant). Cela nous a permis de déterminer les populations microbiennes corrélées à la production des AGV, leur forme d'adhérence au substrat (libre ou liée) mais aussi de mesurer l'activité enzymatique libre et liée à la biomasse. Nous avons également défini en tant que contrôle la cinétique d'un réacteur « blanc » non inoculé afin de mesurer l'activité de la population microbienne endogène au substrat puisque les résidus de maïs n'ont pas été stérilisés. Ces travaux nous ont permis de consolider et de développer nos connaissances sur le fonctionnement des communautés bactériennes lignocellulolytiques.

## **B. Enrichment, stability and scale-up of a cow-rumen derived microbial consortium producing carboxylates from corn stover**

Emeline Auer<sup>1</sup>, Emilie Alaux<sup>1</sup>, Guillaume Caillou<sup>1</sup>, Diego P. Morgavi<sup>2</sup>, Claire Dumas<sup>1</sup>, Guillermina Hernandez-Raquet<sup>1</sup>

<sup>1</sup>) TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

<sup>2</sup>) INRAE, UMR1213 Herbivores, Site de Theix, 63122, Saint Genès Champanelle, France

---

### **1. Abstract**

The bioconversion of lignocellulose for the production of biofuel and chemicals is an attractive alternative to the use of petroleum resources. Enriched microbial consortia with enhanced lignocellulose degradation capacities have been proposed as an effective tool for lignocellulose bioconversion. In this study, two strictly anaerobic mesophilic microbial consortia were enriched by sequencing batch reactor process (SBR), using unpretreated corn stover as sole carbon source and cow rumen as inoculum source. After 8 cycles of enrichment, two stable lignocellulolytic carboxylates-producing microbial consortia were obtained. The two consortia showed a fairly similar community composition, being mainly constituted by members of Bacteroidetes and Firmicutes phyla, but were characterized by different carboxylates production capacities. The most performing enriched community, designated as RCS consortium, efficiently converted lignocellulosic biomass into carboxylates over 21 days. Xylanase activity was mainly produced during the early steps of biomass degradation, correlating with maximum rates of carboxylates production, and was mostly detected on the cellbound fraction. RCS consortium was successfully scaled up to 30 L bioreactor after a cryo-preservation step at -80°C, maintaining its lignocellulose degradation capacity and microbial community diversity. Overall, these results demonstrate the potential of RCS consortium to convert raw corn stover into useful synthons.

### **2. Introduction**

Due to its high polysaccharides content, lignocellulosic biomass, such as agricultural residues, is the most abundant source of renewable carbon on Earth and is considered to be a viable alternative to fossil carbon resources for the sustainable production of industrial chemicals and

biofuels (Cherubini and Strømman, 2011; Kim and Dale, 2004). According to the Food and Agriculture Organization of the United Nations (FAO), global corn production exceeded 1.14 billion tons in 2019 around the world, making it the second world largest crop, after sugar cane. Furthermore, corn stover is an attractive resource for biorefinery as there is two tons of lignocellulosic residues (mainly stems, leaves and stripped cobs) per ton of corn grain produced.

In ruminant animals, the decomposition of plant cell wall is performed by a complex microbiome (bacteria, archaea, fungi and protozoa) which deploy a unique set of hydrolytic enzymes such as cellulases, hemicellulases and lignin-degrading enzymes to hydrolyze the lignocellulose and to transform polysaccharides into short-chain volatile fatty acids (VFA or carboxylates), providing nutrients for the host animal. In the biorefinery context, carboxylates represent building blocks for the production of polyhydroxyalkanoates (PHA), precursors for bioplastics (Serafim et al., 2008), or biofuels (Agler et al., 2011; Torella et al., 2013). Thus, bacteria from rumen microbiome can be exploited in controlled bioreactors to improve lignocellulose bioconversion into VFA. The use of natural ecosystem-derived microbial consortia, with a large arsenal of hydrolytic enzymes, is a promising alternative to the use of the secretome of single microorganisms. Moreover, compared to single strain systems, microbial consortia can be used in non-sterile systems where they show stronger stability and robustness in broad range of conditions.

The enrichment culture is an interesting method to simplify microbial consortia with specific functional properties such as VFA production capacities. Enriched microbial consortia also represent a simplified model to decipher the dynamic of microorganisms and enzymes involved in lignocellulose bioconversion. In previous studies, microbial communities from diverse environments such as soils (de Lima Brossi et al., 2016), compost (Kinet et al., 2015), cow rumen (Lazuka et al., 2015) and termite gut (Auer et al., 2017), were enriched under aerobic or anaerobic conditions on different biomass (wheat straw, switchgrass, corn stover). Some of these enrichments were designed to improve methane production by anaerobic digestion, or to improve the VFA production by methanogenesis inhibition (Lazuka et al., 2015; Webster et al., 2016). In the present study, enrichment was performed by sequencing batch reactor process (SBR) in duplicate bioreactors under anaerobic mesophilic conditions, using corn stover residues as sole carbon source and cow rumen as inoculum source and inhibiting the methanogenesis. VFA production and 16S rRNA gene diversity were monitored throughout the enrichment process. Stability of the VFA production and community composition after cryo-preservation and scale-up

to 30 L bioreactor were assessed. After cryo-preservation of the best enriched VFA-producing consortium (named RCS), kinetic of lignocellulose bioconversion was detailed to identify the different steps of bioconversion by RCS. To study liquid-solid dynamics during bioconversion, xylanase activity and microbial composition were followed separately in the free and substrate-attached fractions.

### 3. Materials and methods

#### a) Lignocellulosic substrate and inoculum

The whole plant of corn stover (variety P9900) was harvested without cobs at an experimental farm (INRAE, St-Martin-de-Hinx, France). After harvest, corn stover was shredded using a knife mill (Retsch SM 100, Germany), with a screen size of 4 mm and was stored in zipper locked bags at room temperature.

The rumen used as inoculum of the first enrichment cycle was obtained from non dairy Holstein cows raised according to French national standards set by the animal welfare legislation (Certificate of Authorization to Experiment on Living Animals, N°004495, Ministry of Agriculture, France). Cows were fed with a ration composed of corn silage (64% DM), hay (6% DM) and concentrate (30% DM). They were fed *ad libitum* once a day in the morning and rumen inoculum sampling was done just before feeding. The whole rumen content of three cows was collected with rumen cannula, taken from various parts of the rumen and manually homogenized. Subsamples (50g) were snap frozen in liquid nitrogen before storage at -80°C until further use.

#### b) Anaerobic bioreactors

##### (1) Bioreactors setup

Enrichment and kinetics experiments were carried out in duplicate batch anaerobic bioreactors (2L BIOSTAT® A+, Sartorius, Germany), using a mineral medium containing milled corn stover residues as sole carbon source (20 g.L<sup>-1</sup>). Each bioreactor contained (for 1 L) : 20 g of substrate, KH<sub>2</sub>PO<sub>4</sub> : 3,42 g ; K<sub>2</sub>HPO<sub>4</sub> : 0,846 g ; NH<sub>4</sub>Cl : 5,58 g ; NaCl : 0,9 g ; MgCl<sub>2</sub>.6H<sub>2</sub>O : 0,1599 g ; CaCl<sub>2</sub>.2H<sub>2</sub>O : 0,09 g ; Biotine : 0,002 mg ; p-aminobenzoate : 0,01 mg ; thiamine : 0,01 mg ; pantothenate : 0,005 mg ; pyridoxamine : 0,05 mg ; vitamin B12 : 0,02 mg ; nicotinate : 0,02 mg ; H<sub>3</sub>BO<sub>3</sub> : 0,6 mg ; FeSO<sub>4</sub>.7H<sub>2</sub>O : 1,1 mg ; CoCl<sub>2</sub>.6H<sub>2</sub>O : 0,19 mg ; MnCl<sub>2</sub>.4H<sub>2</sub>O : 0,05 mg ; ZnCl<sub>2</sub> : 0,042 mg ; NiCl<sub>2</sub>.6H<sub>2</sub>O

: 0,024 mg ; NaMoO<sub>4</sub>.2H<sub>2</sub>O : 0,018 mg ; CuCl<sub>2</sub>.2H<sub>2</sub>O : 0,002 mg ; and was completed with distilled water to obtain a final corn stover concentration of 2% (w/v).

The total volume of the bioreactors were 3L : 2.4 L were used as working volume and 0.6 L as headspace. Bioreactors were operated under mesophilic conditions (35°C) and controlled pH (6.15) by NaOH (1N) addition to favour carboxylates production, as detailed by Lazuka et al., (2015a). To ensure anaerobic conditions, the headspace was flushed with nitrogen gas for 5 minutes at the beginning of each bioreactor. The absence of dissolved oxygen was continuously monitored with a polarographic dissolved oxygen probe (AppliSens). Bromoethanosulfonate (BES, 1mM) was added at the beginning of each bioreactor to inhibit methanogenesis (Webster et al., 2016). If methane was detected in the headspace during the experiment, BES was spiked until a maximum concentration of 5 mM in the bioreactor.

## (2) Enrichment process

At the beginning of the enrichment process, a first batch bioreactor was inoculated at 2% (w/v) with the liquid and solid content of cow rumen (48 g pooled from three cow rumen samples thawed overnight at 4°C). The first cycle of each replicate-enrichment bioreactor was incubated 14 days before removing 10% (v/v, eq 240 mL) of the whole content of the bioreactor (composed of solid and liquid fractions), which was used as inoculum for a new subsequent batch bioreactor. In this manner, eight successive cycles were performed until VFA reached stable concentration (variable duration from 9 to 14 days). After the 8<sup>th</sup> cycle, samples of 240 mL of the whole content of the bioreactor (solid and liquid), which contained the enriched community displaying the best VFA production capacity, were snap frozen in liquid nitrogen and stored at -80°C without cryo-protectant until further use (for a period ranging from 2 weeks to 1 year).

## (3) Cryo-preservation, scale-up and characterization

To determine the impact of cryo-preservation on the microbial diversity and VFA production of the enriched consortium, a ninth cycle of culture was carried out in duplicate bioreactors (2 L BIOSTAT® A+, Sartorius, Germany), using as inoculum RCS frozen sample (10% v/v, eq 240 mL) stored at -80°C and completely thawed at 4°C overnight. To evaluate the stability of VFA production and community composition of RCS consortium during scale-up, a fermentation experiment was conducted in a 30 L-reactor (30 L BIOSTAT® C+, Sartorius, Germany) with 24 L as working volume and 6 L as headspace. The culture medium and conditions described previously

was adapted to the volume of the reactor in the same proportion, with an inoculum volume of 2.5 L (completely thawed at 4°C overnight) and 500 g of corn stover substrate. Lignocellulose bioconversion kinetics of RCS consortium was deeply characterized along twenty-one days incubations in the same conditions and bioreactors (2 L BIOSTAT® A+, Sartorius, Germany) as described above. An uninoculated reactor was also performed to measure the activity of endogenous microorganisms possibly present in the corn stover used as substrate.

### **c) Biochemical analysis**

#### **(1) Volatile fatty acids production**

VFA production (acetate, propionate, iso-butyrate, n-butyrate, iso-valerate, n-valerate and caproate) was determined in the supernatant by gas chromatography using a Varian 3900 chromatograph as described by Cavaillé et al., (2013).

#### **(2) Xylanase activity**

Samples from the duplicate bioreactors were regularly withdrawn and centrifuged (7,197 g, 10 minutes, 4°C) to separate the supernatant and the solid pellet. The solid pellet was suspended in 6 mL of acetate buffer solution (50 mM, pH 6) containing sodium azide (0.05% w/v) and sonicated on ice four times for 20 secondes using a sonication device (Bandelin Sonoplus HD 2070, MS73 probe) at 40%.

The xylanase activity of the supernatant and pellet of each sample, containing respectively extracellular and cell-bound xylanases, were measured using 1% (w/v) xylan beechwood (Sigma Aldrich) dissolved in buffer solution. Xylanase activity was estimated by measuring the release of reducing sugars after 1h incubation at 35°C. Reducing sugars were determined using the dinitrosalicylic acid (DNS) method, measuring absorbance at 570 nm using a UV/VIS spectrophotometer (Multiskan Ascent, Thermo Scientific). One unit of xylanase activity (UA) was defined as the amount of enzyme that produces 1 µmol of reducing sugars per minute. Xylanase activity was expressed for each time point as average values obtained on quadruplicate samples.

### **d) Microbial diversity analysis**

At each sampling point, two 1 mL samples were collected for microbial diversity analysis. The first one was directly ultra-centrifuged (13,000 g, 5 min, 4°C), its supernatant was removed and its



pellet was snap frozen into liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until nucleic acid extraction. This sample will determine the microbial community in the bioreactor (independently of the solid or liquid fraction). The second one was fractionated, using similar methods to Mouriño et al., (2001) and Jensen et al., (2008), into attached biomass (final pellet), free biomass (first supernatant) or weakly attached biomass (supernatant after resuspension of the pellet in 1 mL of 0.9% NaCl and a second low speed centrifugation) by two centrifuging cycles at low speed (150 *g*, 10 min,  $4^{\circ}\text{C}$ ) before ultra-centrifugation (13,000 *g*, 5 min,  $4^{\circ}\text{C}$ ). Each fraction-pellet was snap frozen into liquid nitrogen. For nucleic acids extraction, DNA and RNA were extracted from the frozen pellets using the PowerMicrobiome RNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA) following the manufacturer's instructions. Cell lysis was carried out using a FastPrep Instrument (MP Biomedicals, Santa Ana, CA, USA;  $4 \times 30$  s at  $6 \text{ m.s}^{-1}$ ). Then, DNA purification was performed using the AllPrep DNA/RNA mini kit (Qiagen) following the manufacturer's instruction. The hypervariable V3-V4 region of the 16S rRNA gene was amplified by Illumina MiSeq sequencing using the GenoToul Genomics and Transcriptomics platform (Toulouse, France) and conditions and primers previously described (Auer et al., 2017). A minimum of 21,000 high quality reads were generated for each sample. Joined-pair reads of samples were simultaneously imported in FROGS (Find Rapidly OTUs with Galaxy solution) pipeline according to its guidelines (Escudie et al., 2017). Briefly, data preprocessing removed sequences outside a 380-460 bp range, presenting mismatch in primers or N calls. Aggregation distance was set to  $d=3$ , and clustering was performed with Swarm algorithm (Mahé et al., 2014) after a denoising step. Operational taxonomic units (OTUs) with a total abundance lower than 10 sequences across the overall dataset were discarded using a control synthetic mock community added in the sequencing run as calibration of the sequencing noise threshold (Bokulich et al., 2013). Chimera were removed with VSEARCH (Rognes et al., 2016) using default FROGS parameters, and taxonomic affiliation was performed by Blast using Silva 132 16S with a pintail of 100 as reference database (Ashelford et al., 2005; Quast et al., 2013). All *Ralstonia* sequences were considered as contaminants and were removed from the dataset.

### **e) Data analysis and visualization**

R CRAN software (version 3.5.0) was used to analyze the obtained data. Abundance table was manipulated using the R packages *dplyr* and *tidyr*. The 16S data processing followed guidelines in Cao et al., (2016) and consisted of (i) total sum scaling (TSS) normalization to account for uneven sequencing depth across samples (each OTU count is divided by the total number of counts in each individual sample), (ii) aggregating OTUs at families level (sum of relative abundance of OTUs of

the same family), (iii) filtering and gathering low abundance families in "Others" category and (iv) transforming TSS data using Centered Log Ratio (CLR) transformation to account compositional data generated by TSS normalization using an offset of 1 (Calle, 2019). Alpha diversity indexes (richness, Simpson and Shannon) and  $\beta$ -diversity distance (Bray-Curtis distance) were obtained using the R package *vegan* (version 2.5-7, Oksanen et al., (2020)). Non-Metric Multidimensional Scaling (NMDS) ordinations were performed using *metaMDS* function of R package *vegan* (v2.5-7) Hierarchical clustering was performed using R function *hclust*. Barplots and boxplots were plotted using R's *ggplot2* package (v3.3.5). Significance differences were tested using non-parametric Wilcoxon rank sum tests (also called Mann-Whitney test, Xia and Sun, (2017)) or using two-tailed Student's t-test and p-value resulting from pairwise comparisons were adjusted using False Discovery Rate (FDR) adjustment method using *ggpubr* package (v0.4.0). VFA and bacterial families' dynamics were thus integrated using R package *timeOmics* (v1.4.0), dedicated to time course studies of microbial data (Bodein et al., 2019). Smoothed data and derivatives were obtained after polynomial regression on the raw data.

#### 4. Results and discussion

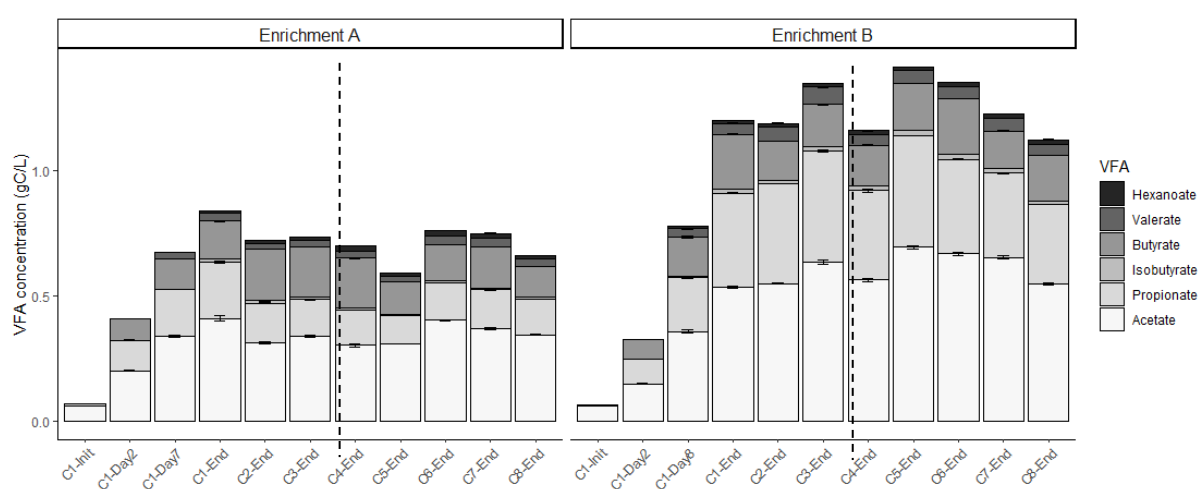
##### a) Enrichment of a corn stover-degrading microbial consortium

To enrich a stable lignocellulolytic microbial consortium specialized on corn stover transformation from cow rumen, an enrichment strategy based on a sequential batch culture process was deployed. For this purpose, two independent enrichments (A and B) were performed into bioreactors (2.4 L) using corn stover as sole carbon source at 2% TS and inoculating at 2% (w/v) with cow rumen. To obtain maximal degradation, the duration of the 1<sup>st</sup> cycle was set at 14 days as fibrolytic bacteria require long periods of time culture (Chassard et al., 2005; Mirande et al., 2010).

Surprisingly, although identical conditions were applied to both bioreactors, enrichments A and B quickly diverged regarding their VFA production profiles (Figure V-1). In both enrichments, VFA production and composition were relatively stable since the end of the first cycle of enrichment, but displayed notable differences between them. While VFA production was similar at the beginning of the 1<sup>st</sup> cycle on days 2 and 7/8 in both enrichments, a difference became apparent at the end of the 1<sup>st</sup> cycle, with enrichment B showing 50% higher VFA production than enrichment A, reaching  $1,200.5 \pm 9.7$  mgC.L<sup>-1</sup> and  $839.2 \pm 24.2$  mgC.L<sup>-1</sup> over 14 days, respectively (Figure V-1).

## Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

This difference in VFA production was maintained over 8 enrichment cycles, as enrichments A and B reached an average production from cycle 1 to cycle 8, respectively,  $720.5 \pm 73.1 \text{ mgC.L}^{-1}$  and  $1250.4.1 \pm 107.0 \text{ mgC.L}^{-1}$  (Figure V-1). At each end cycle, the major VFA produced was acetate with a similar average fraction in enrichments A ( $48.6 \pm 4.1\%$ ) and B ( $48.4 \pm 2.6\%$ ). Regarding the composition of other carboxylates, enrichment A displayed a propionate:butyrate mean ratio of 0.96:1 while enrichment B produced relatively more propionate than butyrate with a ratio of 2.1:1. These observations showed that lignocellulose bioconversion, and particularly the VFA production capacities, were different between these enriched communities and were probably related to differential taxa selection in the two communities that occurred during the 1<sup>st</sup> enrichment cycle.



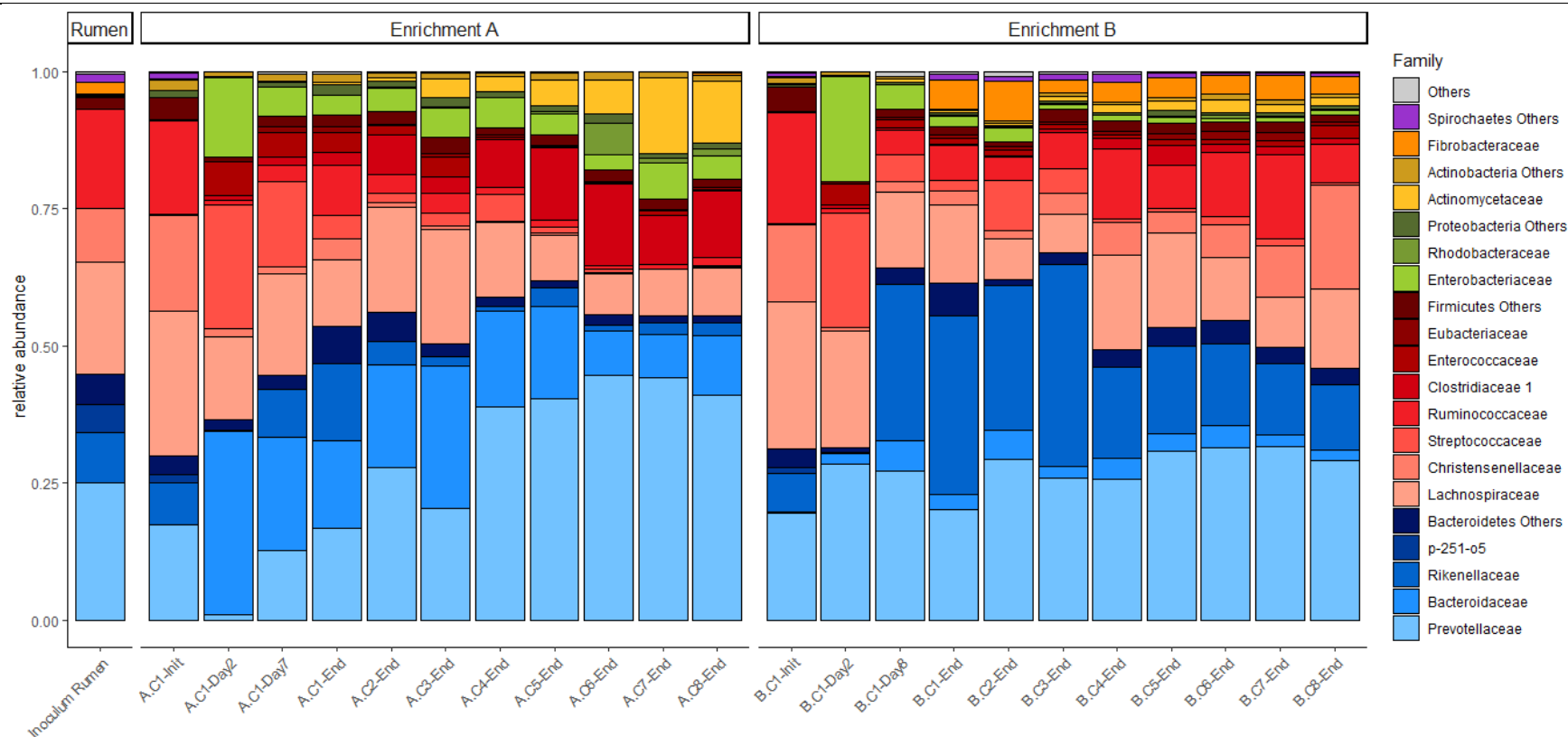
**Figure V-1 : Volatile fatty acids (VFA) production (gC/L) at the end of each cycle of the two corn-stover enriched communities throughout the enrichment process. For cycle 1, the initial (day 0), intermediate (day 2 and 7/8) and final cycle measurements are shown. VFAs acetate, propionate, isobutyrate, butyrate, valerate and hexanoate are distinctly colored. Error bars are standard deviations of technical replicates ( $n \geq 3$  technical replicates).**

To gain insights on the changes occurring in the community composition, we applied 16S rRNA gene amplicon sequencing to characterize the diversity of the cow rumen inoculum, the different sampling points of the 1<sup>st</sup> enrichment cycle (days 0, 2, 7/8 and 14) and the end of each cycle of both enrichments. The observed richness, Shannon and Simpson indexes (Additional file 1: Supplementary Table S1) dropped drastically between the inoculum and the end of cycles 1 of both enrichments reflecting the loss of some rumen species. This diversity decrease was strongly noticed after the 2<sup>nd</sup> day of culture in cycle 1, before a slight re-increase until the end of cycle 1. Subsequently, there were only minor changes in the diversity of communities for both enrichments: alpha diversity indexes remained relatively stable from cycle 2 to 8 for both

enrichments, but richness was slightly higher in enrichment B (Additional file 1: Supplementary Table S1). It can be noted that the greatest diversity change occurred during the initial enrichment phase, when the microbial community structure is simplified. This result can certainly be attributed to the adaptation of substrate and to the acclimatization of the rumen community to the culture conditions in the bioreactors fostering the development of species more adapted to them. Indeed, analysis of the alpha (Additional file 1: Supplementary Table S1) and beta diversity (Bray-Curtis distance – Additional file 2 : Figure S1) indexes clearly showed that the two enrichments diverged at the end of the first cycle of enrichment and these two corn stover degrading communities stabilized around the 4<sup>th</sup>-6<sup>th</sup> enrichment cycle. Rapid stabilization of the community structure has already been reported by Lazuka et al., (2015), with stabilization of a cow rumen-derived consortium after five cycles of enrichment on raw wheat straw. Similar results have also been reported for other enriched lignocellulolytic microbial consortia (Eichorst et al., 2013; Yu et al., 2015). This rapid stabilization highlights the stringency of the enrichment conditions and confirms that using lignocellulosic biomass as a sole carbon source is a satisfying selection pressure to select a simplified lignocellulose-degrading microbial community after a small number of subculture steps.

Further analysis of the community composition (Figure V-2) showed that both enrichments were characterized by high relative abundance of OTUs related to the phylum Bacteroidetes (56.0±3.4% and 55.5±7.4% in enrichments A and B, respectively), followed by Firmicutes (A = 30.4±6.6% and B = 35.3±7.9%). These major phyla were also very abundant in the initial rumen inoculum (44.9% for Bacteroidetes and 50.5% for Firmicutes) in agreement with previous studies identifying Bacteroidetes and Firmicutes as the dominant phyla in cow rumen (Hess et al., 2011) and in lignocellulose-degrading and carboxylates-producing anaerobic communities (DeAngelis et al., 2012; Lazuka et al., 2015). In contrast, strong differences were observed for minor phyla found initially in the rumen inoculum, some of which were absent in the enrichments and others behaved differently in each enrichment. While enrichment A showed a low relative abundance of Proteobacteria (6.9±1.7%) and Actinobacteria (6.5±4.9%), these two phyla were even less abundant in enrichment B (2.0±0.5% and 1.9±0.9%, respectively). For instance, while the rumen inoculum showed OTUs related to *Fibrobacteraceae* and Spirochaetes, these taxa displayed dissimilar relative abundances in both enrichments. Enrichment B contained Fibrobacteres (4.2±1.5%) and Spirochaetes (0.8±0.3%), which were present at very low abundance in enrichment A (0.1±0.2% for Fibrobacteres and 0.1±0.0% for Spirochaetes).

Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs



**Figure V-2 : Relative distribution of the bacterial taxa (Family level) of the two corn-stover enriched communities throughout the enrichment process.** Relative abundances of OTUs were aggregated at the Family level for stacked barplot representation. “Others” group consist of Phylum with low relative abundances (< 1%) and “Phylum Others” groups consist of Families with low relative abundance (< 5%) within major Phyla identified in the dataset. OTUs belonging to the same bacterial Phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Actinobacteria (yellow), Fibrobacteres (orange), Spirochaetes (purple). Each bar corresponds to an experiment (Inoculum rumen, Enrichment A, Enrichment B), a cycle, and a time point, annotated with the corresponding day at the bottom of the graph.

## Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

In order to identify the species that contribute the most to the differences observed in the community composition of both enrichments, we compared the abundance of bacterial taxa between them (Additional file 2 : Figure S2). Wilcoxon statistical analysis showed significant differences ( $p < 0.05$ ) for certain families such as “*p-251-o5*”, *Rikenellaceae* and *Bacteroidaceae* (Bacteroidetes), *Christensenellaceae*, *Clostridiaceae 1* and Firmicutes Others (Firmicutes). Enrichment A-consortium, the lower VFA producer, was characterized by a higher content of *Bacteroidaceae* ( $15.2 \pm 6.1\%$  and  $3.2 \pm 1.2\%$  for enrichment A and B, respectively) and *Clostridiaceae 1* ( $8.8 \pm 4.5\%$  and  $1.4 \pm 1.1\%$  for enrichment A and B, respectively) families than in the Enrichment B-consortium. These families are frequently reported as key bacterial taxa involved in lignocellulose fermentation systems (Zhu et al., 2016; Chukwuma et al., 2021). For instance, *Bacteroidaceae* members are known as coding wide range of CAZymes organized as Polysaccharide Utilization Loci (Naas et al., 2014; Grondin et al., 2017; Lapébie et al., 2019) while those of *Clostridiaceae* family include well-known lignocellulose degrading and VFA-producer species such as *Clostridium* (Desvaux, 2005). This last is frequently selected in lignocellulose-degrading bioreactors and has been reported to be involved in the production of CAZymes and proteins related to VFA production (Wang et al., 2011; Lazuka et al., 2015). *Bacteroidaceae* and *Clostridiaceae* families were previously found in the RWS consortium which was enriched using cow rumen inoculum and same enrichment conditions than those applied here, and efficiently produced VFAs from wheat straw (Lazuka et al., 2015). Nevertheless, the abundances of such families in the RWS consortium was greater than that observed in present enrichments obtained using corn stover as substrate. Indeed, in RWS consortium, *Bacteroidaceae* family was dominant ( $38.3 \pm 7.1\%$ ) throughout wheat straw bioconversion, followed by *Clostridiaceae* ( $6.3 \pm 4.7\%$ ).

Enrichment B-consortium, which was the higher VFA-producing enrichment, displayed higher abundance of *Ruminococcaceae* (3-fold), *Rikenellaceae* (5.5-fold) and *Christensenellaceae* (8-fold), and also a low but stable abundance of *Fibrobacteraceae*, present in remarkably higher abundance (42 fold) in enrichment B compared to enrichment A. *Fibrobacter* and *Ruminococcus* are well-known ruminal genera involved in the production of cellulolytic enzymes, playing critical role in biomass deconstruction (Dai et al., 2015; Neumann et al., 2017). In the rumen ecosystem, members of these families have been reported as coding large parts of the putative glycoside hydrolases with cellulases and hemicellulases activities (Dai et al., 2015) and as succinate producer with lesser amounts of acetate (Ransom-Jones et al., 2012; La Reau and Suen, 2018). However, they are strict anaerobes with low growth rates, making them difficult to maintain under

Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

---

bioreactor conditions (Ransom-Jones et al., 2012; La Reau and Suen, 2018), which explains their low abundances in these enrichments. *Christensenellaceae* were described recently (Waters and Ley, 2019), and few data are available concerning their ecology, excepting in human gut where they were described as highly correlated with *Rikenellaceae* and associated with lean phenotypes (Tavella et al., 2021). Isolates of the genus *Christensenella* displayed saccharolytic capacities with acetate and butyrate as fermentation end-products (Morotomi et al., 2012; Zou et al., 2021). *Rikenellaceae* is a small family of phylum Bacteroidetes with only a few sequenced genomes. Isolated species displayed fermentation capacities on starch and pectins, with acetate, propionate and succinate as end-products (Graf, 2014; Su et al., 2014).

The temporal dynamics of the major families throughout the first enrichment cycle for the two enrichments (Additional file 2 : Figure S3) revealed that, although *Fibrobacteres* were present in low abundance in the rumen inoculum and at the beginning of the incubation in both enrichments, dilution and growing of other taxa made them disappear in a few days, before they were detected again after day 8 in enrichment B only. Similarly, *Rikenellaceae* was present in greater abundance at the end of cycle 1 in enrichment B only. Higher VFA production by Enrichment B-consortium suggests that the presence of these microbial groups in greater abundance in enrichment B could provide a greater capacity for corn stover bioconversion into VFA under the bioreactor conditions. Although our experimental design was not set up to study biological replicates heterogeneity, previous studies have suggested that differences between biological replicates result from stochastic processes occurring under strong selection pressure (Vanwonterghem et al., 2014; Zhou and Ning, 2017) These differences may also result from microscale heterogeneity that may prevails in the initial rumen inoculum, independently of its macroscopic homogenization (Dann et al., 2018). Further experiments including a larger number of biological replicates of the first cycle of enrichment would be needed to validate this hypothesis.

### ***b) Stability after cryo-preservation and scale-up of RCS consortium***

Due to its higher bioconversion rate, Enrichment B-consortium was selected for further experiments and was named RCS consortium. In order to characterize the cryo-preservation stability of the enriched RCS consortium, we determined the activity (VFA production) and the community composition of RCS after storage at -80°C. Two biological replicate bioreactors were performed using end-sample of Cycle 8 of the enrichment B that were snap frozen in liquid



nitrogen without cryo-protectant, stored at  $-80^{\circ}\text{C}$  and defrosted overnight at  $4^{\circ}\text{C}$  as inoculum.

Compared to enrichment cycles 6 to 8, culture end points of the cryo-preserved bioreactors showed no significant differences in VFA production ( $1.18\pm 0.03$  gC/L, Additional file 2 : Figure S4A) and similar VFA profile (acetate 52.6%, propionate 28.7%, butyrate 13.3%, valerate 3.1%, isobutyrate 1.1%, hexanoate 1.1%). Similarly, bacterial communities appeared unaffected by freezing as alpha and beta diversities indexes and relative abundances of families described previously were similar (Additional file 2 : Figure S4B and Figure S4C). Major phyla found in RCS consortium, such as Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria, could explain this robustness as they were previously described as resilient to cryo-preservation at  $-80^{\circ}\text{C}$  in liquid nitrogen without cryo-protectant (Yu et al., 2015). Most of the protocols existing for long-term preservation of microorganisms have indeed been developed for pure culture microorganisms. Kerckhof et al., (2014) showed that cryo-preservation of different mixed microbial communities was favoured by the addition of cryo-preservation agents (dimethyl sulfoxide -DMSO- supplemented or not with trehalose and tryptic soy broth) that avoid cellular damage. Nevertheless, the net benefit of the addition of protective agents varies in function of the microbial community under study being very positive for methanotrophic ecosystems but not needed for complex faecal microbiomes (Kerckhof et al., 2014). Yu et al., (2015) showed that the community structure of a thermophilic switchgrass-degrading microbial consortia shifted after cryo-preservation, regardless the addition of protective agent during cryo-preservation. The authors highlighted a strong impact of cryo-preservation on specific microbial taxa (such as Chloroflexi and Planctomycetes) while other taxa, such as Firmicutes and Actinobacteria, showed a strong tolerance to cryo-preservation and species related to Acidobacteria, Bacteroidetes, Proteobacteria and Verrucomicrobia were not pertubated. These observations suggest that the effect of cryo-preservation is a function of the initial community composition and, in the case of the RCS consortium, the taxa present were cryo-preservation tolerant. In addition, preservation of the community on the enrichment medium, as we have done for the RCS consortium, has been reported to improve tolerance to cryo-preservation (Haruta et al., 2002).

A scaling-up experiment in a 30 L bioreactor (with working volume of 25 L) was also performed to assess the ability of RCS consortium to be cultivated at a larger scale. Compared to the final cycles of enrichment B, the cryo-preservation and scale-up experiments showed similar VFA production levels with  $1.18\pm 0.03$  gC/L and 1.5 gC/L, respectively. A slight decrease of richness was observed in the scale-up experiment but the relative abundances of main RCS families remained at similar



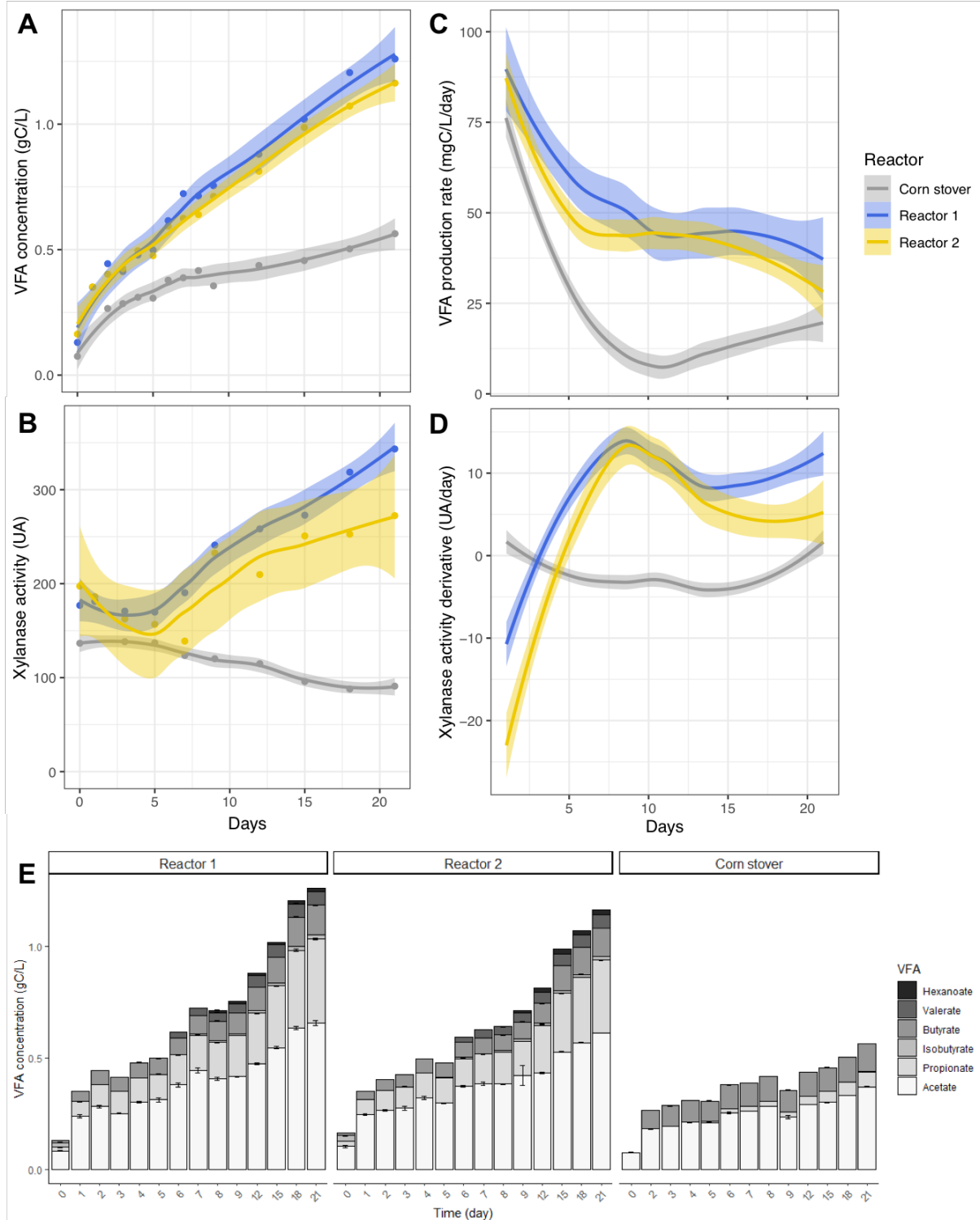
levels. Despite these small variations, the microbial consortium remained highly stable: C9a and C9b, obtained after cryo-conservation of RCS consortium, clustered with cycles 6 to 8 of enrichment B in both NMDS analysis and ward-D2 hierarchical clustering based on Bray-Curtis dissimilarity (Additional file 2 : Figure S4D). The scaled-up reactor was slightly more distant in NMDS analysis, but hierarchical clustering showed that it was still grouped with Enrichment B cycles 6 to 8, and C9a and C9b. Small deviation of the scale-up reactor could be due to differences in reactor design: while media composition and temperature were the same, up-scaling in a different reactor setup may have implied more subtle variations, such as modifications of gas headspace volume, gas transfer or stirring parameters. The RCS consortium also appeared to be functionally and structurally robust for scaling-up. Some researchers found perturbations by upscaling bioreactor and reported significant yield loss in terms of enzyme production by isolate strain (Kumar et al., 2009a). Here, the VFA production capacity and community composition of the RCS consortium showed that both parameters were not significantly affected by RCS culture at larger scale. These results provide additional insights on the robustness of enriched microbial consortia for the production of useful synthons from lignocellulosic biomass.

### **c) *Corn stover bioconversion dynamics of RCS consortium***

#### **(1) *VFA production and xylanase activity dynamics***

The temporal dynamic of VFA production from corn stover for both RCS-inoculated experiments (Reactors 1 & 2) were similar (Figure V-3A, blue and yellow curves), reaching a maximum average VFA concentration of  $1.2 \pm 0.02$  gC/L, with a similar VFA composition (Figure V-3E), in agreement with the values obtained during the enrichment process. The produced VFA were composed of volatile fatty acids of two to six carbon molecules in decreasing proportions ranging from acetate (C2, 52.4%), propionate (C3, 29.0%), butyrate (C4, 10.9%), valerate (C5, 5.0%), isobutyrate (C4, 1.4%) to hexanoate (C6, 1.4%). While bioconversion dynamic of the Reactor 2 seemed slightly delayed compared to Reactor 1, both dynamics were well replicated. The kinetic of production of volatile fatty acids by inoculated bioreactors is divided into 3 periods: from the beginning to 5-6 days with a decreasing production rate ( $87.5-50$  mgC.L<sup>-1</sup>.Day<sup>-1</sup>) before stabilizing between 5-6 and 15 days at a rate around  $50$  mgC.L<sup>-1</sup>.Day<sup>-1</sup> (Figure V-3C, blue and yellow curves). Finally after 21 days, the VFA production rate decreases around  $30$  mgC.L<sup>-1</sup>.Day<sup>-1</sup>. It can be noticed that valerate and hexanoate production only started at days 5 and 6, at the same time that xylanase activity rised after an initial slight decrease of activity (Figure V-3D, blue and yellow curves), corresponding

also to the shift of VFA production rate. In contrast, uninoculated reactor (control) reached a maximum VFA concentration that stabilized around 0.3 gC/L after 5 days and reached 0.56 gC/L after 21 days (Figure V-3A, grey curve). The VFA profile of uninoculated reactor was composed mainly of acetate (65.6 %), butyrate (21.9 %), propionate (11.8%) and small amount of isobutyrate (0.7%) (Figure V-3E). This control reactor showed basal VFA production and a low and decreasing xylanase activity (Figure V-3B-C-D, grey curve), probably resulting from abiotic hydrolysis of corn stover and activity of the endogenous microflora present on the substrate, which was not sterilized to simulate industrial processes.



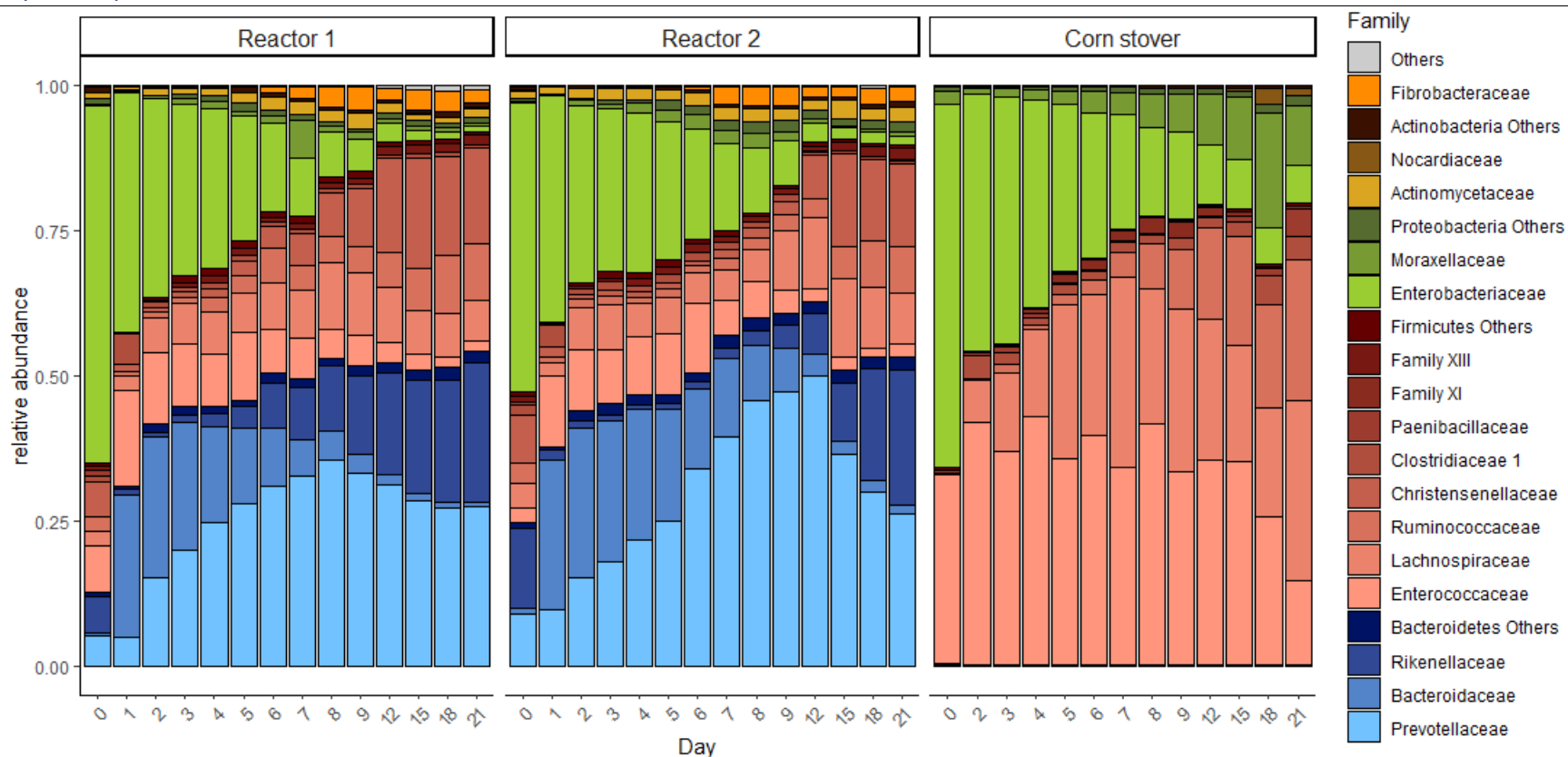
**Figure V-3 : Temporal dynamic of RCS corn stover bioconversion into VFAs and xylanase activity in duplicate inoculated bioreactors (Reactors 1 and 2) and one uninoculated control (Corn stover).** VFA concentration in gC/L (A) and xylanase activity in UA (B) are detailed for each bioreactor and sampling point, curves representing smoothed values. VFA production rate in gC.L<sup>-1</sup>.day<sup>-1</sup> (C) and xylanase derivative in UA.day<sup>-1</sup> (D) were calculated using the smoothed values presented in (A) and (B). Shaded colors represent the standard error of the

smoothing models. (E) VFA cumulative composition of each bioreactor at each time point. Error bars are standard deviations of technical replicates ( $n \geq 3$ ). Each bar corresponds to time point (days). The two RCS inoculated reactors are named "Reactor 1 or 2", the uninoculated reactor is named "Corn stover".

## (2) Diversity analysis and roles of RCS community members

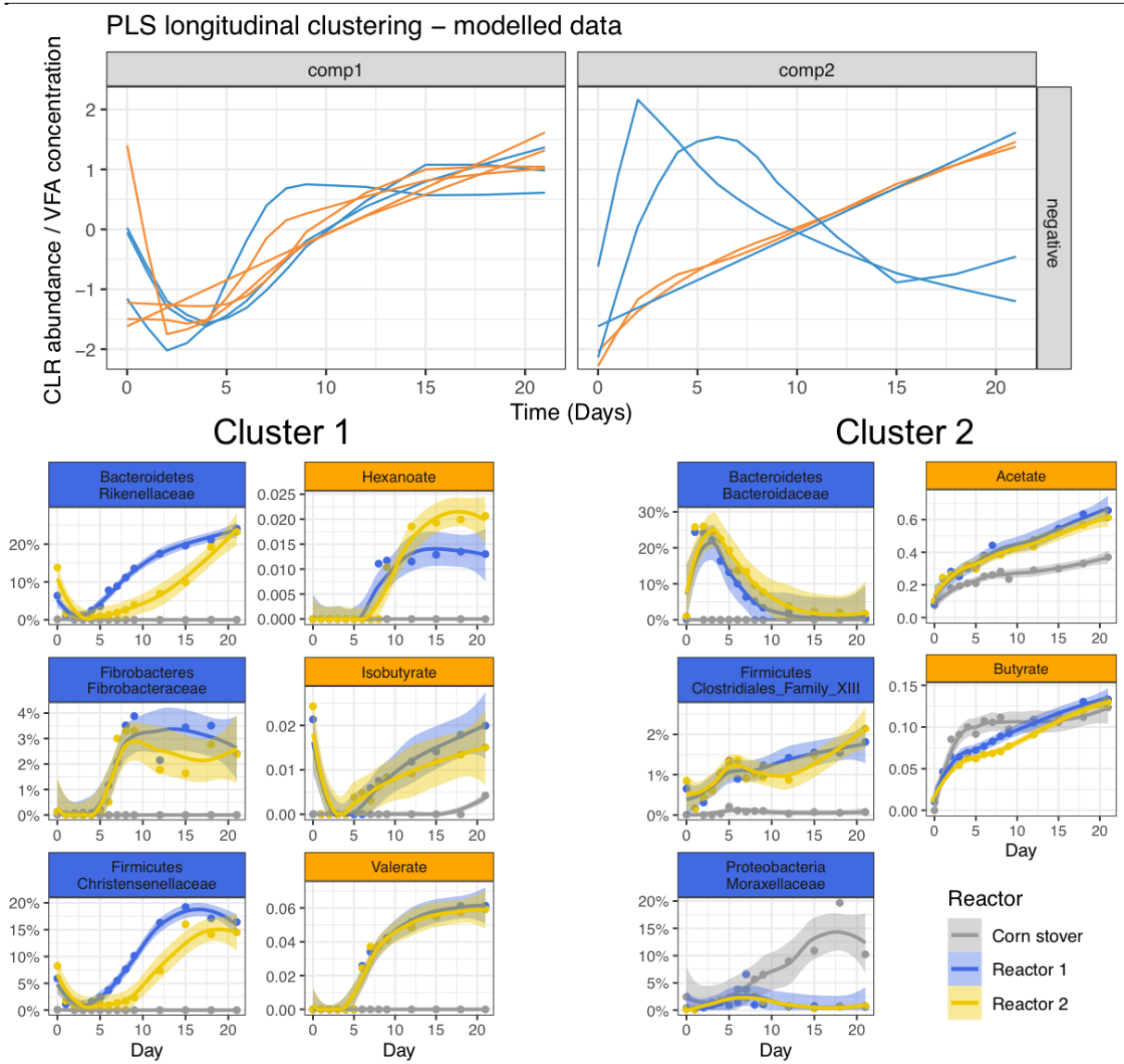
Amplicon sequencing of the 16S rRNA gene (Figure V-4) from the both RCS-inoculated bioreactors showed a clear dominance of Bacteroidetes members followed by Firmicutes (Figure V-4). NMDS analysis confirmed the similar temporal dynamics followed by the RCS-microbial community in the two replicated bioreactors (Additional file 2 : Figure S5), supporting the stability and reproducibility of RCS enriched consortium. Uninoculated reactor showed that the substrate endogenous bacterial diversity was low (less than 200 species), mainly composed of *Enterococcaceae* (32.6%) (Firmicutes), *Enterobacteriaceae* (62.7%) and *Moraxellaceae* (2.4%) (Proteobacteria), described as phyllosphere taxa (Williams and Marco, 2014; Patz et al., 2019). Others phyla, particularly Bacteroidetes and Fibrobacteres, were completely absent in the control reactor. Some taxa that were not detected on the substrate or were rare at the beginning of the incubation in this control reactor, increased with time; for example, *Lachnospiraceae* found at 0.09% on day 0 reached 30.9% on day 21 and *Ruminococcaceae*, found at 0.04% on day 0, increased to 24.4% abundance on day 21. In contrast, *Enterococcaceae* and *Enterobacteriaceae* decreased over time from 32.6% and 62.7% at day 0 to 15% and 7% respectively at day 21. Comparison of the RCS-inoculated bioreactors with the uninoculated bioreactor (Figure V-4) showed that at the beginning of the incubation, the endogenous bacteria that were abundant in the substrate dominated those present in the RCS consortium. For example, *Enterobacteriaceae*, which were highly dominant in the substrate and present in very low abundance in the RCS inoculum were at similar levels in the inoculated and uninoculated reactors at the start of incubation. At the end of incubation, the main differences between RCS-inoculated bioreactors and the uninoculated control stemmed from the presence of Bacteroidetes members, which became dominant in two days in the RCS-inoculated bioreactors, representing about 50% of the total abundance at the end of incubation. *Actinomycetaceae* (Actinobacteria) and Fibrobacteres, whose abundance become detected after day 6 in the RCS-inoculated bioreactors, were absent in the uninoculated control. The temporal dynamics of RCS consortium bacteria is similar to that found during *in situ* incubation of switchgrass in the cow rumen (Piao et al., 2014). This previous study showed that biomass degradation in the cow rumen appeared to be biphasic and coordinated with microbial colonization : after a peak of Proteobacteria species at initial time and

a period where no significant change to biomass was measured, one shift occurred simultaneously with Fibrobacteres and Spirochaetes species and significant fiber degradation was visible by scanning electron microscopy (Piao et al., 2014). The analysis of the uninoculated control reactor allowed us to identify processes directly related with the RCS inoculum. Strong differences were observed in inoculated compared to uninoculated reactors, particularly after 5- 6 incubation days, when xylanase activity increased and longer fatty acids appeared (isobutyrate, valerate and hexanoate). These observations were concomitant with the increase in the abundance of Fibrobacteres, suggesting that this taxon, well-known for its ability to degrade cellulose and produce large amount of succinate which is converted by other bacteria to propionate, could be a key player in the bioconversion of corn stover.



**Figure V-4 : Dynamics of the bacterial taxa (Family level) throughout the corn stover bioconversion by RCS consortium.** Relative abundance of OTUs were aggregated at the Family level for stacked barplot representation. “Others” group consist of Phyla with low relative abundances (< 1%) and “Phylum Others” consist of Families with low relative abundances (< 2%) within the major phyla of the dataset. OTUs belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Actinobacteria (yellow), and Fibrobacteres (orange). Each bar corresponds to time point (days). The two RCS inoculated reactors are named “Reactor 1 or 2”, the uninoculated reactor is named “Corn stover”.

In order to identify the main RCS-community members (at the family level which reached relative abundance at least 2% in one sample) that followed a similar temporal behavior that the production of individual fatty acids and xylanase activity, an integrative longitudinal analysis was performed using *timeOmics* R package (Figure V-5). This approach enabled to identify two main clusters of microbial families. Cluster 1 included members of the families *Rikenellaceae* (Bacteroidetes), *Fibrobacteraceae* (Fibrobacteres) and *Christensenellaceae* (Firmicutes) which followed a similar time profile than isobutyrate, valerate and hexanoate production. These families and fatty acids showed low abundances or concentrations at the beginning of the incubation but a strong increase after 3-5 incubation days. This gathering suggests that these bacterial families, at least *Fibrobacteraceae*, were involved (directly or indirectly) in the late production of long chain VFA molecules (C5-C6) as previously described in co-culture experiment (Kenealy et al., 1995). Cluster 2 gathered members of the families *Bacteroidaceae* (Bacteroidetes), *Clostridiales\_Family\_XIII* (Firmicutes) and *Moraxellaceae* (Protobacteria) with acetate and butyrate production. The temporal dynamics of these bacterial families were characterized by abundance increase at the beginning of the incubation followed by a sharp decrease thereafter. This temporal pattern was related to the production rates of acetate and butyrate, which increased sharply at the beginning of the incubation, but staggered over time.

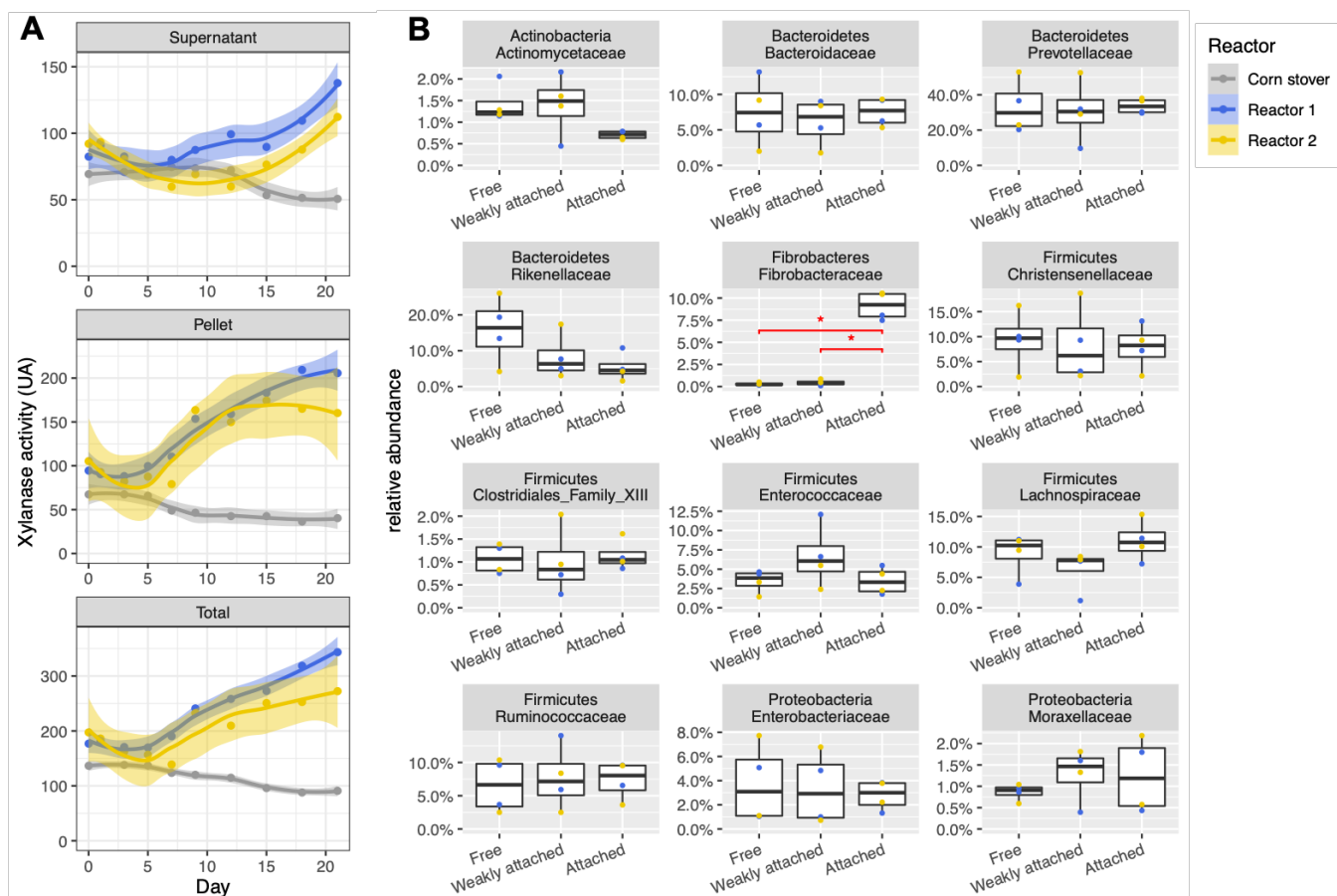


**Figure V-5 : Longitudinal analysis integrating dynamics of RCS main bacterial families (CLR transformed data) and individual fatty acid produced in each reactor, using *timeOmics R* package following the authors guidelines (Bodein et al., 2019). At the top : variables (individual fatty acids and main bacterial families) were modelled using linear mixed effect model, then the curves were clustered using Partial Least Square (PLS) regression. Propionate and uninoculated reactor were not considered due to modelling issues. Modelling curves corresponding to Taxa are colored in blue and those corresponding to volatile fatty acids in orange. At the bottom: Relative abundances of RCS main bacterial families and VFA concentration were plotted, smoothed curves of the uninoculated control (“Corn stover”) and the RCS-inoculated reactors 1 & 2 are colored in grey, blue and yellow, respectively. Shaded colors correspond to the standard error of the smoothing model.**



(3) Role of substrate-attached bacteria on xylanase activity

In natural fermentation systems, such as rumen or the gut of lignocellulolytic insects, enzymes or microorganisms can be attached to the substrate or free in the medium (Deusch et al., 2017). For instance, fibrolytic species such as *Ruminococcus* and *Fibrobacter* species require adhesion to potentiate their enzyme activity and hydrolyze cellulose (Karri et al., 2021). This need of adhesion to the substrate may explain why some key lignocellulolytic taxa can hardly be maintained in stirred bioreactors. Similarly, enzymes can be secreted into the liquid fraction by bacteria, adhered to the substrate or remain firmly attached to their producing bacteria, which are themselves free or attached to the substrate. In order to determine whether bacterial taxa and xylanolytic enzymes were differentially attached to the substrate, xylanase activity and bacterial diversity were observed in different sample fractions: free, weakly attached and strongly attached to the substrate, as detailed in the Materials and methods section (Figure V-6). For the five first days, xylanase activity in the supernatant and pellet fractions showed similar levels (Figure V-6A). Subsequently, xylanase activity doubled in the substrate-bound fraction from about 100 UA on day 5 to about 200 UA on day 15; while in the supernatant, xylanase activity also increased, but to a lesser extent, from about 70 UA on day five to 140 UA at the end of the incubation. The uninoculated control reactor showed a decreasing xylanase activity (from about 140UA to 95UA), shared in relatively equal parts between the supernatant and pellet fractions (about 50 UA in both fractions at the end of incubation).



**Figure V-6 : Xylanase activity and community composition in each sample fraction (free, weakly attached and attached to the substrate). A:** total xylanase activity (UA) and xylanase activity in the supernatant and pellet fractions. Curves are colored by reactor (uninoculated in grey, RCS inoculated reactors 1 & 2 in blue and yellow) and represented smoothed values. **B:** Relative abundances of main bacterial families in the different sample fractions (free, weakly attached and attached), in late (9 and 21 days) sampling points of RCS-inoculated reactors. Asterisks indicate adjusted T-tests significant differences ( $p < 0.05$ , FDR ajustement) between sample fractions for each bacterial family.

Distribution of main bacterial families were compared between the different sample fractions (free, weakly and strongly attached to the substrate, Figure V-6B and Additional file 2 : Figure S6). Due to the small number of fractioned samples sequenced (days 1, 5, 9 and 21), longitudinal PLS analysis could not be performed on fractioned data and because of the strong temporal effects between sample points for adherent fibrolytic bacteria, significance statistical tests were performed using only days 9 and 21 samples. (Figure V-6B). The uninoculated control reactor showed a similar community composition regardless sample fraction (data not show). RCS-inoculated reactors displayed remarkable differences for *Fibrobacteraceae* which was the only

family with significant differences between free and attached fractions. Indeed, while it was at quite low abundance in total communities, it was one of the most abundant attached family, representing more than 10% of attached fraction whereas it was almost not detected (<0.5%) in free or weakly attached fraction. As mentioned earlier, *Fibrobacter* (the only represented genus of the *Fibrobacteraceae* family) is a well-known fibrolytic genus in rumen ecosystems. Its temporal profile in RCS-inoculated bioreactors and its exclusively substrate-bound growth seems to be linked to substrate-bound xylanase activity, and could be linked with the late increased VFA production. *Fibrobacter* thus may play a key role in the capacity of RCS for corn stover bioconversion. Nevertheless, its role in valerate and hexanoate production is not clear as its metabolic capacities linked to these VFA production were not sufficiently described, so it probably acts as an architect member of the bacterial community, allowing other taxa to use its hydrolysis and/or fermentation products. Further metatranscriptomics and/or metaproteomics analysis of corn stover RCS-mediated bioconversion would be required to confirm its role and to gain access to the full repertory of carbohydrate active enzymes and VFA-related proteins produced by *Fibrobacter*, and their temporal dynamics, in the RCS-enriched consortium. Surprisingly, *Ruminococcaeae*, other fiber-adherent family known in cow rumen, is not particularly dedicated here to specific fraction whereas it would be expected to be more abundant in the solid fraction like *Fibrobacteraceae*. Other studies showed that the genus *Ruminococcus* was loosely adherent and was more reported both in the solid and the liquid phase of the rumen rather than attached to fiber particles while *Fibrobacter* was higher in the solid fraction in cow rumen (Klevenhusen et al., 2017; Li et al., 2020b). Additionally, although anaerobic rumen fungi (*Neocallimastigomycota*) were not detected in qPCR amplification (data not shown), it would be interesting to check the presence of other important microorganisms usually found in the rumen such as archaea and protozoa which play an important role in the fermentation of LC biomass.

## 5. Conclusion

Through the deployed enrichment procedure, RCS, an enriched consortium efficiently converted corn stover into VFA was obtained starting from cow rumen. It appeared to be specialized to corn stover as it showed noticeable bacterial composition differences compared to similar rumen derived consortium obtained on different substrates (especially wheat straw). Moreover, it displayed a great stability: it was highly tolerant to cryo-preservation, conserved its VFA production and bacterial diversity in scaled-up bioreactor, and showed good reproducibility in bioconversion dynamics. RCS bacterial diversity included some taxa which were rarely described

in bioreactor context, such as *Fibrobacteraceae*, which displayed remarkable properties: strictly adherent to the substrate, their growth was late but correlated with the late production of specific VFA and with late increase of substrate-attached xylanase activity. All of these elements suggested that *Fibrobacteraceae* could play a key role in RCS bioconversion capacities, but this hypothesis requires metaproteomics confirmation.

**6. Additional files**

**Additional file 1: Supplementary Table S1**

**Tableau V-S1 : Richness and diversity indices (Inverse Simpson and Shannon) of the initial cow rumen inoculum, and throughout the enrichment process.**

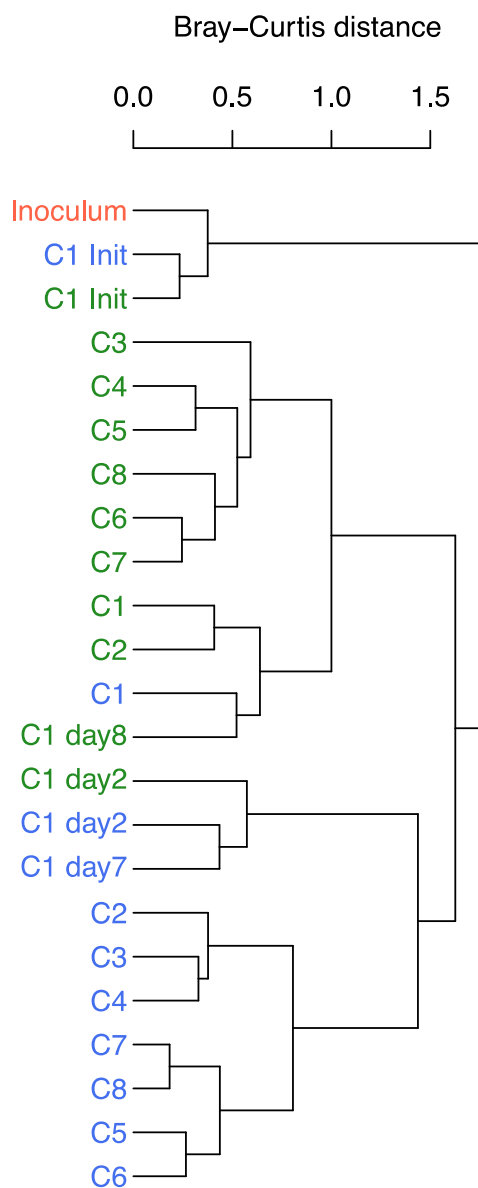
		Cycle												
	<i>Cow rumen</i>	<i>Enrichment</i>	<i>C1- Init</i>	<i>C1- Day2</i>	<i>C1- Day7/8</i>	<i>C1- end</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	<i>Mean±SD C2-&gt;C8</i>
<b>Richness</b>	1495	A	1579	552	607	672	473	470	380	434	380	325	497	422.7±62.7
		B	1512	469	680	676	462	530	555	529	574	561	663	553.4±60.5
<b>Shannon index</b>	6.3	A	6.2	2.8	3.7	4.2	3.9	3.8	3.5	3.3	3.4	3.4	3.6	3.6±0.2
		B	6.3	3.1	4.1	4.2	3.6	3.3	3.9	4.1	4.0	3.9	4.2	3.9±0.3
	161.3	A	102.4	6.7	16.9	31.0	24.9	18.5	17.3	11.3	14.7	15.5	18.5	17.2±4.2

Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

	Cycle													
	<i>Cow rumen</i>	<i>Enrichment</i>	<i>C1- Init</i>	<i>C1- Day2</i>	<i>C1- Day7/8</i>	<i>C1- end</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	<i>Mean±SD C2-&gt;C8</i>
<b>Inv. Simpson index</b>		B	155.6	10.5	25.6	18.2	13.0	7.5	21.9	25.5	23.7	21.1	25.4	19.7±6.9

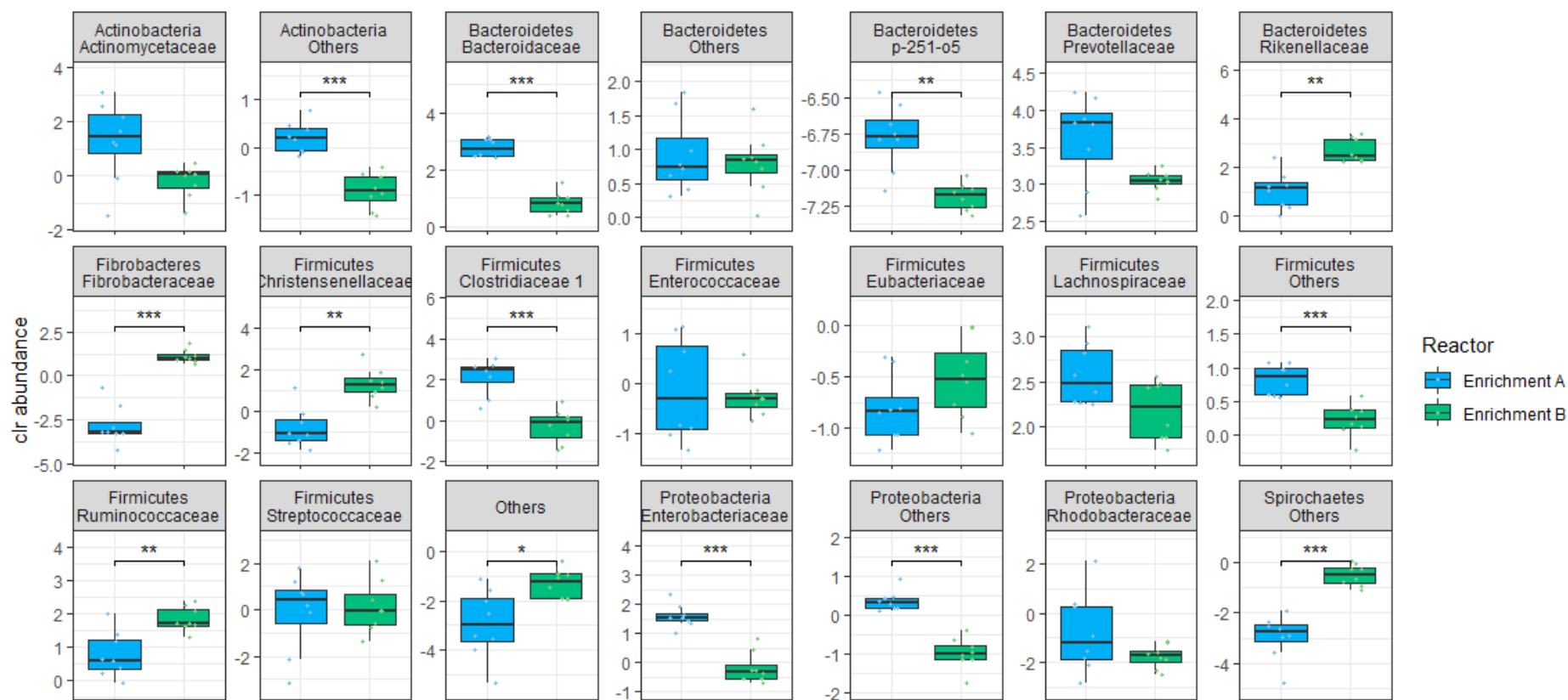
**Additional file 2: Supplementary Figures S1 to S6**

Figure S1



**Figure V-S1 : Hierarchical cluster dendrogram of the cow rumen inoculum (red) and enrichments A (blue) and B (green) microbial communities throughout the enrichment process.** For cycle 1, the initial (day 0), intermediate (day 2 and 7/8) and final cycle measurements are shown. Others cycles are represented by end time points. Clustering based on the Bray-Curtis distance. Scale bar indicates Bray-Curtis distance value between the samples.

Figure S2



**Figure V-S2 : Abundance of major bacterial families (> 5%) at the end of Cycle 1 to Cycle 8 of enrichments A (blue) and B (green).** Relative abundances of OTUs were aggregated at the Family level and CLR transformed. Median values and interquartile ranges are indicated in the plots. Significant difference between enrichment A and B are indicated (Wilcoxon test, FDR-adjusted p-values \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001).



Figure S3

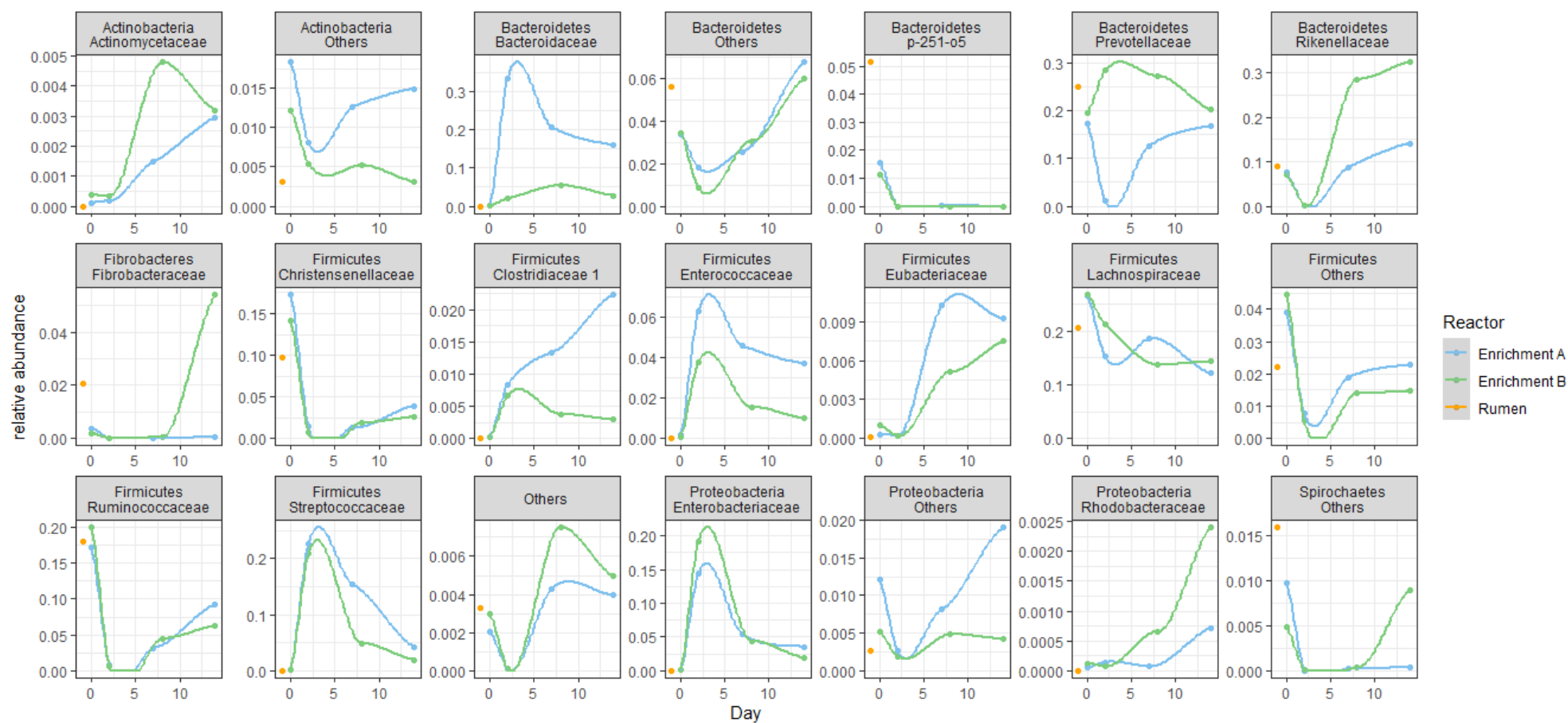
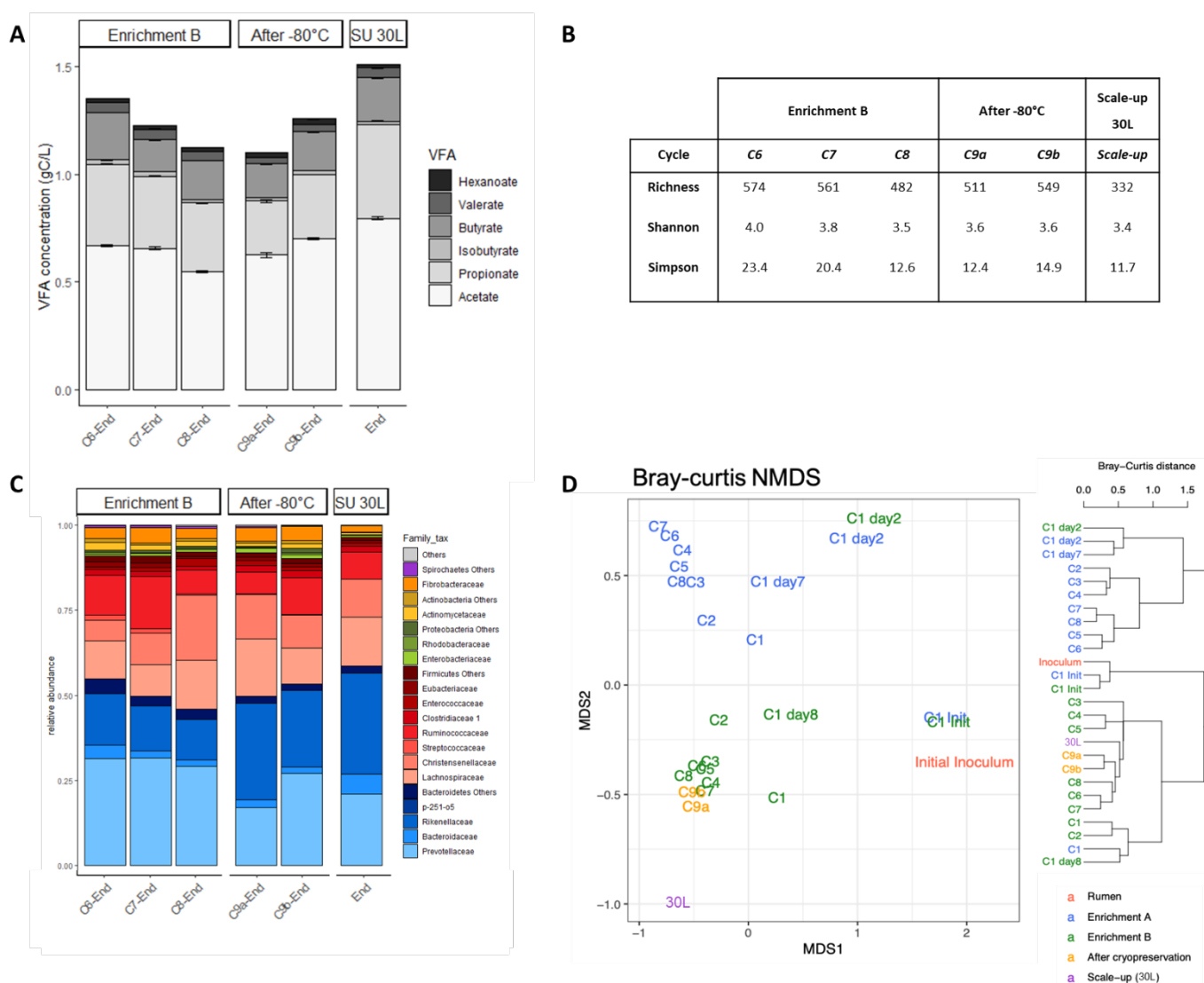


Figure V-S3 : Temporal dynamics of major families (>5%) throughout the 1st cycle (days 0, 2, 7 or 8, 14) of enrichment A (blue) and B (green).

Figure S4



**Figure V-S4 : Impact of cryo-preservation and scaling-up on the lignocellulolytic activity and diversity of RCS consortium. Comparisons of VFA production (A), bacterial diversity (B) and community composition (C) at end points of the three last cycles (from Cycle 6 to Cycle 8) of Enrichment B, and two replicated reactors (C9a and C9b) starting from frozen Cycle 8 end samples. “Others” group consists of Phylum with low relative abundances (< 1%) and “Phylum Others” groups consist of Families with low relative abundances (< 5%) within major Phyla in the dataset. OTUs belonging to the same bacterial Phylum were represented with the same color palette: Bacteroidetes (blue), Firmicutes (red), Proteobacteria (green), Actinobacteria (yellow), Fibrobacteres (orange), Spirochaetes (purple). Each bar corresponds to a time point, annotated with the corresponding day and experiment at the bottom of the graph. No significant difference was obtained between Enrichment, Cryopreserved or scaled-up reactor (Wilcoxon test, FDR adjustment, p-values >0.05 for Shannon and Simpson). D : Non-metric multidimensional scaling (NMDS) and hierarchical cluster dendrogram of bacterial community structures of initial cow rumen (red) duplicate enrichments (Enrichment A in blue and Enrichment B in green), cryo-preservation (orange) and scaling-up of RCS**

**consortium experiments (purple).** The ordination is based on Bray-Curtis coefficient matrix of relative abundances. For cycle 1, the initial (day 0), intermediate (day 2 and 7/8) and final cycle measurements are shown. Others cycles are represented by end time points. Hierarchical clustering based on the Bray-Curtis distance. Scale bar indicates Bray-Curtis distance value between the samples.

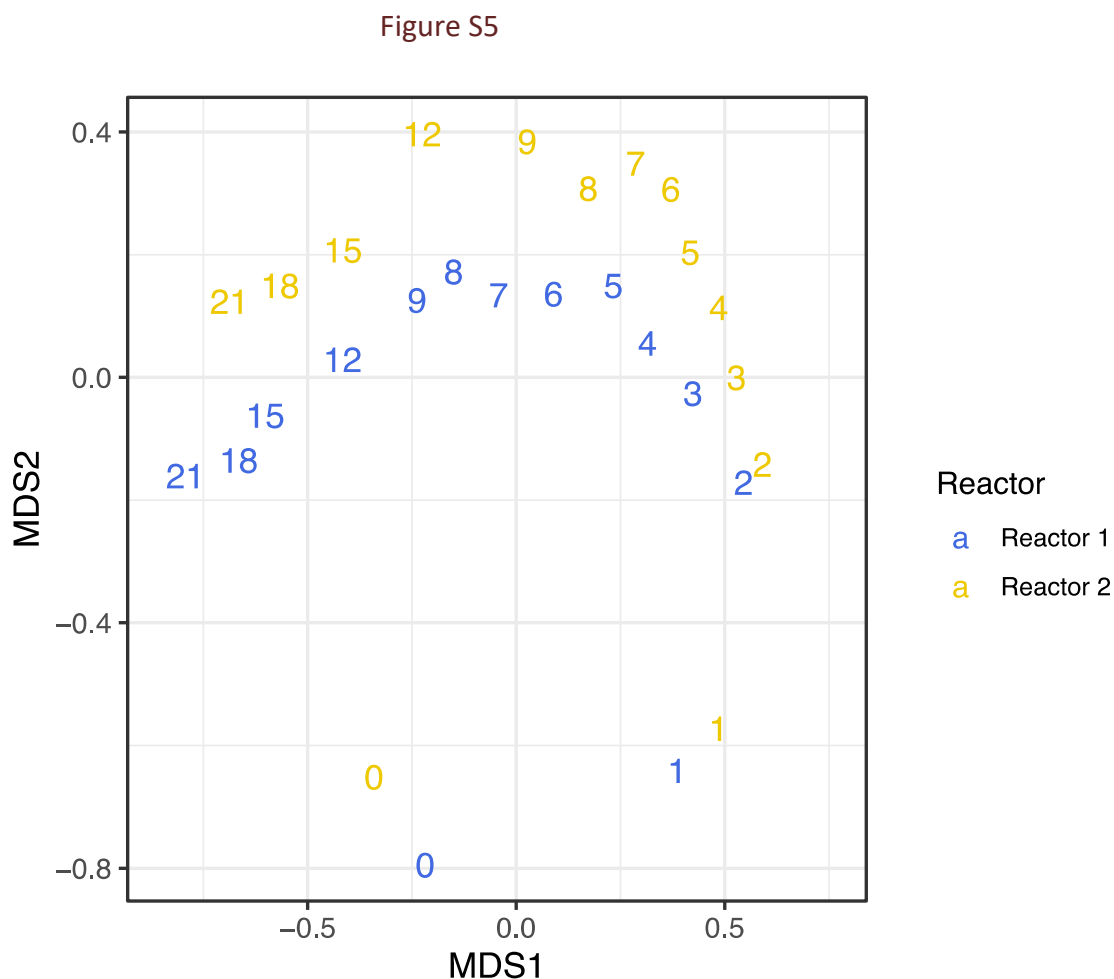


Figure V-S5 : Non-metric multidimensional scaling (NMDS) of bacterial community structures of RCS consortium throughout corn stover bioconversion in duplicate inoculated bioreactors (Reactor 1 in blue, Reactor 2 in yellow). The ordination is based on Bray-Curtis coefficient matrix of relative abundances. The numbers indicate days of incubation.

Figure S6

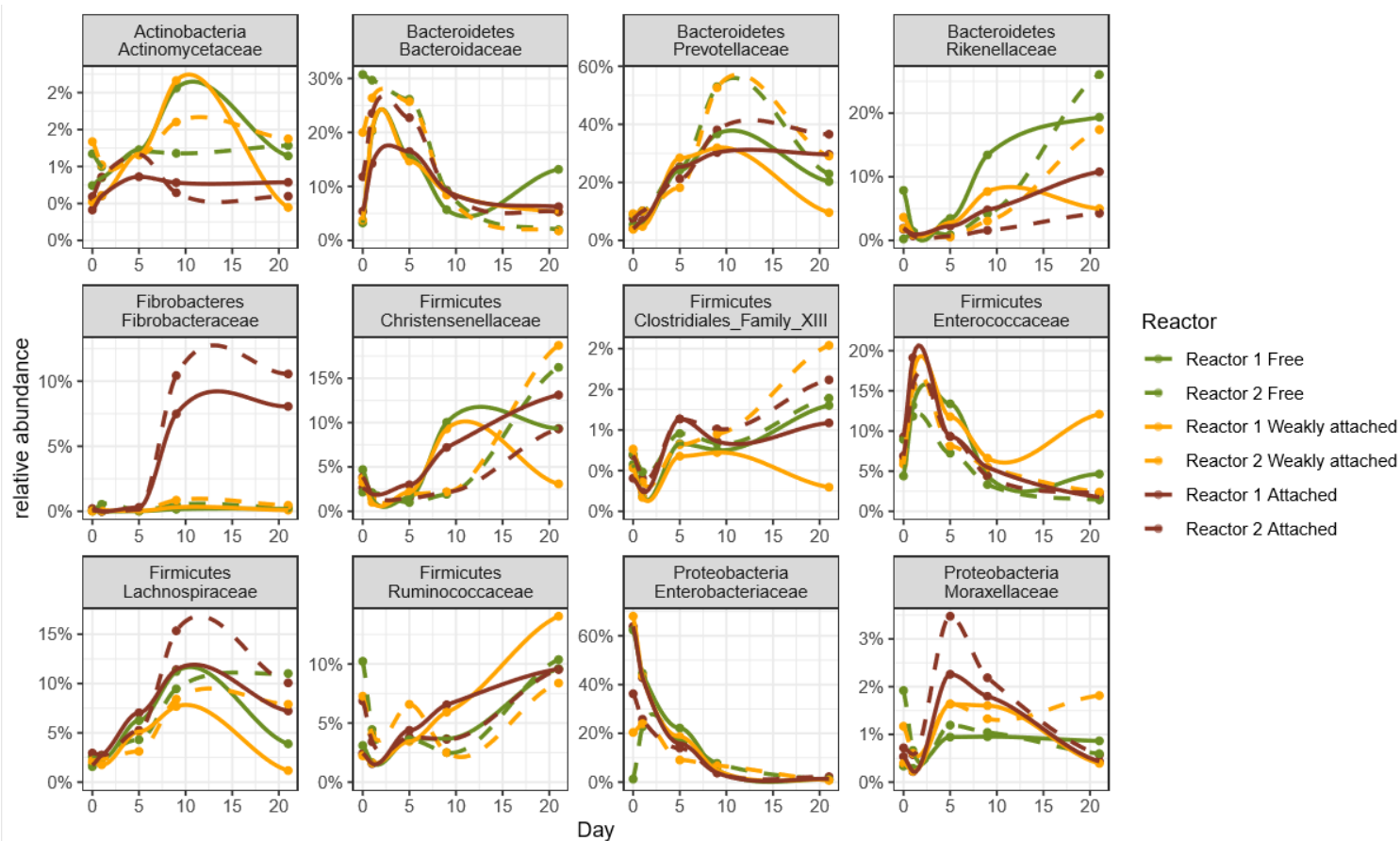


Figure V-S6 : Relative abundances of the main bacterial families throughout corn stover bioconversion in duplicate RCS-inoculated bioreactors (Reactors 1 & 2 in full line and dash respectively) and fraction samples (free (green), weakly attached (orange) and attached (brown) bacteria).

## **7. Acknowledgements**

This research was conducted as part of the Hi-Solids project supported by the French National Research Agency (ANR). The authors thank the Genomics and Transcriptomics (GET) INRAE platform for the 16S sequencing analysis. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources. C. Roy-Camille, A. Lazuka, L. Auer, M. Abadie, M. Bounouba, E. Mengelle and S. Dubos are acknowledged for their assistance and advices with experiments and analytical support.

## **8. Funding**

This work was funded by the French National Research Agency (ANR) under the Project Hi-Solids ANR-14-CE19-0013-01, and the TRANSFORM department of INRAE.

### C. Conclusion

Ces travaux de thèse ont permis de réaliser l'enrichissement en anaérobie d'un consortium microbien appelé RCS (*Rumen Corn Stover*) issu de rumen de vache, à faible teneur de résidus de maïs (2% TS). L'enrichissement a été conduit durant 8 cycles d'une durée variable de 9 à 15 jours avec des prélèvements réguliers permettant de suivre la bioconversion de la lignocellulose et de décrire la dynamique bactérienne au cours de l'enrichissement. La durée des cycles du processus d'enrichissement a été fixée à un minimum de 9 jours en fonction du souhait d'enrichir des taxons tardifs intéressants et présents initialement dans le rumen de vache (Chassard et al., 2005; Mirande et al., 2010). Ces taxons, comme les *Fibrobacteraceae*, montrent un caractère « tardif » lors de la colonisation du substrat dans le rumen de vache et une croissance lente, et sont essentiels à la digestion de la cellulose dans le rumen de vache (Piao et al., 2014). A partir du 2<sup>ème</sup> cycle, un nouveau cycle d'enrichissement démarrait dès lors que la concentration en acides gras volatils atteignait un plateau.

A la fin de cet enrichissement mené en duplicat, deux consortia microbiens ont été obtenus et ont présenté des capacités de production d'acides gras volatils différentes après 14 jours ( $839.2 \pm 24.2 \text{ mgC.L}^{-1}$  pour le A et  $1,200.5 \pm 9.7 \text{ mgC.L}^{-1}$  pour le B). Le profil des AGV produits est également différent entre les deux consortia puisque bien que l'acétate fut l'AGV majoritairement produit (48%) au sein des deux communautés, l'enrichissement A a présenté un ratio propionate : butyrate de 0.96:1 alors qu'il était de 2.1 :1 pour l'enrichissement B. Ces différences en termes de production d'AGV sont visibles dès la fin du 1<sup>er</sup> cycle d'enrichissement et sont conservées à chaque fin des 7 cycles suivants. L'activité xylanase présentait un profil également différent, supérieur dans l'enrichissement B, mais n'a pas été présenté dans la 1<sup>ère</sup> partie de ce chapitre. Grâce à l'analyse des communautés présentes dans l'inoculum de rumen de vache et durant le 1<sup>er</sup> cycle d'enrichissement réalisé en duplicat, nous avons pu observer que la vitesse et la force de sélection sont telles que la présence ou non de certaines espèces dans le consortium final se joue dès les premiers jours du premier cycle, avec une diminution drastique de la richesse. Bien qu'il existe une certaine redondance fonctionnelle entre les espèces au sein d'un consortium (Taxis et al., 2015; Weimer, 2015), nous constatons ici que le procédé d'enrichissement à partir d'un écosystème naturel très riche peut conduire à un potentiel fonctionnel distinct entre les consortia enrichis et que ce potentiel fonctionnel semble être lié à la présence ou non de populations microbiennes. De forts effets aléatoires peuvent expliquer ces observations, et sont déjà visibles avec seulement des dupliquas de bioréacteurs. En s'appuyant

## Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

sur les résultats obtenus ici, une étude pour caractériser ces effets aléatoires et pour sélectionner le meilleur consortium pour une fonction donnée nécessiterait la conduite de nombreux répliquas, mais pourrait se concentrer sur le seul premier cycle où l'essentiel de la sélection se joue et où le potentiel du consortium pour une fonction semble déjà se profiler. Une expérience similaire a été conduite pour déterminer les effets stochastiques dans la mise en place de la structure de la communauté microbienne dans des réacteurs électrochimiques : 14 répliquas, issus du même inoculum et conduits dans des conditions similaires ont engendré des communautés microbiennes différentes avec des capacités distinctes de production de dihydrogène (Zhou et al., 2013).

Le consortium microbien sélectionné pour ses capacités de production d'AGV a été appelé RCS (*Rumen Corn Stover*) et sa communauté se compose principalement de familles de micro-organismes issues des phyla Bacteroidetes et Firmicutes telles que *Bacteroidaceae*, *Prevotellaceae*, *Rikenellaceae* (Bacteroidetes), *Christensenellaceae*, *Enterococcaceae*, *Lachnospiraceae*, *Ruminococcaceae* (Firmicutes) qui sont connues pour être impliquées dans la bioconversion de la lignocellulose en acides gras volatils. Le consortium RCS est surtout défini par la présence d'OTUs de la famille des *Fibrobacteraceae* et du phylum Spirochaetes. La caractérisation fine des dynamiques des populations microbiennes de la communauté RCS au cours de la dégradation de résidus de maïs durant 21 jours a montré des dynamiques différenciées selon les taxons, avec des taxons précoces, tardifs ou d'autres plus constants, et ressemble fortement à celle observée *in situ* dans le rumen de vache à chaque ingestion de switchgrass (Piao et al., 2014). Dans notre étude, certains taxons ont été reliés aux dynamiques de production de certains acides gras volatils par analyse longitudinale. Les acides gras volatils à plus longue chaîne (C5-C6) apparaissent plus tardivement vers les jours 5-6, de façon synchrone avec des taxons tardifs et faiblement abondants, comme les *Fibrobacteraceae*. Ceux-ci, spécifiques de l'enrichissement le plus productif, sont également identifiés dans l'analyse fractionnée de la biomasse, où leur dynamique exclusivement attachée au substrat coïncide aussi avec l'augmentation tardive de l'activité xylanase. Malgré leur croissance lente et leur faible abondance relative, plusieurs indices semblent donc leur donner un rôle central dans la bioconversion, qu'une analyse par approche omique (par métabolomique comme dans les chapitres 3 et 4 ou par métranscriptomique) pourrait confirmer. Des échantillons ont été prélevés et analysés par chromatographie liquide couplée à la spectrométrie de masse en tandem, mais l'analyse bioinformatique n'a pu être réalisée dans le temps imparti de ces travaux de thèse, et pourrait se heurter à leur faible abondance. Il pourrait également être intéressant d'observer par FISH

## Chapitre 5 – Enrichissement, essais de stabilité et de mise à l'échelle d'un consortium microbien dérivé de rumen de vache pour la production de carboxylates à partir de résidus de maïs

(*Fluorescence In Situ Hybridization*) ou par microscopie électronique la dynamique de dégradation des résidus de maïs et le lien entre la présence effective de certaines espèces microbiennes sélectionnées dans les deux consortia enrichis dans un but d'améliorer les connaissances des écosystèmes lignocellulolytiques et leurs adaptations en bioréacteur.

Enfin, le consortium microbien RCS obtenu par enrichissement, présente des caractéristiques très intéressantes pour être un inoculum candidat à la bioconversion de la lignocellulose en carboxylates via des procédés industriels, particulièrement pour l'industrie agroalimentaire. Outre sa robustesse face à la cryo-conservation et à la non-stérilisation du substrat puis sa stabilité face au changement d'échelle du bioréacteur, le consortium sélectionné est davantage spécialisé pour la production d'acétate et de propionate que pour la production de butyrate. A l'heure actuelle, face à l'évolution des modes de vie et des attentes des consommateurs, la demande de conservateurs naturels pour préserver les aliments explose et a des effets directs sur la demande en acides acétique et propionique (Baumann and Westermann, 2016). L'action combinée de prétraitements sur les résidus de maïs, comme le procédé d'extrusion-réactive qui permet de coupler les prétraitements mécanique, chimique et/ou enzymatique, pourrait permettre d'augmenter les capacités de rendement de RCS en AGV à plus longue chaîne ( $C_{\geq 3}$ ), et même de le spécialiser davantage dans la production d'un type d'AGV économiquement plus intéressant que l'acétate (Baumann and Westermann, 2016). Dans une perspective économique et environnementale durable, les procédés de prétraitement et de bioconversion doivent idéalement s'effectuer à de fortes teneurs en solides (>15% TS) afin de réduire les coûts de production et d'augmenter la concentration en produits. Il serait donc important d'étudier l'impact de la teneur en solides sur la structure de la communauté microbienne du consortium RCS et ses capacités de bioconversion.





## VI. A propos de la transformation CLR des données compositionnelles

Les approches omiques engendrent des tables de comptages (nombre de *reads*) ou des tables d'abondances relatives. Ces dernières sont obtenues par une normalisation appelée en anglais *Total Sum Scaling* (TSS), qui permet d'avoir une somme des abondances par échantillon égale à 100% (division de chaque valeur par la somme totale des valeurs de l'échantillon). Les valeurs sont donc liées entre elles et la variation d'une valeur induit une variation des autres. Ces données sont qualifiées de données compositionnelles car elles sont contenues dans un simplexe dont la somme égale 1 (Aitchison, 1982). L'analyse des données compositionnelles comporte de ce fait un risque de « corrélation fallacieuse » (*spurious correlation* en anglais) et de mauvaise interprétation des données. Si cette problématique est encore peu prise en compte dans le domaine de la biologie, il y a une réelle prise de conscience du problème de la compositionnalité des données omiques, particulièrement en écologie microbienne (Weiss et al., 2017; Quinn et al., 2018).

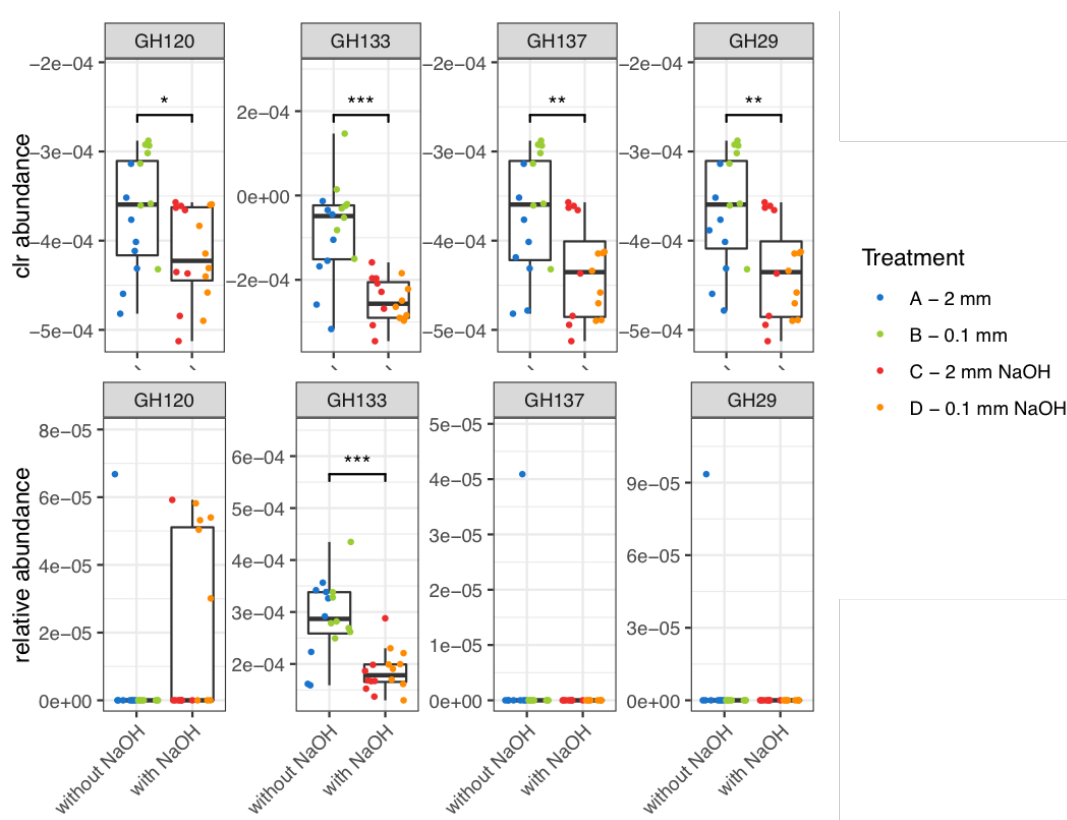
Pour réaliser ces travaux de thèse, la transformation CLR (*Centered Log Ratio*) suggérée par Aitchinson (Aitchison, 1982) et la plus utilisée en écologie microbienne, a été employée pour traiter la problématique des données compositionnelles (ici données de séquençage ARNr 16S et de métaprotéomique). Cette transformation a été discutée dans ce manuscrit, et au vu des interrogations quant à son utilisation et son interprétation, une réflexion plus générale s'impose. En théorie, la transformation CLR est simple : elle consiste, pour une variable donnée (OTU, taxon, protéine ou fonction), à diviser le logarithme de sa valeur (comptage ou abondance relative) par la moyenne géométrique de l'échantillon. Si la méthode est censée atténuer le biais des données compositionnelles, plusieurs manières d'effectuer cette transformation existent et peuvent avoir un impact sur l'interprétation des résultats des analyses.

Tout d'abord, comment gérer la forte occurrence des valeurs nulles dans les données -omiques ? Les données omiques ont une forte occurrence de valeurs nulles, mais toutes n'ont pas la même signification. Une variable (OTU ou protéine) peut prendre une valeur nulle en raison de sa non-détection par la technique utilisée : sa présence est probable mais en dessous du seuil de détection, on parle de zéro d'échantillonnage. Elle peut également prendre une valeur nulle du

fait d'une non-présence réelle dans l'échantillon : on parle alors de zéro structurel s'il distingue des groupes, ou de zéro aberrant (*outlier*) (Kaul et al., 2017). A des fins d'analyses multivariées (factorielles), il est nécessaire de distinguer les zéros structurels des autres valeurs nulles, sans significativité dans l'étude. Des méthodes d'estimation des valeurs nulles existent pour remplacer les zéros, avec leurs avantages et inconvénients (Lubbe et al., 2021). Pour ces travaux de thèse, nous avons obtenu respectivement 33.5% et 72.2% de valeurs nulles dans les données de séquençage 16S et de métaprotéomique du chapitre 3, 76.6% dans le chapitre 4 et 85.6% dans le chapitre 5. Nous avons fait le choix de ne pas utiliser de méthodes permettant d'estimer les zéros car nous avons considéré dans le cas de la métaprotéomique qu'une valeur est réellement nulle si nous n'avons pas détecté la protéine dans au moins 2/3 des réplicats techniques. Dans le cas des données 16S, nous n'avons pas de méthode appropriée pour corriger cette erreur éventuellement induite par la profondeur de séquençage, notamment dans le chapitre 5 où des espèces peuvent réellement disparaître au cours de l'enrichissement ou être considérées comme très rares dans un écosystème aussi riche et complexe qu'est le rumen de vache. Cependant, comme la transformation log interdit les valeurs nulles, il est nécessaire de les éliminer. Dans la littérature, diverses solutions sont employées pour traiter les zéros avant la transformation log : entre autres, le filtrage de chaque variable présentant des zéros (Lovell et al., 2015), l'ajout d'une valeur constante positive (*offset*) à chaque comptage ou à chaque abondance relative (Martín-Fernández et al., 2003; Mandal et al., 2015), l'utilisation d'un modèle de distribution des valeurs ou d'apprentissage profond (*deep learning*) (Templ, 2021). Nous avons choisi d'utiliser la deuxième solution, mais, une nouvelle fois, deux stratégies existent : soit on ajoute à chaque comptage (ou abondance relative) une valeur inférieure entre zéro et la valeur de la limite de détection, ou soit on ajoute 1 à chaque valeur. Ce choix est arbitraire mais il a une incidence sur la valeur de la moyenne géométrique de chaque échantillon (par laquelle les valeurs de chaque échantillon vont être divisées). Ainsi, il a été démontré qu'il avait un impact par exemple, sur les résultats de clustering (Costea et al., 2014). Dans cet article, les auteurs conseillent même d'adapter l'*offset* en fonction du traitement des données, particulièrement dans le cas des données normalisées TSS (comme pour nos données de métaprotéomique qui sont des abondances relatives). Dans ce cas, il semble préférable de définir une valeur bien inférieure à la valeur minimale observée avant transformation. Les analyses exploratoires des données 16S (comptage) et métaprotéomique (relatif) avaient mis le doigt sur cette problématique, puis par soucis de cohérence dans l'analyse des données présentées dans ce manuscrit, nous avons fait le choix de définir l'*offset* à 1 pour toutes les analyses. Il serait intéressant de définir l'*offset* de

l'analyse métabolomique sur une valeur inférieure, de l'ordre de  $10^{-6}$ , afin d'observer son impact sur les conclusions des chapitres 3 et 4. Enfin, le choix d'agréger les variables de faibles abondances dans une catégorie « Others » a permis de réduire fortement le nombre de variables (donc les dimensions des données) et drastiquement le nombre de zéros, ce qui a eu également un impact majeur sur le calcul des moyennes géométriques de chaque échantillon.

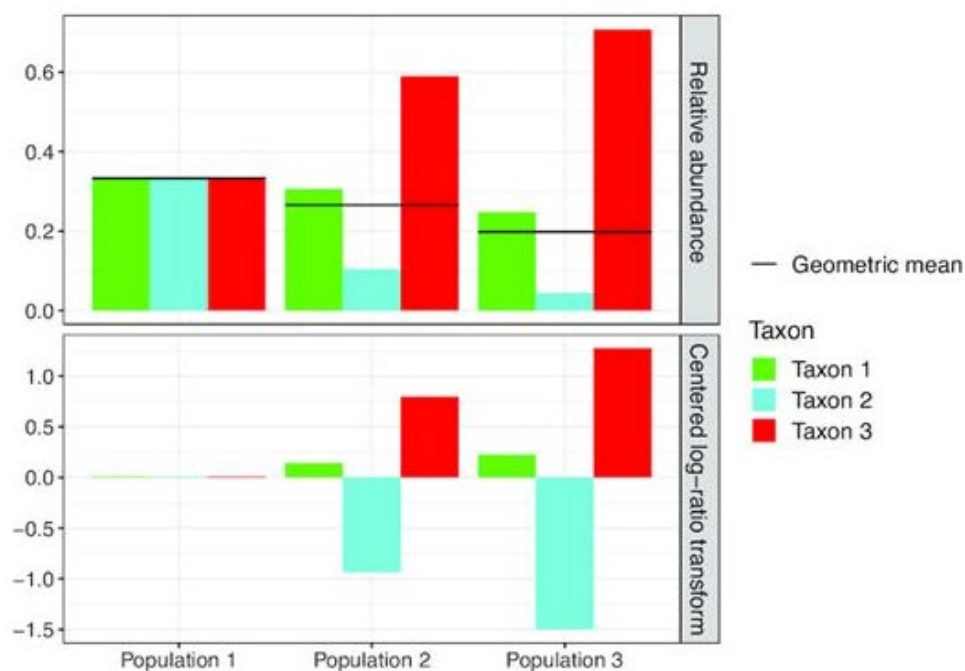
Par ailleurs, les variations de moyenne géométrique peuvent être significatives entre échantillons, ce qui peut parfois induire des conclusions fallacieuses. En effet les abondances relatives nulles sont toutes égales après ajout d'un offset, mais lorsqu'il s'agit de faire la transformation CLR et de les diviser par des moyennes géométriques significativement différentes, elles donnent des valeurs CLR significativement différentes (exemple : GH137 et GH29 dans la Figure VI-1). Quel sens donner à ces différences, statistiquement significatives, alors qu'elles sont basées uniquement sur des valeurs d'abondances nulles ? De plus, on peut observer dans la Figure VI-1, une diminution significative des valeurs CLR entre les traitements A-B et C-D, alors qu'il y a davantage de valeurs d'abondances relatives non nulles (et supérieures) dans les traitements C et D (exemple GH120).



**Figure VI-1 : Limites de la transformation CLR. Exemple de données issues du Chapitre 4 lors de la comparaison entre substats prétraités (A-B vs C-D) de l'expression de quelques familles CAZy. Les valeurs transformées CLR et les abondances relatives (NSAF) sont respectivement**

représentées en haut et en bas. Les comparaisons sans NaOH / avec NaOH significativement différentes sont indiquées par des astérisques (test de Wilcoxon, ajustement FDR).

Ce problème d'interprétation des données CLR a déjà été soulevé par d'autres auteurs, statisticiens et spécialistes des données compositionnelles (Hawinkel et al., 2020). Dans l'exemple théorique présenté ci-après (Figure VI-2), les auteurs font remarquer que lorsque la distribution des abondances des taxons s'écarte d'une distribution homogène, la moyenne géométrique diminue (de gauche à droite sur la figure). Néanmoins, comme l'abondance du taxon 1 (en vert) diminue, mais de façon moins importante que la moyenne géométrique, le taxon 1 apparaît augmenter en valeur CLR, ce qui peut conduire à une interprétation confuse.



**Figure VI-2: Limites de la transformation CLR. Exemple théorique illustrant l'effet de la compositionnalité et des différences de moyenne géométrique sur les interprétations de données CLR.** Figure issue des travaux de Hawinkel et al., (2020).

Au vu des difficultés rencontrées lors de ces travaux et dans les quelques articles spécialisés qui abordent ses limites, la transformation CLR ne semble remplir que très imparfaitement son rôle. Si la « correction » du biais de compositionnalité des données est essentielle, d'autres méthodes publiées récemment, semblent prometteuses : la normalisation CSS (*Cumulative Sum Scaling*

*normalisation*), plutôt recommandée pour l'analyse de données transcriptomiques, consiste à diviser les comptages par la somme cumulative des comptages jusqu'à un centile déterminé à partir des données (Paulson et al., 2013). Cette normalisation fait l'hypothèse que la majorité des observations ne sont pas différentiellement représentées entre échantillons et que la fraction stable sert de référence pour normaliser les échantillons. Elle semble donc peut adaptée à l'étude des dynamiques de communautés microbiennes par métabarcoding ciblée au cours d'un enrichissement par exemple.



## Conclusions générales et perspectives

Au cours de cette thèse, j'ai principalement utilisé deux stratégies d'analyses métagénomiques, la métagénomique amplicon et la métagénomique *shotgun label-free*, appliquées aux consortia microbiens lignocellulolytiques. Alors que ces approches rapportent habituellement le résultat d'une « photo » à un instant donné, souvent en point final, ici nous nous sommes attachés à explorer la communauté dans son ensemble au cours des différentes étapes de la dynamique de bioconversion de la lignocellulose. L'étude des omiques par analyses multivariées et leur intégration avec les données macrocinétiques par analyse longitudinale ont ainsi permis de mieux comprendre les mécanismes déployés par les communautés lignocellulolytiques une fois stables et enrichies mais aussi lorsqu'elles font face à des perturbations (cryo-conservation, mise à l'échelle, prétraitement du substrat).

### Etude fonctionnelle des communautés microbiennes par approche métagénomique

Les dynamiques de diversité obtenues par métagénomique amplicon ciblé (séquençage du gène codant l'ARNr 16S) ont beaucoup apporté à l'écologie microbienne, surtout pour des écosystèmes complexes difficilement cultivables comme le rumen de vache, afin de connaître les espèces présentes au sein d'une communauté. Cependant, leur intérêt pour la caractérisation fonctionnelle des communautés reste limité. Les méthodes permettant de prédire le potentiel fonctionnel à partir de données de séquençage 16S, telles que PICRUSt (Douglas et al., 2018) et Tax4Fun (Aßhauer et al., 2015), se sont développées ces dernières années. Bien qu'elles présentent des avantages pour prédire le potentiel génomique d'une communauté microbienne, elles ne sont pas très adaptées pour l'étude des dynamiques fonctionnelles temporelles au cours de la dégradation de la lignocellulose où les expressions de gènes diffèrent au cours du temps. Ces méthodes sont également limitées par les bases de données puisqu'il est nécessaire que les génomes des organismes y soient référencés. Ainsi, l'étude fine des dynamiques fonctionnelles requière *a minima* une approche métagénomique ou, comme présenté ici, une approche métagénomique. La métagénomique est apparue comme un excellent outil d'étude du fonctionnement *in vivo* ou *in situ* des communautés pour la bioconversion de la lignocellulose (D'haeseleer et al., 2013; Zhu et al., 2016; Deusch et al., 2017; Speda et al., 2017) en répondant à trois questions majeures : qui ? fait quoi ? et quand ? Nous avons montré que dans le cas de l'étude de consortia microbiens lignocellulolytiques en bioréacteurs (avec des diversités faibles



par rapport à des écosystèmes naturels), des résultats intéressants et robustes pouvaient être obtenus par métaprotéomique *shotgun* sans marquage pour caractériser le profil hydrolytique (CAZymes) et les enzymes impliquées dans la production des AGV chez deux consortia distincts enrichis sur paille de blé au cours des phases de dégradation, mais aussi l'impact de prétraitements du substrat sur la communauté structurale, la régulation de l'expression des CAZymes et les capacités de bioconversion. Néanmoins, la technique présente des verrous technologiques non négligeables : (i) les protocoles d'extraction et de purification des protéines sont lourds et limitent le nombre d'échantillons dans l'analyse, (ii) l'identification des spectres et l'annotation des protéines reposent sur les séquences référencées dans les bases de données, (iii) la profondeur d'analyse n'est pas suffisante pour une analyse exhaustive des protéines des populations moins abondantes. Afin d'améliorer notre analyse métaprotéomique et dans la perspective de l'utiliser dans la suite du projet Hi-Solids pour caractériser l'impact du prétraitement par extrusion-réactive ou de la teneur en solides sur le consortium RCS, la réalisation du métagénome de la communauté enrichie RCS pourrait être réalisée en amont et *a minima*, il serait nécessaire de réduire la taille de la base de données UniprotKB/TrEMBL aux seules espèces détectées par métabarcoding 16S. En effet, les analyses métaprotéomiques présentées dans ce manuscrit ont été réalisées avec la version de la base de données UniprotKB/TrEMBL d'avril 2016 qui présentait plus de 40 millions de séquences, soit presque 4 fois moins que la version actuelle (juin 2021) qui contient 151 millions de séquences bactériennes. Des ajustements seront aussi à prévoir au niveau du pipeline d'annotation Prophaner car une nouvelle version existe depuis peu avec l'implémentation de diamond à la place de BlastP pour réaliser l'annotation taxonomique. Outre ces changements de méthodologie, on pourrait approfondir les premiers résultats observés dans le chapitre 5 et faire une analyse métaprotéomique fractionnée (libre, faiblement attachée et attachée au substrat) pour mieux comprendre les mécanismes hydrolytiques dans chaque fraction du consortium enrichi RCS puisque certaines populations les distinguent.

### **Complémentarité fonctionnelle au sein des écosystèmes lignocellulolytiques bactériens**

La complémentarité fonctionnelle supposée dans les travaux de thèse d'A. Lazuka et de L. Auer entre représentants des phyla Bacteroidetes et Firmicutes (sur la base des données de séquençage de l'ADNr 16S) a été confirmée en métaprotéomique par l'expression des protéines au cours de la

bioconversion de la lignocellulose, principalement par « un partage des tâches » entre les genres *Bacteroidetes* et *Firmicutes*, et particulièrement entre *Bacteroides* et *Clostridium*. Les travaux de métaprotéomique comparative appliquée à l'étude de l'effet de prétraitements sur le métaprotéome présentés dans le Chapitre 4 ont fourni des résultats inédits au niveau des abondances protéiques concernant la probable régulation de l'expression génique en réponse à des modifications physico-chimiques du substrat. En effet, une régulation spécifique de l'expression de familles CAZymes a pu être mise en évidence, plus particulièrement chez certaines familles CAZy connues pour leurs activités arabinoxylanolytiques et produites par les *Bacteroides*. Ces observations ont été mises en lien avec la structure particulière des génomes des *Bacteroidetes*, organisés en clusters de gènes dédiés à un type de sucre (PULs) et régulés de façon très spécifique par et pour la dégradation d'un type de polysaccharide. Nous avons fait l'hypothèse que le prétraitement alcalin favorisait la libération d'une forte concentration de motifs d'arabinoxylans qui activeraient les PULs dédiés à l'hydrolyse de ce polymère et coderaient toute la machinerie enzymatique nécessaire. Concernant l'étude des voies métaboliques de production des AGV par métaprotéomique, l'approche confirme à large échelle un impact du substrat au travers d'un changement métabolique de production d'acétate au profit du butyrate mais ne montre pas d'effet de régulation des producteurs. Nous avons fait l'hypothèse que l'hydrolyse enzymatique des arabinoxylans par les espèces saccharolytiques favorisait la fermentation des pentoses en butyrate chez d'autres espèces.

Aussi, lors de l'enrichissement du consortium microbien RCS, nous avons observé un potentiel fonctionnel distinct entre les consortia enrichis qui semblait être lié à la présence ou non de certaines populations microbiennes au sein des consortia, particulièrement les espèces fibrolytiques adhérentes telles que les *Fibrobacteraceae* qui possèdent une stratégie unique d'adhésion au substrat et d'hydrolyse de la cellulose. Ces taxons ont aussi été reliés au consortium présentant les meilleures capacités de production d'AGV mais aussi aux dynamiques de production des acides gras volatils à plus longue chaîne comme le valérate ou le caproate par analyse longitudinale. Il existe là encore une réelle complémentarité fonctionnelle au sein du consortium puisqu'on peut supposer que les *Fibrobacteraceae*, en hydrolysant la cellulose en formate, acétate et succinate permettent à d'autres espèces du consortium de produire davantage d'AGV, dont du valérate et du caproate. En plus d'être l'étape limitante de la digestion anaérobie, les mécanismes d'hydrolyse et ses produits générés ont un réel impact sur la

bioconversion de la lignocellulose en acides gras volatils et reposent sur la complémentarité fonctionnelle entre espèces.

Il est à noter dans le résultat de ces études que les familles CAZymes connues pour leurs activités hémicellulolytiques semblent centrales dans l'hydrolyse de la lignocellulose et apparaissent comme significativement impactées par les stratégies d'optimisation de production de carboxylates à de nombreux niveaux d'analyses, mais les familles associées à d'autres activités enzymatiques (cellulases par exemple) sortent plus rarement significatives et sont aussi pratiquement indétectables en dosage colorimétrique enzymatique (CMCase) alors que l'hydrolyse de la cellulose est constatée par mesure de la fraction cellulosique au sein du substrat prélevé. Plusieurs explications sont possibles à cette observation :

- La diversité d'enzymes nécessaire à l'hydrolyse des hémicelluloses est beaucoup plus importante que pour la cellulose ;
- L'adsorption des enzymes type cellulases au substrat ;
- L'activité par unité d'enzyme pourrait être beaucoup plus forte pour les cellulases que pour les hémicellulases ;
- La « faible » activité cellulase nécessaire à l'hydrolyse pourrait être non limitante dans le système : si l'hydrolyse de la cellulose est limitée en raison de sa faible accessibilité (cristallinité ou accessibilité limitée par les hémicelluloses et la lignine), les activités cellulases pourraient être considérées en excès et ne pas être impactées par les différents paramètres d'optimisation de l'hydrolyse (choix de l'inoculum ou prétraitement du substrat).

### **Enrichissement d'un consortium microbien dérivé de rumen de vache pour la bioconversion de résidus de maïs en acides gras volatils dans le cadre du projet Hi-Solids**

L'utilisation d'un nouveau substrat modèle dans le projet Hi-Solids nécessitait l'obtention d'un nouvel inoculum, adapté et enrichi spécifiquement sur des résidus de maïs. Le rumen bovin ayant fait ses preuves lors de l'enrichissement d'un consortium sur paille de blé en bioreacteurs

discontinus séquentiels (Lazuka et al., 2015), c'est à partir de cette source d'inoculum et dans des conditions expérimentales similaires qu'a été obtenu un nouveau consortium microbien enrichi et présentant une certaine robustesse et stabilité face à la congélation et à un essai de mise à l'échelle. Des prélèvements réguliers au cours des cycles d'enrichissement ont permis d'étudier les dynamiques de la diversité microbienne et de production d'AGV au cours de celui-ci et d'identifier de potentiels forts effets aléatoires pendant le premier cycle d'enrichissement. En effet, l'enrichissement d'un consortium microbien issu d'un écosystème naturel aussi riche, complexe et diversifié que le rumen de vache sur un substrat spécifique dans des conditions contrôlées similaires implique des perturbations de la structure de la communauté initiale, incluant une sélection positive comme négative de certaines espèces présentes initialement. Bien qu'il existe une certaine redondance fonctionnelle au sein des écosystèmes naturels, l'absence de certaines espèces, colonisatrices dites « primaires », au cours de l'assemblage de la communauté et de la colonisation du substrat peut avoir des conséquences sur la présence ou non de colonisateurs dits « secondaires », et un impact conséquent sur le potentiel fonctionnel du consortium enrichi. Pour pallier à ces effets et dans un but de sélectionner le consortium présentant les meilleures capacités de bioconversion du substrat, il serait important de réaliser plusieurs réplicats de bioréacteurs lors du 1<sup>er</sup> cycle d'enrichissement. Cette expérience pourrait également servir à mieux comprendre les mécanismes d'adaptation et d'assemblage mis en jeu au cours de la colonisation du substrat en bioréacteur par un inoculum lignocellulolytique complexe et de déterminer les colonisateurs « primaires » et « secondaires », essentiels dans un consortium pour une fonction souhaitée, par une approche de réseaux d'associations entre espèces microbiennes. Ici, l'analyse de la diversité 16S a pu mettre en évidence l'importance d'enrichir des colonisateurs dits « secondaires » (définis aussi comme essentiels dans l'hydrolyse du substrat dans le rumen de vache) : les *Fibrobacteraceae*, qui apparaissent comme caractéristiques de la fraction attachée au substrat d'un consortium présentant la meilleure capacité de production d'AGV et semblent avoir un rôle clé dans la bioconversion de résidus de maïs, malgré leur faible abondance relative. Les résultats de l'analyse métabolomique initiée pourront être intégrés aux résultats déjà obtenus par analyse longitudinale et pourront confirmer l'hypothèse de leur rôle clé au sein de ce consortium.

### **Perspectives du projet Hi-Solids**

D'autres expérimentations ont été menées dans le cadre du projet Hi-Solids mais les résultats n'ont pas été présentés ici. Parmi eux, des cinétiques de bioconversion à faible teneur en solides ont été réalisés pour examiner huit substrats de résidus de maïs prétraités ou non par extrusion-réactive (avec le consortium enrichi RCS). Les premiers résultats de ces expériences ont montré une amélioration significative de la production d'AGV dans le seul cas du prétraitement combinant la pré-hydrolyse du substrat par une xylanase commerciale et l'extrusion (1.3 g/L d'AGV pour le substrat brut contre 3.3 g/L d'AGV pour le substrat extrudé avec une xylanase commerciale). Un changement du ratio des AGV produits a également été constaté avec davantage de butyrate produit et moins d'acétate et de propionate. Des résultats similaires ont également été décrits à forte teneur en solides (15% TS) par les collaborateurs du projet (Marone et al., 2019) et la condition de prétraitement (pré-hydrolyse par une xylanase commerciale puis extrusion) a été choisie pour la suite du projet.

Un important travail expérimental a ensuite été mené pour la conduite de cinétiques de bioconversion de résidus de maïs prétraités par la condition d'extrusion-réactive sélectionnée avec le consortium enrichi RCS à faible (2%) et à haute (15%) teneur en solides. Des bioréacteurs inoculés en dupliquas et un réacteur contrôle non inoculé ont été conduits pour chaque condition de concentration de substrat (2 et 15% TS), et plusieurs prélèvements ont été réalisés pour analyser l'impact du prétraitement et de la teneur en solides sur les capacités de bioconversion du consortium à un niveau macrocinétique (AGV, dégradation du substrat) mais aussi à l'échelle de la communauté (colonisation du substrat par FISH, métabarcoding 16S et métabotéomique). Ce travail a été précédé d'un conséquent travail de conception et de construction de deux bioréacteurs capables d'agiter de façon homogène à cette condition de teneur en solides (15% TS), et la mise au point d'un nouveau système de prélèvement.

Les résultats préliminaires sont prometteurs et intéressants : les rendements de bioconversion sont nettement supérieurs avec cette condition de prétraitement à faible teneur en solides et sont légèrement inférieurs à forte teneur en solides. Cette baisse de rendement à forte teneur en solides pourrait s'expliquer par une inadaptation du consortium RCS et de nouvelles pressions de sélection conséquentes exercées sur l'inoculum puisque ce consortium est adapté à la bioconversion à faible teneur en solide (la réalisation d'un enrichissement à haute teneur en solide avait un temps été envisagée, mais a été écartée car irréalisable dans le cadre de cette thèse). Les

premières analyses de la diversité 16S montrent respectivement une présence et une absence des *Fibrobacteraceae* dans la communauté à 21 jours dans le cadre des cinétiques à 2 et 15% de solides. Au vu des éléments discutés au cours de l'enrichissement et la caractérisation de RCS, cette absence, étonnante car on s'attendait à ce que cette teneur favorise les espèces fibrolytiques, pourrait expliquer un rendement plus faible. De plus, on constate également qu'avec l'augmentation de la teneur en solides, la proportion du butyrate augmente nettement au sein du ratio des AGV produits. Cette observation a également été faite par les collaborateurs du projet lors d'essai à différentes teneur en solides. Une des hypothèses serait que l'eau libre étant moins disponible pour les bactéries au sein du bioréacteur, les bactéries favoriseraient la fermentation butyrique et alcoolique qui ne requiert pas d'eau à la place de l'acidogénèse qui nécessite de l'eau. Les nouvelles données du projet Hi-Solids, obtenues par approches omiques apporteront encore de nouveaux éléments de compréhension des mécanismes mis en jeu par les communautés lignocellulolytiques en bioréacteurs en situation stable et face à des perturbations extérieures (prétraitement, teneur en solides) en vue de les piloter pour une utilisation à l'échelle industrielle.



## Références bibliographiques

Abdallah Ismail, N., Ragab, S.H., Abd ElBaky, A., Shoeib, A.R.S., Alhosary, Y., and Fekry, D. (2011). Frequency of Firmicutes and Bacteroidetes in gut microbiota in obese and normal weight Egyptian children and adults. *Arch. Med. Sci. AMS* 7, 501–507.

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A., and Levin, D.B. (2011). Biomass pretreatment: Fundamentals toward application. *Biotechnol. Adv.* 29, 675–685.

Aghapour Aktij, S., Zirehpour, A., Mollahosseini, A., Taherzadeh, M.J., Tiraferri, A., and Rahimpour, A. (2020). Feasibility of membrane processes for the recovery and purification of bio-based volatile fatty acids: A comprehensive review. *J. Ind. Eng. Chem.* 81, 24–40.

Agler, M.T., Wrenn, B.A., Zinder, S.H., and Angenent, L.T. (2011). Waste to bioproduct conversion with undefined mixed cultures: the carboxylate platform. *Trends Biotechnol.* 29, 70–78.

Agler, M.T., Spirito, C.M., Usack, J.G., Werner, J.J., and Angenent, L.T. (2012). Chain elongation with reactor microbiomes: upgrading dilute ethanol to medium-chain carboxylates. *Energy Environ. Sci.* 5, 8189–8192.

Agrawal, D., Tsang, A., and Chadha, B.S. (2021). Economizing the lignocellulosic hydrolysis process using heterologously expressed auxiliary enzymes feruloyl esterase D (CE1) and  $\beta$ -xylosidase (GH43) derived from thermophilic fungi *Scytalidium thermophilum*. *Bioresour. Technol.* 339, 125603.

Agua, I. del, Usack, J.G., and Angenent, L.T. (2015). Comparison of semi-batch vs. continuously fed anaerobic bioreactors for the treatment of a high-strength, solids-rich pumpkin-processing wastewater. *Environ. Technol.* 36, 1974–1983.

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B Methodol.* 44, 139–177.

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2000). Logratio Analysis and Compositional Distance. *Math. Geol.* 32, 271–275.

Alessi, A.M., Bird, S.M., Bennett, J.P., Oates, N.C., Li, Y., Dowle, A.A., Polikarpov, I., Young, J.P.W., McQueen-Mason, S.J., and Bruce, N.C. (2017). Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. *Sci. Rep.* 7, 2356.

Alessi, A.M., Bird, S.M., Oates, N.C., Li, Y., Dowle, A.A., Novotny, E.H., deAzevedo, E.R., Bennett, J.P., Polikarpov, I., Young, J.P.W., et al. (2018). Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. *Biotechnol. Biofuels* 11, 166.



- Allgaier, M., Reddy, A., Park, J.I., Ivanova, N., D'haeseleer, P., Lowry, S., Sapra, R., Hazen, T.C., Simmons, B.A., VanderGheynst, J.S., et al. (2010). Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. *PLoS ONE* 5.
- Alvira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M.J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresour. Technol.* 101, 4851–4861.
- Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46.
- Anwar, Z., Gulfranz, M., and Irshad, M. (2014). Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: A brief review. *J. Radiat. Res. Appl. Sci.* 7, 163–173.
- Armengaud, J., Hartmann, E.M., and Bland, C. (2013). Proteogenomics for environmental microbiology. *PROTEOMICS* 13, 2731–2742.
- Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.* 71, 7724–7736.
- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12, 186.
- Aßhauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882–2884.
- Auer, L., Lazuka, A., Sillam-Dussès, D., Miambi, E., O'Donohue, M., and Hernandez-Raquet, G. (2017). Uncovering the Potential of Termite Gut Microbiome for Lignocellulose Bioconversion in Anaerobic Batch Bioreactors. *Front. Microbiol.* 8, 2623.
- Aydin, S., Yesil, H., and Tugtas, A.E. (2018). Recovery of mixed volatile fatty acids from anaerobically fermented organic wastes by vapor permeation membrane contactors. *Bioresour. Technol.* 250, 548–555.
- Baba, Y., Tada, C., Fukuda, Y., and Nakai, Y. (2013). Improvement of methane production from waste paper by pretreatment with rumen fluid. *Bioresour. Technol.* 128, 94–99.
- Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D.T., et al. (2017). Galactomannan Catabolism Conferred by a Polysaccharide Utilization Locus of *Bacteroides ovatus*. *J. Biol. Chem.* 292, 229–243.
- Baker, G.C., Smith, J.J., and Cowan, D.A. (2003). Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* 55, 541–555.
- Banerjee, G., Car, S., Scott-Craig, J.S., Hodge, D.B., and Walton, J.D. (2011). Alkaline peroxide pretreatment of corn stover: effects of biomass, peroxide, and enzyme loading and composition on yields of glucose and xylose. *Biotechnol. Biofuels* 4, 16.
- Barakat, A., de Vries, H., and Rouau, X. (2013). Dry fractionation process as an important step in current and future lignocellulose biorefineries: A review. *Bioresour. Technol.* 134, 362–373.

- Barakat, A., Chuetor, S., Monlau, F., Solhy, A., and Rouau, X. (2014a). Eco-friendly dry chemo-mechanical pretreatments of lignocellulosic biomass: Impact on energy and yield of the enzymatic hydrolysis. *Appl. Energy* *113*, 97–105.
- Barakat, A., Mayer-Laigle, C., Solhy, A., Arancon, R.A.D., Vries, H. de, and Luque, R. (2014b). Mechanical pretreatments of lignocellulosic biomass: towards facile and environmentally sound technologies for biofuels production. *RSC Adv.* *4*, 48109–48127.
- Bastidas-Oyanedel, J.-R., and Schmidt, J.E. (2018). Increasing Profits in Food Waste Biorefinery—A Techno-Economic Analysis. *Energies* *11*, 1551.
- Baumann, I., and Westermann, P. (2016). Microbial Production of Short Chain Fatty Acids from Lignocellulosic Biomass: Current Processes and Market. *BioMed Res. Int.* *2016*, e8469357.
- Bayané, A., and Guiot, S.R. (2011). Animal digestive strategies versus anaerobic digestion bioprocesses for biogas production from lignocellulosic biomass. *Rev. Environ. Sci. Biotechnol.* *10*, 43–62.
- Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* *8*, 548–557.
- Bayer, E.A., Lamed, R., White, B.A., and Flint, H.J. (2008). From cellulosomes to cellulosomes. *Chem. Rec. N. Y. N* *8*, 364–377.
- Bayer, E.A., Shoham, Y., and Lamed, R. (2013). Lignocellulose-Decomposing Bacteria and Their Enzyme Systems. In *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Berlin, Heidelberg: Springer), pp. 215–266.
- Béraud, P. (2013). La (méta)génomique des microorganismes du rumen et ses applications à la production des ruminants.
- Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, *Bacteroides thetaiotaomicron*, to the Suckling Period \*. *J. Biol. Chem.* *281*, 36269–36279.
- Blouzard, J.-C., Coutinho, P.M., Fierobe, H.-P., Henrissat, B., Lignon, S., Tardif, C., Pagès, S., and de Philip, P. (2010). Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* *10*, 541–554.
- Bodein, A., Chapleur, O., Droit, A., and Lê Cao, K.-A. (2019). A Generic Multivariate Framework for the Integration of Microbiome Longitudinal Studies With Other Data Types. *Front. Genet.* *10*, 963.
- Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., and Caporaso, J.G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* *10*, 57–59.

- Braga, C.M.P., Delabona, P. da S., Lima, D.J. da S., Paixão, D.A.A., Pradella, J.G. da C., and Farinas, C.S. (2014). Addition of feruloyl esterase and xylanase produced on-site improves sugarcane bagasse hydrolysis. *Bioresour. Technol.* *170*, 316–324.
- Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 1948–1953.
- Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* *12*, 168–180.
- Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., and Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol.* *184*, 449–456.
- Bugg, T.D., Ahmad, M., Hardiman, E.M., and Singh, R. (2011). The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol.* *22*, 394–400.
- Burnum, K.E., Callister, S.J., Nicora, C.D., Purvine, S.O., Hugenholtz, P., Warnecke, F., Scheffrahn, R.H., Smith, R.D., and Lipton, M.S. (2011). Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *ISME J.* *5*, 161–164.
- Calle, M.L. (2019). Statistical Analysis of Metagenomics Data. *Genomics Inform.* *17*.
- Calusinska, M., Marynowska, M., Bertucci, M., Untereiner, B., Klimek, D., Goux, X., Sillam-Dussès, D., Gawron, P., Halder, R., Wilmes, P., et al. (2020). Integrative omics analysis of the termite gut system adaptation to *Miscanthus* diet identifies lignocellulose degradation enzymes. *Commun. Biol.* *3*.
- Cao, K.-A.L., Costello, M.-E., Lakis, V.A., Bartolo, F., Chua, X.-Y., Brazeilles, R., and Rondeau, P. (2016). MixMC: A Multivariate Statistical Framework to Gain Insight into Microbial Communities. *PLOS ONE* *11*, e0160169.
- Cardona, M.J., Tozzi, E.J., Karuna, N., Jeoh, T., Powell, R.L., and McCarthy, M.J. (2015). A process for energy-efficient high-solids fed-batch enzymatic liquefaction of cellulosic biomass. *Bioresour. Technol.* *198*, 488–496.
- Carrard, G., Koivula, A., Söderlund, H., and Béguin, P. (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci.* *97*, 10342–10347.
- Cavaillé, L., Grousseau, E., Pocquet, M., Lepeuple, A.-S., Uribelarrea, J.-L., Hernandez-Raquet, G., and Paul, E. (2013). Polyhydroxybutyrate production by direct use of waste activated sludge in phosphorus-limited fed-batch culture. *Bioresour. Technol.* *149*, 301–309.
- Chandel, A.K., and Singh, O.V. (2011). Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of “Biofuel.” *Appl. Microbiol. Biotechnol.* *89*, 1289–1303.

- Chang, V.S., and Holtzapple, M.T. (2000). Fundamental factors affecting biomass enzymatic reactivity. *Appl. Biochem. Biotechnol.* *84*, 5–37.
- Chassard, C., and Bernalier-Donadille, A. (2006). H<sub>2</sub> and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. *FEMS Microbiol. Lett.* *254*, 116–122.
- Chassard, C., Gaillard-Martinie, B., and Bernalier-Donadille, A. (2005). Interaction between H<sub>2</sub>-producing and non-H<sub>2</sub>-producing cellulolytic bacteria from the human colon. *FEMS Microbiol. Lett.* *242*, 339–344.
- Chassard, C., Goumy, V., Leclerc, M., Del'homme, C., and Bernalier-Donadille, A. (2007). Characterization of the xylan-degrading microbial community from human faeces. *FEMS Microbiol. Ecol.* *61*, 121–131.
- Chen, H. (2014). Chemical Composition and Structure of Natural Lignocellulose. In *Biotechnology of Lignocellulose: Theory and Practice*, H. Chen, ed. (Dordrecht: Springer Netherlands), pp. 25–71.
- Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., and Hwang, C.-C. (2013). Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLOS ONE* *8*, e62856.
- Cherubini, F., and Strømman, A.H. (2011). Chemicals from lignocellulosic biomass: opportunities, perspectives, and potential of biorefinery systems. *Biofuels Bioprod. Biorefining* *5*, 548–561.
- Chu, Y., Hao, Z., Wang, K., Tu, T., Huang, H., Wang, Y., Bai, Y.G., Wang, Y., Luo, H., Yao, B., et al. (2019). The GH10 and GH48 dual-functional catalytic domains from a multimodular glycoside hydrolase synergize in hydrolyzing both cellulose and xylan. *Biotechnol. Biofuels* *12*, 279.
- Chukwuma, O.B., Rafatullah, M., Tajarudin, H.A., and Ismail, N. (2021). A Review on Bacterial Contribution to Lignocellulose Breakdown into Useful Bio-Products. *Int. J. Environ. Res. Public. Health* *18*, 6001.
- Chynoweth, D.P., and Pullammanappallil, P. (1996). Anaerobic Digestion of Municipal Solid Wastes. In *Microbiology of Solid Waste*, (CRC Press), p.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* *42*, D633–D642.
- Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretailade, E., Bayat, A.R., Shingfield, K.J., Peyret, P., and Forano, E. (2017). Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Front. Microbiol.* *8*.
- Cortes-Tolalpa, L., Jiménez, D.J., de Lima Brossi, M.J., Salles, J.F., and van Elsas, J.D. (2016). Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. *Appl. Microbiol. Biotechnol.* *100*, 7713–7725.

- Cortes-Tolalpa, L., Salles, J.F., Elsas, V., and Dirk, J. (2017). Bacterial Synergism in Lignocellulose Biomass Degradation – Complementary Roles of Degraders As Influenced by Complexity of the Carbon Source. *Front. Microbiol.* *8*.
- da Costa Sousa, L., Chundawat, S.P., Balan, V., and Dale, B.E. (2009). ‘Cradle-to-grave’ assessment of existing lignocellulose pretreatment technologies. *Curr. Opin. Biotechnol.* *20*, 339–347.
- Costea, P.I., Zeller, G., Sunagawa, S., and Bork, P. (2014). A fair comparison. *Nat. Methods* *11*, 359–359.
- Cragg, S.M., Beckham, G.T., Bruce, N.C., Bugg, T.D., Distel, D.L., Dupree, P., Etxabe, A.G., Goodell, B.S., Jellison, J., McGeehan, J.E., et al. (2015). Lignocellulose degradation mechanisms across the Tree of Life. *Curr. Opin. Chem. Biol.* *29*, 108–119.
- Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinforma. Oxf. Engl.* *20*, 1466–1467.
- Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., Liu, L., Luo, Y., Liu, D., Zheng, H., et al. (2015). Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide Degradation by Microorganisms in the Cow Rumen. *Appl. Environ. Microbiol.* *81*, 1375–1386.
- Dann, L.M., McKerral, J.C., Smith, R.J., Tobe, S.S., Paterson, J.S., Seymour, J.R., Oliver, R.L., and Mitchell, J.G. (2018). Microbial micropatches within microbial hotspots. *PLOS ONE* *13*, e0197224.
- De Baere, L. (2000). Anaerobic digestion of solid waste: state-of-the-art. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* *41*, 283–290.
- DeAngelis, K.M., Fortney, J.L., Borglin, S., Silver, W.L., Simmons, B.A., and Hazen, T.C. (2012). Anaerobic decomposition of switchgrass by tropical soil-derived feedstock-adapted consortia. *MBio* *3*, e00249-11.
- Dehority, B.A. (1991). Effects of microbial synergism on fibre digestion in the rumen. *Proc. Nutr. Soc.* *50*, 149–159.
- Delannoy-Bruno, O., Desai, C., Raman, A.S., Chen, R.Y., Hibberd, M.C., Cheng, J., Han, N., Castillo, J.J., Couture, G., Lebrilla, C.B., et al. (2021). Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* *595*, 91–95.
- Delmas, S., Pullan, S.T., Gaddipati, S., Kokolski, M., Malla, S., Blythe, M.J., Ibbett, R., Campbell, M., Liddell, S., Aboobaker, A., et al. (2012). Uncovering the Genome-Wide Transcriptional Responses of the Filamentous Fungus *Aspergillus niger* to Lignocellulose Using RNA Sequencing. *PLOS Genet.* *8*, e1002875.
- Denef, V.J., Kalnejais, L.H., Mueller, R.S., Wilmes, P., Baker, B.J., Thomas, B.C., VerBerkmoes, N.C., Hettich, R.L., and Banfield, J.F. (2010). Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc. Natl. Acad. Sci.* *107*, 2383–2390.
- Desvaux, M. (2005). *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* *29*, 741–764.

- Deusch, S., Tilocca, B., Camarinha-Silva, A., and Seifert, J. (2015). News in livestock research — use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. *Comput. Struct. Biotechnol. J.* *13*, 55–63.
- Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutschord, M., and Seifert, J. (2017). A Structural and Functional Elucidation of the Rumen Microbiome Influenced by Various Diets and Microenvironments. *Front. Microbiol.* *8*.
- D’haeseleer, P., Gladden, J.M., Allgaier, M., Chain, P.S.G., Tringe, S.G., Malfatti, S.A., Aldrich, J.T., Nicora, C.D., Robinson, E.W., Paša-Tolić, L., et al. (2013). Proteogenomic Analysis of a Thermophilic Bacterial Consortium Adapted to Deconstruct Switchgrass. *PLoS ONE* *8*.
- Dijkstra, J. (1994). Production and absorption of volatile fatty acids in the rumen. *Livest. Prod. Sci.* *39*, 61–69.
- Ding, S.Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V., Shoham, Y., Bayer, E.A., and Flint, H.J. (2001). Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* *183*, 1945–1953.
- Doi, R.H., and Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* *2*, 541–551.
- Douglas, G.M., Beiko, R.G., and Langille, M.G.I. (2018). Predicting the Functional Potential of the Microbiome from Marker Genes Using PICRUSt. In *Microbiome Analysis: Methods and Protocols*, R.G. Beiko, W. Hsiao, and J. Parkinson, eds. (New York, NY: Springer), pp. 169–177.
- Dumond, L., Lam, P.Y., van Erven, G., Kabel, M., Mounet, F., Grima-Pettenati, J., Tobimatsu, Y., and Hernandez-Raquet, G. (2021). Termite Gut Microbiota Contribution to Wheat Straw Delignification in Anaerobic Bioreactors. *ACS Sustain. Chem. Eng.* *9*, 2191–2202.
- Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., and Flint, H.J. (2002). Acetate Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-Producing Bacteria from the Human Large Intestine. *Appl. Environ. Microbiol.* *68*, 5186–5190.
- Eggeman, T., and Elander, R.T. (2005). Process and economic analysis of pretreatment technologies. *Bioresour. Technol.* *96*, 2019–2025.
- Eichorst, S.A., and Kuske, C.R. (2012). Identification of cellulose-responsive bacterial and fungal communities in geographically and edaphically different soils by using stable isotope probing. *Appl. Environ. Microbiol.* *78*, 2316–2327.
- Eichorst, S.A., Varanasi, P., Stavila, V., Zemla, M., Auer, M., Singh, S., Simmons, B.A., and Singer, S.W. (2013). Community dynamics of cellulose-adapted thermophilic bacterial consortia. *Environ. Microbiol.* *15*, 2573–2587.
- Eichorst, S.A., Joshua, C., Sathitsuksanoh, N., Singh, S., Simmons, B.A., and Singer, S.W. (2014). Substrate-Specific Development of Thermophilic Bacterial Consortia by Using Chemically Pretreated Switchgrass. *Appl. Environ. Microbiol.* *80*, 7423–7432.



- Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214.
- Eng, J.K., Fischer, B., Grossmann, J., and Maccoss, M.J. (2008). A fast SEQUEST cross correlation algorithm. *J. Proteome Res.* 7, 4598–4602.
- Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., and Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J.* 4, 642–647.
- Eryildiz, B., Lukitawesa, and Taherzadeh, M.J. (2020). Effect of pH, substrate loading, oxygen, and methanogens inhibitors on volatile fatty acid (VFA) production from citrus waste by anaerobic digestion. *Bioresour. Technol.* 302, 122800.
- Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., and Pascal, G. (2017). FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinforma. Oxf. Engl.*
- Eun, J.-S., Beauchemin, K.A., Hong, S.-H., and Bauer, M.W. (2006). Exogenous enzymes added to untreated or ammoniated rice straw: Effects on in vitro fermentation characteristics and degradability. *Anim. Feed Sci. Technol.* 131, 87–102.
- Foroughi, F., Rezvani Ghomi, E., Morshedi Dehaghi, F., Borayek, R., and Ramakrishna, S. (2021). A Review on the Life Cycle Assessment of Cellulose: From Properties to the Potential of Making It a Low Carbon Material. *Materials* 14, 714.
- Ge, X., Matsumoto, T., Keith, L., and Li, Y. (2014). Biogas energy production from tropical biomass wastes by anaerobic digestion. *Bioresour. Technol.* 169, 38–44.
- Geng, W., Jin, Y., Jameel, H., and Park, S. (2015). Strategies to achieve high-solids enzymatic hydrolysis of dilute-acid pretreated corn stover. *Bioresour. Technol.* 187, 43–48.
- Gilbert, H.J., Knox, J.P., and Boraston, A.B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* 23, 669–677.
- Gírio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S., and Bogel-Lukasik, R. (2010). Hemicelluloses for fuel ethanol: A review. *Bioresour. Technol.* 101, 4775–4800.
- Gladden, J.M., Eichorst, S.A., Hazen, T.C., Simmons, B.A., and Singer, S.W. (2012). Substrate perturbation alters the glycoside hydrolase activities and community composition of switchgrass-adapted bacterial consortia. *Biotechnol. Bioeng.* 109, 1140–1145.
- Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* 8, 2224.
- Goldbeck, R., Gonçalves, T.A., Damásio, A.R.L., Brenelli, L.B., Wolf, L.D., Paixão, D.A.A., Rocha, G.J.M., and Squina, F.M. (2016). Effect of hemicellulolytic enzymes to improve sugarcane bagasse saccharification and xylooligosaccharides production. *J. Mol. Catal. B Enzym.* 131, 36–46.

- Graf, J. (2014). The Family Rikenellaceae. *Prokaryotes Major Lineages Bact. Archaea* 857–859.
- Gralka, M., Szabo, R., Stocker, R., and Cordero, O.X. (2020). Trophic Interactions and the Drivers of Microbial Community Assembly. *Curr. Biol.* 30, R1176–R1188.
- Grbić-Galić, D. (1983). Anaerobic degradation of coniferyl alcohol by methanogenic consortia. *Appl. Environ. Microbiol.* 46, 1442–1446.
- Grethlein, H.E. (1985). The Effect of Pore Size Distribution on the Rate of Enzymatic Hydrolysis of Cellulosic Substrates. *Bio/Technology* 3, 155–160.
- Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., and Brumer, H. (2017). Polysaccharide Utilization Loci: Fueling Microbial Communities. *J. Bacteriol.* 199.
- Gu, Y., Chen, X., Liu, Z., Zhou, X., and Zhang, Y. (2014). Effect of inoculum sources on the anaerobic digestion of rice straw. *Bioresour. Technol.* 158, 149–155.
- Guimarães, J.L., Frollini, E., da Silva, C.G., Wypych, F., and Satyanarayana, K.G. (2009). Characterization of banana, sugarcane bagasse and sponge gourd fibers of Brazil. *Ind. Crops Prod.* 30, 407–415.
- Guo, H., Wang, X.-D., and Lee, D.-J. (2018). Proteomic researches for lignocellulose-degrading enzymes: A mini-review. *Bioresour. Technol.* 265, 532–541.
- Guo, P., Mochizuki, K., Cheng, W., Zhou, M., Gao, H., Zheng, D., Wang, X., and Cui, Z. (2011). Effects of different pretreatment strategies on corn stalk acidogenic fermentation using a microbial consortium. *Bioresour. Technol.* 102, 7526–7531.
- Harholt, J., Suttangkakul, A., and Vibe Scheller, H. (2010). Biosynthesis of Pectin. *Plant Physiol.* 153, 384–395.
- Hart, E.H., Creevey, C.J., Hitch, T., and Kingston-Smith, A.H. (2018). Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria. *Sci. Rep.* 8, 10504.
- Haruta, S., Cui, Z., Huang, Z., Li, M., Ishii, M., and Igarashi, Y. (2002). Construction of a stable microbial community with high cellulose-degradation ability. *Appl. Microbiol. Biotechnol.* 59, 529–534.
- Hawinkel, S., Bijmans, L., Cao, K.-A.L., and Thas, O. (2020). Model-based joint visualization of multiple compositional omics datasets. *NAR Genomics Bioinforma.* 2.
- Hawley, A.K., Brewer, H.M., Norbeck, A.D., Paša-Tolić, L., and Hallam, S.J. (2014). Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci.* 111, 11395–11400.
- Heeg, K., Pohl, M., Sontag, M., Mumme, J., Klocke, M., and Nettmann, E. (2014). Microbial communities involved in biogas production from wheat straw as the sole substrate within a two-phase solid-state anaerobic digestion. *Syst. Appl. Microbiol.* 37, 590–600.



- Hemsworth, G.R., Johnston, E.M., Davies, G.J., and Walton, P.H. (2015). Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.* *33*, 747–761.
- Hendriks, A.T.W.M., and Zeeman, G. (2009). Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresour. Technol.* *100*, 10–18.
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* *331*, 463–467.
- Hindrichsen, I.K., Wettstein, H.-R., Machmüller, A., Soliva, C.R., Bach Knudsen, K.E., Madsen, J., and Kreuzer, M. (2004). Effects of feed carbohydrates with contrasting properties on rumen fermentation and methane release in vitro. *Can. J. Anim. Sci.* *84*, 265–276.
- Hodge, D.B., Karim, M.N., Schell, D.J., and McMillan, J.D. (2008). Model-Based Fed-Batch for High-Solids Enzymatic Cellulose Hydrolysis. *Appl. Biochem. Biotechnol.* *152*, 88–107.
- Holck, J., Fredslund, F., Møller, M.S., Brask, J., Krogh, K.B.R.M., Lange, L., Welner, D.H., Svensson, B., Meyer, A.S., and Wilkens, C. (2019). A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan. *J. Biol. Chem.* *294*, 17339–17353.
- Hollister, E.B., Forrest, A.K., Wilkinson, H.H., Ebbole, D.J., Malfatti, S.A., Tringe, S.G., Holtzapple, M.T., and Gentry, T.J. (2010). Structure and dynamics of the microbial communities underlying the carboxylate platform for biofuel production. *Appl. Microbiol. Biotechnol.* *88*, 389–399.
- Hollister, E.B., Forrest, A.K., Wilkinson, H.H., Ebbole, D.J., Tringe, S.G., Malfatti, S.A., Holtzapple, M.T., and Gentry, T.J. (2012). Mesophilic and thermophilic conditions select for unique but highly parallel microbial communities to perform carboxylate platform biomass conversion. *PloS One* *7*, e39689.
- Hu, Z.-H., and Yu, H.-Q. (2005). Application of rumen microorganisms for enhanced anaerobic fermentation of corn stover. *Process Biochem.* *40*, 2371–2377.
- Hu, Z., Wang, Y., and Wen, Z. (2008). Alkali (NaOH) Pretreatment of Switchgrass by Radio Frequency-based Dielectric Heating. *Appl. Biochem. Biotechnol.* *148*, 71–81.
- Hu, Z.-H., Wang, G., and Yu, H.-Q. (2004). Anaerobic degradation of cellulose by rumen microorganisms at various pH values. *Biochem. Eng. J.* *21*, 59–62.
- Janeček, Š., Svensson, B., and MacGregor, E.A. (2011). Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb. Technol.* *49*, 429–440.
- Janke, L., Leite, A., Nikolausz, M., Schmidt, T., Liebetrau, J., Nelles, M., and Stinner, W. (2015). Biogas Production from Sugarcane Waste: Assessment on Kinetic Challenges for Process Designing. *Int. J. Mol. Sci.* *16*, 20685–20703.
- Jensen, P.D., Hardin, M.T., and Clarke, W.P. (2008). Measurement and quantification of sessile and planktonic microbial populations during the anaerobic digestion of cellulose. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* *57*, 465–469.

Jiang, Y., Zhang, X., Yuan, H., Huang, D., Wang, R., Liu, H., and Wang, T. (2021). Research progress and the biotechnological applications of multienzyme complex. *Appl. Microbiol. Biotechnol.* *105*, 1759–1777.

Jiménez, D.J., Maruthamuthu, M., and van Elsas, J.D. (2015). Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. *Biotechnol. Biofuels* *8*.

Jiménez, D.J., de Lima Brossi, M.J., Schückel, J., Kračun, S.K., Willats, W.G.T., and van Elsas, J.D. (2016). Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl. Microbiol. Biotechnol.* *100*, 10463–10477.

Jones, J.A.D., Kerr, R.G., Haltli, B.A., and Tinto, W.F. (2018). Temperature and pH effect on glucose production from pretreated bagasse by a novel species of *Citrobacter* and other bacteria. *Heliyon* *4*, e00657.

Jose, V.L., Appoorthy, T., More, R.P., and Arun, A.S. (2017). Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express* *7*.

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* *4*, 923–925.

Käll, L., Storey, J.D., MacCoss, M.J., and Noble, W.S. (2008). Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J. Proteome Res.* *7*, 29–34.

Kamra, D.N. (2005). Rumen microbial ecosystem. *Curr. Sci.* *89*, 124–135.

Karri, S., Vadela, M.B., and Gundi, V.A.K.B. (2021). Chapter 15 - Fiber degradation strategies of bacteria in rumen ecosystem. In *Recent Developments in Applied Microbiology and Biochemistry*, B. Viswanath, ed. (Academic Press), pp. 153–159.

Kaul, A., Mandal, S., Davidov, O., and Peddada, S.D. (2017). Analysis of Microbiome Data in the Presence of Excess Zeros. *Front. Microbiol.* *8*, 2114.

Kenealy, W.R., Cao, Y., and Weimer, P.J. (1995). Production of caproic acid by cocultures of ruminal cellulolytic bacteria and *Clostridium kluyveri* grown on cellulose and ethanol. *Appl. Microbiol. Biotechnol.* *44*, 507–513.

Kerckhof, F.-M., Courtens, E.N.P., Geirnaert, A., Hoefman, S., Ho, A., Vilchez-Vargas, R., Pieper, D.H., Jauregui, R., Vlaeminck, S.E., Wiele, T.V. de, et al. (2014). Optimized Cryopreservation of Mixed Microbial Communities for Conserved Functionality and Diversity. *PLOS ONE* *9*, e99517.

Keunen, J.E., Plaizier, J.C., Kyriazakis, L., Duffield, T.F., Widowski, T.M., Lindinger, M.I., and McBride, B.W. (2002). Effects of a Subacute Ruminant Acidosis Model on the Diet Selection of Dairy Cows. *J. Dairy Sci.* *85*, 3304–3313.

- Khan, M.U., and Ahring, B.K. (2019). Lignin degradation under anaerobic digestion: Influence of lignin modifications -A review. *Biomass Bioenergy* 128, 105325.
- Kim, M., and Day, D.F. (2011). Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. *J. Ind. Microbiol. Biotechnol.* 38, 803–807.
- Kim, S., and Dale, B.E. (2004). Global potential bioethanol production from wasted crops and crop residues. *Biomass Bioenergy* 26, 361–375.
- Kim, N.-J., Park, G.W., Kang, J., Kim, Y.-C., and Chang, H.N. (2013). Volatile fatty acid production from lignocellulosic biomass by lime pretreatment and its applications to industrial biotechnology. *Biotechnol. Bioprocess Eng.* 18, 1163–1168.
- Kinet, R., Destain, J., Hiligsmann, S., Thonart, P., Delhalle, L., Taminiau, B., Daube, G., and Delvigne, F. (2015). Thermophilic and cellulolytic consortium isolated from composting plants improves anaerobic digestion of cellulosic biomass: Toward a microbial resource management approach. *Bioresour. Technol.* 189, 138–144.
- Kivaisi, A.K., and Eliapenda, S. (1995). Application of rumen microorganisms for enhanced anaerobic degradation of bagasse and maize bran. *Biomass Bioenergy* 8, 45–50.
- Kivaisi, A.K., Gijzen, H.J., Op den Camp, H.J., and Vogels, G.D. (1992). Conversion of cereal residues into biogas in a rumen-derived process. *World J. Microbiol. Biotechnol.* 8, 428–433.
- Kleiner, M., Thorson, E., Sharp, C.E., Dong, X., Liu, D., Li, C., and Strous, M. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* 8, 1558.
- Klevenhusen, F., Petri, R.M., Kleefisch, M.-T., Khiaosa-ard, R., Metzler-Zebeli, B.U., and Zebeli, Q. (2017). Changes in fibre-adherent and fluid-associated microbial communities and fermentation profiles in the rumen of cattle fed diets differing in hay quality and concentrate amount. *FEMS Microbiol. Ecol.* 93.
- Kobayashi, H., Ohta, H., and Fukuoka, A. (2012). Conversion of lignocellulose into renewable chemicals by heterogeneous catalysis. *Catal. Sci. Technol.* 2, 869–883.
- Kumar, K.S., Manimaran, A., Permaul, K., and Singh, S. (2009a). Production of  $\beta$ -xylanase by a *Thermomyces lanuginosus* MC 134 mutant on corn cobs and its application in biobleaching of bagasse pulp. *J. Biosci. Bioeng.* 107, 494–498.
- Kumar, P., Barrett, D.M., Delwiche, M.J., and Stroeve, P. (2009b). Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. *Ind. Eng. Chem. Res.* 48, 3713–3729.
- Kumar, R., Singh, S., and Singh, O.V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* 35, 377–391.
- La Reau, A.J., and Suen, G. (2018). The Ruminococci: key symbionts of the gut ecosystem. *J. Microbiol.* 56, 199–208.

- Lamed, R., Setter, E., Kenig, R., and Edward, B. (1983). The cellulosome: A discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. In *Biotechnology Bioengineering Symposium*, p.
- Lapébie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019). Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* *10*, 2043.
- Larsson, P.T., Wickholm, K., and Iversen, T. (1997). A CP/MAS13C NMR investigation of molecular ordering in celluloses. *Carbohydr. Res.* *302*, 19–25.
- Laureano-Perez, L., Teymouri, F., Alizadeh, H., and Dale, B.E. (2005). Understanding factors that limit enzymatic hydrolysis of biomass. *Appl. Biochem. Biotechnol.* *124*, 1081–1099.
- Lazuka, A., Auer, L., Bozonnet, S., Morgavi, D.P., O'Donohue, M., and Hernandez-Raquet, G. (2015). Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. *Bioresour. Technol.* *196*, 241–249.
- Lazuka, A., Roland, C., Barakat, A., Guillon, F., O'Donohue, M., and Hernandez-Raquet, G. (2017). Ecofriendly lignocellulose pretreatment to enhance the carboxylate production of a rumen-derived microbial consortium. *Bioresour. Technol.* *236*, 225–233.
- Lazuka, A., Auer, L., O'Donohue, M., and Hernandez-Raquet, G. (2018). Anaerobic lignocellulolytic microbial consortium derived from termite gut: enrichment, lignocellulose degradation and community dynamics. *Biotechnol. Biofuels* *11*, 284.
- Lê Cao, K.-A., Boitard, S., and Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* *12*, 253.
- Lemée, L., Kpogbemabou, D., Pinard, L., Beauchet, R., and Laduranty, J. (2012). Biological pretreatment for production of lignocellulosic biofuel. *Bioresour. Technol.* *117*, 234–241.
- Leth, M.L., Ejby, M., Workman, C., Ewald, D.A., Pedersen, S.S., Sternberg, C., Bahl, M.I., Licht, T.R., Aachmann, F.L., Westereng, B., et al. (2018). Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* *3*, 570–580.
- Li, J., Zhong, H., Ramayo-Caldas, Y., Terrapon, N., Lombard, V., Potocki-Veronese, G., Estellé, J., Popova, M., Yang, Z., Zhang, H., et al. (2020a). A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. *GigaScience* *9*, giaa057.
- Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., and Pan, C. (2012). Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* *11*, 1582–1590.
- Li, Z., Mu, C., Xu, Y., Shen, J., and Zhu, W. (2020b). Changes in the Solid-, Liquid-, and Epithelium-Associated Bacterial Communities in the Rumen of Hu Lambs in Response to Dietary Urea Supplementation. *Front. Microbiol.* *11*, 244.

- de Lima Brossi, M.J., Jiménez, D.J., Cortes-Tolalpa, L., and van Elsas, J.D. (2016). Soil-Derived Microbial Consortia Enriched with Different Plant Biomass Reveal Distinct Players Acting in Lignocellulose Degradation. *Microb. Ecol.* *71*, 616–627.
- Lin, Y.-W., Tuan, N.N., and Huang, S.-L. (2016). Metaproteomic analysis of the microbial community present in a thermophilic swine manure digester to allow functional characterization: A case study. *Int. Biodeterior. Biodegrad.* *115*, 64–73.
- Linder, M., and Teeri, T.T. (1997). The roles and function of cellulose-binding domains. *J. Biotechnol.* *57*, 15–28.
- Liu, D., Li, M., Xi, B., Zhao, Y., Wei, Z., Song, C., and Zhu, C. (2015). Metaproteomics reveals major microbial players and their biodegradation functions in a large-scale aerobic composting plant. *Microb. Biotechnol.* *8*, 950–960.
- Liu, D., Keiblinger, K.M., Schindlbacher, A., Wegner, U., Sun, H., Fuchs, S., Lassek, C., Riedel, K., and Zechmeister-Boltenstern, S. (2017). Microbial functionality as affected by experimental warming of a temperate mountain forest soil—A metaproteomics survey. *Appl. Soil Ecol.* *117–118*, 196–202.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* *42*, D490-495.
- López-Mondéjar, R., Zühlke, D., Větrovský, T., Becher, D., Riedel, K., and Baldrian, P. (2016a). Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer *Paenibacillus O199*. *Biotechnol. Biofuels* *9*, 104.
- López-Mondéjar, R., Zühlke, D., Becher, D., Riedel, K., and Baldrian, P. (2016b). Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci. Rep.* *6*.
- López-Mondéjar, R., Algora, C., and Baldrian, P. (2019). Lignocellulolytic systems of soil bacteria: A vast and diverse toolbox for biotechnological conversion processes. *Biotechnol. Adv.* *37*, 107374.
- Lovell, D., Pawlowsky-Glahn, V., Egozcue, J.J., Marguerat, S., and Bähler, J. (2015). Proportionality: A Valid Alternative to Correlation for Relative Data. *PLOS Comput. Biol.* *11*, e1004075.
- Lü, F., Chai, L., Shao, L., and He, P. (2017). Precise pretreatment of lignocellulose: relating substrate modification with subsequent hydrolysis and fermentation to products and by-products. *Biotechnol. Biofuels* *10*, 88.
- Lubbe, S., Filzmoser, P., and Templ, M. (2021). Comparison of zero replacement strategies for compositional data with large numbers of zeros. *Chemom. Intell. Lab. Syst.* *210*, 104248.
- Ludwig, D., Michael, B., Hirth, T., Rupp, S., and Zibek, S. (2013). High Solids Enzymatic Hydrolysis of Pretreated Lignocellulosic Materials with a Powerful Stirrer Concept. *Appl. Biochem. Biotechnol.* *172*, 1699–1713.
- Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002). Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.* *66*, 506–577.

- Mackenzie, A.K., Naas, A.E., Kracun, S.K., Schückel, J., Fangel, J.U., Agger, J.W., Willats, W.G.T., Eijsink, V.G.H., and Pope, P.B. (2015). A polysaccharide utilization locus from an uncultured bacteroidetes phylotype suggests ecological adaptation and substrate versatility. *Appl. Environ. Microbiol.* *81*, 187–195.
- Mafa, M.S., Malgas, S., and Pletschke, B.I. (2021). Feruloyl esterase (FAE-1) sourced from a termite hindgut and GH10 xylanases synergy improves degradation of arabinoxylan. *AMB Express* *11*, 21.
- Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* *2*.
- Malgas, S., Thoresen, M., van Dyk, J.S., and Pletschke, B.I. (2017). Time dependence of enzyme synergism during the degradation of model and natural lignocellulosic substrates. *Enzyme Microb. Technol.* *103*, 1–11.
- Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., and Peddada, S.D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* *26*, 10.3402/mehd.v26.27663.
- Mansfield, S.D., Mooney, C., and Saddler, J.N. (1999). Substrate and Enzyme Characteristics that Limit Cellulose Hydrolysis. *Biotechnol. Prog.* *15*, 804–816.
- Marone, A., Trably, E., Carrère, H., Prompsy, P., Guillon, F., Joseph-Aimé, M., Barakat, A., Fayoud, N., Bernet, N., and Escudié, R. (2019). Enhancement of corn stover conversion to carboxylates by extrusion and biotic triggers in solid-state fermentation. *Appl. Microbiol. Biotechnol.* *103*, 489–503.
- Marriott, P.E., Gómez, L.D., and McQueen-Mason, S.J. (2016). Unlocking the potential of lignocellulosic biomass through plant science. *New Phytol.* *209*, 1366–1381.
- Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *J. Biol. Chem.* *284*, 24673–24677.
- Martin, M.A. (2010). First generation biofuels compete. *New Biotechnol.* *27*, 596–608.
- Martín-Fernández, J.A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Math. Geol.* *35*, 253–278.
- Martins, L.H. da S., Rabelo, S.C., and da Costa, A.C. (2015). Effects of the pretreatment method on high solids enzymatic hydrolysis and ethanol fermentation of the cellulosic fraction of sugarcane bagasse. *Bioresour. Technol.* *191*, 312–321.
- Marynowska, M., Goux, X., Sillam-Dussès, D., Rouland-Lefèvre, C., Roisin, Y., Delfosse, P., and Calusinska, M. (2017). Optimization of a metatranscriptomic approach to study the lignocellulolytic potential of the higher termite gut microbiome. *BMC Genomics* *18*.
- Marynowska, M., Goux, X., Sillam-Dussès, D., Rouland-Lefèvre, C., Halder, R., Wilmes, P., Gawron, P., Roisin, Y., Delfosse, P., and Calusinska, M. (2020). Compositional and functional



characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites. *Microbiome* 8, 96.

McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J., and Knox, J.P. (2006). Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4765–4770.

McMillan, J.D. (1994). Pretreatment of Lignocellulosic Biomass. In *Enzymatic Conversion of Biomass for Fuels Production*, (American Chemical Society), pp. 292–324.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8, e61217.

Meng, X., Foston, M., Leisen, J., DeMartini, J., Wyman, C.E., and Ragauskas, A.J. (2013). Determination of porosity of lignocellulosic biomass before and after pretreatment by using Simons' stain and NMR techniques. *Bioresour. Technol.* 144, 467–476.

Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016). Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. *Appl. Environ. Microbiol.* 82, 1686–1692.

Mille-Lindblom, C., and Tranvik, L.J. (2003). Antagonism between Bacteria and Fungi on Decomposing Aquatic Plant Litter. *Microb. Ecol.* 45, 173–182.

Mirande, C., Kadlecikova, E., Matulova, M., Capek, P., Bernalier-Donadille, A., Forano, E., and Béra-Maillet, C. (2010). Dietary fibre degradation and fermentation by two xylanolytic bacteria *Bacteroides xylanisolvens* XB1AT and *Roseburia intestinalis* XB6B4 from the human intestine. *J. Appl. Microbiol.* 109, 451–460.

Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001). Invited Review: Adhesion Mechanisms of Rumen Cellulolytic Bacteria. *J. Dairy Sci.* 84, 1294–1309.

Mkabayi, L., Malgas, S., Wilhelmi, B.S., and Pletschke, B.I. (2020). Evaluating Feruloyl Esterase—Xylanase Synergism for Hydroxycinnamic Acid and Xylo-Oligosaccharide Production from Untreated, Hydrothermally Pre-Treated and Dilute-Acid Pre-Treated Corn Cobs. *Agronomy* 10, 688.

Mockaitis, G., Bruant, G., Guiot, S.R., Peixoto, G., Foresti, E., and Zaiat, M. (2020). Acidic and thermal pre-treatments for anaerobic digestion inoculum to improve hydrogen and volatile fatty acid production using xylose as the substrate. *Renew. Energy* 145, 1388–1398.

Modenbach, A.A., and Nokes, S.E. (2012). The use of high-solids loadings in biomass pretreatment—a review. *Biotechnol. Bioeng.* 109, 1430–1442.

Modenbach, A.A., and Nokes, S.E. (2013). Enzymatic hydrolysis of biomass at high-solids loadings – A review. *Biomass Bioenergy* 56, 526–544.

Moletta, R. (2011). *La méthanisation* (Éd. Tec & doc).

Monlau, F., Barakat, A., Steyer, J.P., and Carrere, H. (2012). Comparison of seven types of thermo-chemical pretreatments on the structural features and anaerobic digestion of sunflower stalks. *Bioresour. Technol.* 120, 241–247.

- Morotomi, M., Nagai, F., and Watanabe, Y. (2012). Description of *Christensenella minuta* gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* *62*, 144–149.
- Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M., and Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* *96*, 673–686.
- Motte, J.-C., Trably, E., Escudié, R., Hamelin, J., Steyer, J.-P., Bernet, N., Delgenes, J.-P., and Dumas, C. (2013). Total solids content: a key parameter of metabolic pathways in dry anaerobic digestion. *Biotechnol. Biofuels* *6*, 164.
- Motte, J.-C., Trably, E., Hamelin, J., Escudié, R., Bonnafous, A., Steyer, J.-P., Bernet, N., Delgenes, J.-P., and Dumas, C. (2014). Total solid content drives hydrogen production through microbial selection during thermophilic fermentation. *Bioresour. Technol.* *166*, 610–615.
- Mouriño, F., Akkarawongsa, R., and Weimer, P.J. (2001). Initial pH as a determinant of cellulose digestion rate by mixed ruminal microorganisms in vitro. *J. Dairy Sci.* *84*, 848–859.
- Munir, R., and Levin, D.B. (2016). *Enzyme Systems of Anaerobes for Biomass Conversion*. *Adv. Biochem. Eng. Biotechnol.*
- Munir, R.I., Schellenberg, J., Henrissat, B., Verbeke, T.J., Sparling, R., and Levin, D.B. (2014). Comparative Analysis of Carbohydrate Active Enzymes in *Clostridium termitidis* CT1112 Reveals Complex Carbohydrate Degradation Ability. *PLoS ONE* *9*.
- Murphy, M.R., Baldwin, R.L., and Koong, L.J. (1982). Estimation of Stoichiometric Parameters for Rumen Fermentation of Roughage and Concentrate Diets. *J. Anim. Sci.* *55*, 411–421.
- Murray, W.D. (1986). Symbiotic Relationship of *Bacteroides cellulosolvens* and *Clostridium saccharolyticum* in Cellulose Fermentation. *Appl. Environ. Microbiol.* *51*, 710–714.
- Muth, T., Kolmeder, C.A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F.J., Rensen, S.S., Reichl, U., de Vos, W.M., Rapp, E., et al. (2015). Navigating through metaproteomics data: a logbook of database searching. *Proteomics* *15*, 3439–3453.
- Naas, A.E., Mackenzie, A.K., Mravec, J., Schückel, J., Willats, W.G.T., Eijsink, V.G.H., and Pope, P.B. (2014). Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* *5*, e01401-01414.
- Naik, S.N., Goud, V.V., Rout, P.K., and Dalai, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* *14*, 578–597.
- Neumann, A.P., and Suen, G. (2018). The Phylogenomic Diversity of Herbivore-Associated *Fibrobacter* spp. Is Correlated to Lignocellulose-Degrading Potential. *MSphere* *3*, e00593-18.
- Neumann, A.P., McCormick, C.A., and Suen, G. (2017). *Fibrobacter* communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen. *Environ. Microbiol.* *19*, 3768–3783.



- Nishiyama, T., Ueki, A., Kaku, N., Watanabe, K., and Ueki, K. (2009). *Bacteroides graminisolvens* sp. nov., a xylanolytic anaerobe isolated from a methanogenic reactor treating cattle waste. *Int. J. Syst. Evol. Microbiol.* *59*, 1901–1907.
- Nitsos, C.K., Choli-Papadopoulou, T., Matis, K.A., and Triantafyllidis, K.S. (2016). Optimization of Hydrothermal Pretreatment of Hardwood and Softwood Lignocellulosic Residues for Selective Hemicellulose Recovery and Improved Cellulose Enzymatic Hydrolysis. *ACS Sustain. Chem. Eng.* *4*, 4529–4544.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2020). *vegan: Community Ecology Package*.
- O'Sullivan, A.C. (1997). Cellulose: the structure slowly unravels. *Cellulose* *4*, 173–207.
- Overmann, J., Fischer, U., and Pfennig, N. (1992). A new purple sulfur bacterium from saline littoral sediments, *Thiorhodovibrio winogradskyi* gen. nov. and sp. nov. *Arch. Microbiol.* *157*, 329–335.
- Ozbayram, E.G., Kleinstuber, S., Nikolausz, M., Ince, B., and Ince, O. (2018). Enrichment of lignocellulose-degrading microbial communities from natural and engineered methanogenic environments. *Appl. Microbiol. Biotechnol.* *102*, 1035–1043.
- Pan, C., and Banfield, J.F. (2014). Quantitative Metaproteomics: Functional Insights into Microbial Communities. In *Environmental Microbiology: Methods and Protocols*, I.T. Paulsen, and A.J. Holmes, eds. (Totowa, NJ: Humana Press), pp. 231–240.
- Pandey, M.P., and Kim, C.S. (2011). Lignin Depolymerization and Conversion: A Review of Thermochemical Methods. *Chem. Eng. Technol.* *34*, 29–41.
- Paoletti, A.C., Parmely, T.J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci.* *103*, 18928–18933.
- Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* *18*, 1403–1414.
- Parawira, W., Murto, M., Read, J.S., and Mattiasson, B. (2005). Profile of hydrolases and biogas production during two-stage mesophilic anaerobic digestion of solid potato waste. *Process Biochem.* *40*, 2945–2952.
- Patz, S., Witzel, K., Scherwinski, A.-C., and Ruppel, S. (2019). Culture Dependent and Independent Analysis of Potential Probiotic Bacterial Genera and Species Present in the Phyllosphere of Raw Eaten Produce. *Int. J. Mol. Sci.* *20*, 3661.
- Paulova, L., Patakova, P., Branska, B., Rychtera, M., and Melzoch, K. (2015). Lignocellulosic ethanol: Technology design and its impact on process efficiency. *Biotechnol. Adv.* *33*, 1091–1107.
- Paulson, J., Stine, O., Corrada Bravo, H., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* *10*.

- Pavlostathis, S.G., and Giraldo-Gomez, E. (1991). Kinetics of Anaerobic Treatment. *Water Sci. Technol.* *24*, 35–59.
- Pereira, G.V., Abdel-Hamid, A.M., Dutta, S., D’Alessandro-Gabazza, C.N., Wefers, D., Farris, J.A., Bajaj, S., Wawrzak, Z., Atomi, H., Mackie, R.I., et al. (2021). Degradation of complex arabinoxylans by human colonic Bacteroidetes. *Nat. Commun.* *12*, 459.
- Pérez, J., Muñoz-Dorado, J., de la Rubia, T., and Martínez, J. (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int. Microbiol. Off. J. Span. Soc. Microbiol.* *5*, 53–63.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* *47*, D442–D450.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* *20*, 3551–3567.
- Pfennig, N., and Trüper, H.G. (1992). The Family Chromatiaceae. p.
- Piao, H., Lachman, M., Malfatti, S., Sczyrba, A., Knierim, B., Auer, M., Tringe, S.G., Mackie, R.I., Yeoman, C.J., and Hess, M. (2014). Temporal dynamics of fibrolytic and methanogenic rumen microorganisms during in situ incubation of switchgrass determined by 16S rRNA gene profiling. *Front. Microbiol.* *5*, 307.
- Pinheiro, B.A., Gilbert, H.J., Sakka, K., Sakka, K., Fernandes, V.O., Prates, J.A.M., Alves, V.D., Bolam, D.N., Ferreira, L.M.A., and Fontes, C.M.G.A. (2009). Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem. J.* *424*, 375–384.
- Polansky, O., Sekelova, Z., Faldynova, M., Sebkova, A., Sisak, F., and Rychlik, I. (2016). Important Metabolic Pathways and Biological Processes Expressed by Chicken Cecal Microbiota. *Appl. Environ. Microbiol.* *82*, 1569–1576.
- Prasad, S., Singh, A., and Joshi, H.C. (2007). Ethanol as an alternative fuel from agricultural, industrial and urban residues. *Resour. Conserv. Recycl.* *50*, 1–39.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* *40*, D290–D301.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* *41*, D590–D596.
- Quinn, T.P., Erb, I., Richardson, M.F., and Crowley, T.M. (2018). Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* *34*, 2870–2878.

Quiroga, A.G., Costa, A., and Filho, R.M. (2010). Analysis of conversion and operation strategies for enzymatic hydrolysis of lignocellulosic biomass in a series of CSTRs with distributed feeding. *Bioprocess Biosyst. Eng.* *33*, 901–910.

Rabelo, S.C., Andrade, R.R., Maciel Filho, R., and Costa, A.C. (2014). Alkaline hydrogen peroxide pretreatment, enzymatic hydrolysis and fermentation of sugarcane bagasse to ethanol. *Fuel* *136*, 349–357.

Rajeswari, G., Jacob, S., Chandel, A.K., and Kumar, V. (2021). Unlocking the potential of insect and ruminant host symbionts for recycling of lignocellulosic carbon with a biorefinery approach: a review. *Microb. Cell Factories* *20*, 107.

Raman, B., Pan, C., Hurst, G.B., Jr, M.R., McKeown, C.K., Lankford, P.K., Samatova, N.F., and Mielenz, J.R. (2009). Impact of Pretreated Switchgrass and Biomass Carbohydrates on *Clostridium thermocellum* ATCC 27405 Cellulosome Composition: A Quantitative Proteomic Analysis. *PLOS ONE* *4*, e5271.

Ramos-Suarez, M., Zhang, Y., and Outram, V. (2021). Current perspectives on acidogenic fermentation to produce volatile fatty acids from waste. *Rev. Environ. Sci. Biotechnol.* *20*, 439–478.

Ransom-Jones, E., Jones, D.L., McCarthy, A.J., and McDonald, J.E. (2012). The Fibrobacteres: an Important Phylum of Cellulose-Degrading Bacteria. *Microb. Ecol.* *63*, 267–281.

Ransom-Jones, E., Jones, D.L., Edwards, A., and McDonald, J.E. (2014). Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. *Syst. Appl. Microbiol.* *37*, 502–509.

Ratanakhanokchai, K., Waeonukul, R., Pason, P., Tachaapaikoon, C., Kyu, K.L., Sakka, K., Kosugi, A., and Mori, Y. (2013). *Paenibacillus curdlanolyticus* Strain B-6 Multienzyme Complex: A Novel System for Biomass Utilization (IntechOpen).

Ravachol, J., Borne, R., Tardif, C., de Philip, P., and Fierobe, H.-P. (2014). Characterization of All Family-9 Glycoside Hydrolases Synthesized by the Cellulosome-producing Bacterium *Clostridium cellulolyticum*. *J. Biol. Chem.* *289*, 7335–7348.

Ravachol, J., Borne, R., Meynial-Salles, I., Soucaille, P., Pagès, S., Tardif, C., and Fierobe, H.-P. (2015). Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium *Ruminiclostridium cellulolyticum*. *Biotechnol. Biofuels* *8*, 114.

Raven, Evert, and Eichhorn (2014). *Biologie végétale - Raven, Evert, Eichhorn ; traduction de la 8e édition américaine "The biology of plants" par Jules Bouharmont (Bruxelles: De Boeck).*

Ravn, J.L., Glitsø, V., Pettersson, D., Ducatelle, R., Van Immerseel, F., and Pedersen, N.R. (2018). Combined endo- $\beta$ -1,4-xylanase and  $\alpha$ -l-arabinofuranosidase increases butyrate concentration during broiler cecal fermentation of maize glucurono-arabinoxylan. *Anim. Feed Sci. Technol.* *236*, 159–169.

Reddy, A.P., Allgaier, M., Singer, S.W., Hazen, T.C., Simmons, B.A., Hugenholtz, P., and VanderGheynst, J.S. (2011). Bioenergy feedstock-specific enrichment of microbial populations during high-solids thermophilic deconstruction. *Biotechnol. Bioeng.* *108*, 2088–2098.

- Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., Flint, H.J., and Louis, P. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* *8*, 1323–1335.
- Ren, Z., You, W., Wu, S., Poetsch, A., and Xu, C. (2019). Secretomic analyses of *Ruminiclostridium papyrosolvens* reveal its enzymatic basis for lignocellulose degradation. *Biotechnol. Biofuels* *12*, 183.
- Riaz, M.N. (2000). *Extruders in Food Applications* (CRC Press).
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* *4*.
- Rohart, F., Gautier, B., Singh, A., and Cao, K.-A.L. (2017). mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLOS Comput. Biol.* *13*, e1005752.
- Rosnow, J.J., Anderson, L.N., Nair, R.N., Baker, E.S., and Wright, A.T. (2017). Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. *Crit. Rev. Biotechnol.* *37*, 626–640.
- Ryckeboer, J., Mergaert, J., Coosemans, J., Deprins, K., and Swings, J. (2003). Microbiological aspects of biowaste during composting in a monitored compost bin. *J. Appl. Microbiol.* *94*, 127–137.
- Saha, B.C., Iten, L.B., Cotta, M.A., and Wu, Y.V. (2005). Dilute acid pretreatment, enzymatic saccharification, and fermentation of rice hulls to ethanol. *Biotechnol. Prog.* *21*, 816–822.
- Sawanon, S., and Kobayashi, Y. (2006). Synergistic fibrolysis in the rumen by cellulolytic *Ruminococcus flavefaciens* and non-cellulolytic *Selenomonas ruminantium*: Evidence in defined cultures. *Anim. Sci. J.* *77*, 208–214.
- Sawatdeenarunat, C., Surendra, K.C., Takara, D., Oechsner, H., and Khanal, S.K. (2015). Anaerobic digestion of lignocellulosic biomass: Challenges and opportunities. *Bioresour. Technol.* *178*, 178–186.
- Scharf, M.E. (2015). Omic research in termites: an overview and a roadmap. *Front. Genet.* *6*, 76.
- Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* *61*, 263–289.
- Schiebenhoefer, H., Schallert, K., Renard, B.Y., Trappe, K., Schmid, E., Benndorf, D., Riedel, K., Muth, T., and Fuchs, S. (2020). A complete and flexible workflow for metaproteomics data analysis based on MetaProteomeAnalyzer and Prophane. *Nat. Protoc.* *15*, 3212–3239.
- Schneider, T., and Riedel, K. (2010). Environmental proteomics: Analysis of structure and function of microbial communities. *PROTEOMICS* *10*, 785–798.
- Schneider, T., Schmid, E., de Castro, J.V., Cardinale, M., Eberl, L., Grube, M., Berg, G., and Riedel, K. (2011). Structure and function of the symbiosis partners of the lung lichen (*Lobaria pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* *11*, 2752–2756.

- Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., Richter, A., Eberl, L., Zechmeister-Boltenstern, S., and Riedel, K. (2012). Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J.* *6*, 1749–1762.
- Schols, H.A., Visser, R.G.F., and Voragen, A.G.J. (2009). *Pectins and pectinases* (Wageningen Academic Publishers).
- Schwalm, N.D., Townsend, G.E., and Groisman, E.A. (2016). Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium *Bacteroides thetaiotaomicron*. *MBio* *7*, e01342-16.
- Schwarz, W. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* *56*, 634–649.
- Seon Park, J., Russell, J.B., and Wilson, D.B. (2007). Characterization of a family 45 glycosyl hydrolase from *Fibrobacter succinogenes* S85. *Anaerobe* *13*, 83–88.
- Serafim, L.S., Lemos, P.C., Albuquerque, M.G.E., and Reis, M.A.M. (2008). Strategies for PHA production by mixed cultures and renewable waste materials. *Appl. Microbiol. Biotechnol.* *81*, 615–628.
- Seshadri, R., Leahy, S.C., Attwood, G.T., Teh, K.H., Lambie, S.C., Cookson, A.L., Eloe-Fadrosch, E.A., Pavlopoulos, G.A., Hadjithomas, M., Varghese, N.J., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* *36*, 359–367.
- Sethupathy, S., Morales, G.M., Li, Y., Wang, Y., Jiang, J., Sun, J., and Zhu, D. (2021). Harnessing microbial wealth for lignocellulose biomass valorization through secretomics: a review. *Biotechnol. Biofuels* *14*, 154.
- Shinkai, T., Mitsumori, M., Sofyan, A., Kanamori, H., Sasaki, H., Katayose, Y., and Takenaka, A. (2016). Comprehensive detection of bacterial carbohydrate-active enzyme coding genes expressed in cow rumen. *Anim. Sci. J. Nihon Chikusan Gakkaiho* *87*, 1363–1370.
- Simmons, C.W., Reddy, A.P., Simmons, B.A., Singer, S.W., and VanderGheynst, J.S. (2014). Effect of inoculum source on the enrichment of microbial communities on two lignocellulosic bioenergy crops under thermophilic and high-solids conditions. *J. Appl. Microbiol.* *117*, 1025–1034.
- Singh, K.M., Reddy, B., Patel, D., Patel, A.K., Parmar, N., Patel, A., Patel, J.B., and Joshi, C.G. (2014). High Potential Source for Biomass Degradation Enzyme Discovery and Environmental Aspects Revealed through Metagenomics of Indian Buffalo Rumen.
- Singh, R., Bennett, J.P., Gupta, M., Sharma, M., Eqbal, D., Alessi, A.M., Dowle, A.A., McQueen-Mason, S.J., Bruce, N.C., and Yazdani, S.S. (2019). Mining the biomass deconstructing capabilities of rice yellow stem borer symbionts. *Biotechnol. Biofuels* *12*, 265.
- Siriwongrungson, V., Zeng, R.J., and Angelidaki, I. (2007). Homoacetogenesis as the alternative pathway for H<sub>2</sub> sink during thermophilic anaerobic degradation of butyrate under suppressed methanogenesis. *Water Res.* *41*, 4204–4210.

- Snelling, T.J., and Wallace, R.J. (2017). The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiol.* *17*, 9.
- Solli, L., Håvelsrud, O.E., Horn, S.J., and Rike, A.G. (2014). A metagenomic study of the microbial communities in four parallel biogas reactors. *Biotechnol. Biofuels* *7*.
- de Souza, W.R., de Gouvea, P.F., Savoldi, M., Malavazi, I., de Souza Bernardes, L.A., Goldman, M.H.S., de Vries, R.P., de Castro Oliveira, J.V., and Goldman, G.H. (2011). Transcriptome analysis of *Aspergillus niger* grown on sugarcane bagasse. *Biotechnol. Biofuels* *4*, 40.
- Speda, J., Jonsson, B.-H., Carlsson, U., and Karlsson, M. (2017). Metaproteomics-guided selection of targeted enzymes for bioprospecting of mixed microbial communities. *Biotechnol. Biofuels* *10*, 128.
- St John, F.J., González, J.M., and Pozharski, E. (2010). Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett.* *584*, 4435–4441.
- Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng. Des. Sel. PEDS* *19*, 555–562.
- Starke, R., Müller, M., Gaspar, M., Marz, M., Küsel, K., Totsche, K.U., von Bergen, M., and Jehmlich, N. (2017). Candidate Brocadiales dominates C, N and S cycling in anoxic groundwater of a pristine limestone-fracture aquifer. *J. Proteomics* *152*, 153–160.
- Starke, R., Pylro, V.S., and Morais, D.K. (2021). 16S rRNA Gene Copy Number Normalization Does Not Provide More Reliable Conclusions in Metataxonomic Surveys. *Microb. Ecol.* *81*, 535–539.
- Stewart, R.D., Auffret, M.D., Warr, A., Walker, A.W., Roehe, R., and Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* *37*, 953–961.
- Su, X.-L., Tian, Q., Zhang, J., Yuan, X.-Z., Shi, X.-S., Guo, R.-B., and Qiu, Y.-L. (2014). *Acetobacteroides hydrogenigenes* gen. nov., sp. nov., an anaerobic hydrogen-producing bacterium in the family Rikenellaceae isolated from a reed swamp. *Int. J. Syst. Evol. Microbiol.* *64*, 2986–2991.
- Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N.N., Mikhailova, N., Chertkov, O., et al. (2011). The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist. *PLoS ONE* *6*, e18814.
- Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* *83*, 1–11.
- Sun, J., Zhang, L., and Loh, K.-C. (2021). Review and perspectives of enhanced volatile fatty acids production from acidogenic fermentation of lignocellulosic biomass wastes. *Bioresour. Bioprocess.* *8*, 68.



- Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., Dalin, A.-M., Muller, E.E.L., Shah, P., Wilmes, P., et al. (2017). Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *ISME J.* *11*, 2538–2551.
- Tanaka, M., Ikesaka, M., Matsuno, R., and Converse, A.O. (1988). Effect of pore size in substrate and diffusion of enzyme on hydrolysis of cellulosic materials with cellulases. *Biotechnol. Bioeng.* *32*, 698–706.
- Tanca, A., Palomba, A., Fraumene, C., Pagnozzi, D., Manghina, V., Deligios, M., Muth, T., Rapp, E., Martens, L., Addis, M.F., et al. (2016). The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome* *4*, 51.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* *4*, 41.
- Tavella, T., Rampelli, S., Guidarelli, G., Bazzocchi, A., Gasperini, C., Pujos-Guillot, E., Comte, B., Barone, M., Biagi, E., Candela, M., et al. (2021). Elevated gut microbiome abundance of Christensenellaceae, Porphyromonadaceae and Rikenellaceae is associated with reduced visceral adipose tissue and healthier metabolic profile in Italian elderly. *Gut Microbes* *13*, 1880221.
- Taxis, T.M., Wolff, S., Gregg, S.J., Minton, N.O., Zhang, C., Dai, J., Schnabel, R.D., Taylor, J.F., Kerley, M.S., Pires, J.C., et al. (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res.* *43*, 9600–9612.
- Templ, M. (2021). Artificial Neural Networks to Impute Rounded Zeros in Compositional Data. In *Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn*, P. Filzmoser, K. Hron, J.A. Martín-Fernández, and J. Palarea-Albaladejo, eds. (Cham: Springer International Publishing), pp. 163–187.
- Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H.J., and Henrissat, B. (2018). PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* *46*, D677–D683.
- Thapa, S., Mishra, J., Arora, N., Mishra, P., Li, H., O'Hair, J., Bhatti, S., and Zhou, S. (2020). Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Rev. Environ. Sci. Biotechnol.* *19*, 621–648.
- Tian, D., Guo, Y., Hu, J., Yang, G., Zhang, J., Luo, L., Xiao, Y., Deng, S., Deng, O., Zhou, W., et al. (2020). Acidic deep eutectic solvents pretreatment for selective lignocellulosic biomass fractionation with enhanced cellulose reactivity. *Int. J. Biol. Macromol.* *142*, 288–297.
- Tilocca, B., Burbach, K., Heyer, C.M.E., Hoelzle, L.E., Mosenthin, R., Stefanski, V., Camarinha-Silva, A., and Seifert, J. (2017). Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome—from days to weeks. *Microbiome* *5*.

- Tokuda, G., Mikaelyan, A., Fukui, C., Matsuura, Y., Watanabe, H., Fujishima, M., and Brune, A. (2018). Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites. *Proc. Natl. Acad. Sci. U. S. A.* *115*, E11996–E12004.
- Tomazetto, G., Pimentel, A.C., Wibberg, D., Dixon, N., and Squina, F.M. (2020). Multi-omic Directed Discovery of Cellulosomes, Polysaccharide Utilization Loci, and Lignocellulases from an Enriched Rumen Anaerobic Consortium. *Appl. Environ. Microbiol.* *86*, e00199-20.
- Torella, J.P., Ford, T.J., Kim, S.N., Chen, A.M., Way, J.C., and Silver, P.A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 11290–11295.
- Van Dyk, J.S., and Pletschke, B.I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnol. Adv.* *30*, 1458–1480.
- Vandenbossche, V., Brault, J., Vilarem, G., Hernández-Meléndez, O., Vivaldo-Lima, E., Hernández-Luna, M., Barzana, E., Duque, A., Manzanares, P., Ballesteros, M., et al. (2014). A new lignocellulosic biomass deconstruction process combining thermo-mechano chemical action and bio-catalytic enzymatic hydrolysis in a twin-screw extruder. *Ind. Crops Prod.* *55*, 258–266.
- Vanwonterghem, I., Jensen, P.D., Dennis, P.G., Hugenholtz, P., Rabaey, K., and Tyson, G.W. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J.* *8*, 2015–2028.
- Vassilev, S.V., Baxter, D., Andersen, L.K., Vassileva, C.G., and Morgan, T.J. (2012). An overview of the organic and inorganic phase composition of biomass. *Fuel* *94*, 1–33.
- Viborg, A.H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., and Brumer, H. (2019). A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* *294*, 15973–15986.
- Villares Garicochea, A., Moreau, C., Bennati-granier, C., GARAJOVA, S., Foucat, L., Falourd, X., Saake, B., Berrin, J.-G., and Cathala, B. (2017). Lytic polysaccharide monoxygenases disrupt the cellulose fibers structure. *Sci. Rep.* *7*, np.
- Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., and Wu, W. (2016). Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* *9*.
- Wang, W., Yan, L., Cui, Z., Gao, Y., Wang, Y., and Jing, R. (2011). Characterization of a microbial consortium capable of degrading lignocellulose. *Bioresour. Technol.* *102*, 9321–9324.
- Wang, W., Zhuang, X., Yuan, Z., Yu, Q., Qi, W., Wang, Q., and Tan, X. (2012). High consistency enzymatic saccharification of sweet sorghum bagasse pretreated with liquid hot water. *Bioresour. Technol.* *108*, 252–257.
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., et al. (2007). Metagenomic and



functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565.

Watanabe, Y., Nagai, F., and Morotomi, M. (2012). Characterization of *Phascolarctobacterium succinatutens* sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces. *Appl. Environ. Microbiol.* 78, 511–518.

Waters, J.L., and Ley, R.E. (2019). The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biol.* 17, 83.

Webster, T.M., Smith, A.L., Reddy, R.R., Pinto, A.J., Hayes, K.F., and Raskin, L. (2016). Anaerobic microbial community response to methanogenic inhibitors 2-bromoethanesulfonate and propynoic acid. *MicrobiologyOpen* 5, 537–550.

Weimer, P.J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Front. Microbiol.* 6, 296.

Weimer, P.J., and Kohn, R.A. (2016). Impacts of ruminal microorganisms on the production of fuels: how can we intercede from the outside? *Appl. Microbiol. Biotechnol.* 100, 3389–3398.

Weimer, P.J., Russell, J.B., and Muck, R.E. (2009). Lessons from the cow: what the ruminant animal can teach us about consolidated bioprocessing of cellulosic biomass. *Bioresour. Technol.* 100, 5323–5331.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27.

Williams, T.R., and Marco, M.L. (2014). Phyllosphere Microbiota Composition and Microbial Community Transplantation on Lettuce Plants Grown Indoors. *MBio* 5, e01564-14.

Wilmes, P., and Bond, P.L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920.

Wilson, D.B. (2009). Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose* 16, 723–727.

Wilson, J.R. (1993). Organization of Forage Plant Tissues. In *Forage Cell Wall Structure and Digestibility*, (John Wiley & Sons, Ltd), pp. 1–32.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090.

Wongwilaiwalin, S., Laothanachareon, T., Mhuantong, W., Tangphatsornruang, S., Eurwilaichitr, L., Igarashi, Y., and Champreda, V. (2013). Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. *Appl. Microbiol. Biotechnol.* 97, 8941–8954.

Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R., and Lee, Y.Y. (2005). Coordinated development of leading biomass pretreatment technologies. *Bioresour. Technol.* 96, 1959–1966.

- Xia, Y., and Sun, J. (2017). Hypothesis testing and statistical analysis of microbiome. *Genes Dis.* *4*, 138–148.
- Xing, M.-N., Zhang, X.-Z., and Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnol. Adv.* *30*, 920–929.
- Xu, N., Zhang, W., Ren, S., Liu, F., Zhao, C., Liao, H., Xu, Z., Jiangfeng, H., Li, Q., Tu, Y., et al. (2012). Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H<sub>2</sub>SO<sub>4</sub> pretreatments in *Miscanthus*. *Biotechnol. Biofuels* *5*, 58.
- Xu, Q., Gao, W., Ding, S.-Y., Kenig, R., Shoham, Y., Bayer, E.A., and Lamed, R. (2003). The Cellulosome System of *Acetivibrio cellulolyticus* Includes a Novel Type of Adaptor Protein and a Cell Surface Anchoring Protein. *J. Bacteriol.* *185*, 4548–4557.
- Xu, R., Zhang, K., Liu, P., Han, H., Zhao, S., Kakade, A., Khan, A., Du, D., and Li, X. (2018). Lignin depolymerization and utilization by bacteria. *Bioresour. Technol.* *269*, 557–566.
- Yang, Q., and Pan, X. (2016). Correlation between lignin physicochemical properties and inhibition to enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.* *113*, 1213–1224.
- Yang, B., Wang, Y., and Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* *17*, 135.
- Yang, L., Xu, F., Ge, X., and Li, Y. (2015). Challenges and strategies for solid-state anaerobic digestion of lignocellulosic biomass. *Renew. Sustain. Energy Rev.* *44*, 824–834.
- Yang, X., Chen, H., Gao, H., and Li, Z. (2001). Bioconversion of corn straw by coupling ensiling and solid-state fermentation. *Bioresour. Technol.* *78*, 277–280.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., and Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* *12*, 635–645.
- Yu, C., Reddy, A.P., Simmons, C.W., Simmons, B.A., Singer, S.W., and VanderGheynst, J.S. (2015). Preservation of microbial communities enriched on lignocellulose under thermophilic and high-solid conditions. *Biotechnol. Biofuels* *8*, 206.
- Yutin, N., and Galperin, M.Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* *15*, 2631–2641.
- Zhang, B., He, P.-J., Lü, F., Shao, L.-M., and Wang, P. (2007a). Extracellular enzyme activities during regulated hydrolysis of high-solid organic wastes. *Water Res.* *41*, 4468–4478.
- Zhang, M., Chekan, J.R., Dodd, D., Hong, P.-Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci.* *111*, E3708–E3717.
- Zhang, S., Xu, Y., and Hanna, M.A. (2012). Pretreatment of Corn Stover with Twin-Screw Extrusion Followed by Enzymatic Saccharification. *Appl. Biochem. Biotechnol.* *166*, 458–469.

- Zhang, Y., Gao, W., and Meng, Q. (2007b). Fermentation of plant cell walls by ruminal bacteria, protozoa and fungi and their interaction with fibre particle size. *Arch. Anim. Nutr.* *61*, 114–125.
- Zhao, J., and Chen, H. (2013). Correlation of porous structure, mass transfer and enzymatic hydrolysis of steam exploded corn stover. *Chem. Eng. Sci.* *104*, 1036–1044.
- Zhao, Y., Wang, Y., Zhu, J. y., Ragauskas, A., and Deng, Y. (2008). Enhanced enzymatic hydrolysis of spruce by alkaline pretreatment at low temperature. *Biotechnol. Bioeng.* *99*, 1320–1328.
- Zheng, J., and Rehmann, L. (2014). Extrusion pretreatment of lignocellulosic biomass: a review. *Int. J. Mol. Sci.* *15*, 18967–18984.
- Zhivin, O., Dassa, B., Moraïs, S., Utturkar, S.M., Brown, S.D., Henrissat, B., Lamed, R., and Bayer, E.A. (2017). Unique organization and unprecedented diversity of the Bacteroides (Pseudobacteroides) cellulosolvans cellulosome system. *Biotechnol. Biofuels* *10*, 211.
- Zhou, J., and Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiol. Mol. Biol. Rev.* *81*, e00002-17.
- Zhou, J., Liu, W., Deng, Y., Jiang, Y.-H., Xue, K., He, Z., Van Nostrand, J.D., Wu, L., Yang, Y., and Wang, A. (2013). Stochastic Assembly Leads to Alternative Communities with Distinct Functions in a Bioreactor Microbial Community. *MBio* *4*.
- Zhou, M., Yan, B., Wong, J.W.C., and Zhang, Y. (2018). Enhanced volatile fatty acids production from anaerobic fermentation of food waste: A mini-review focusing on acidogenic metabolic pathways. *Bioresour. Technol.* *248*, 68–78.
- Zhu, N., Yang, J., Ji, L., Liu, J., Yang, Y., and Yuan, H. (2016). Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. *Biotechnol. Biofuels* *9*.
- Zhu, Y., Lee, Y.Y., and Elander, R.T. (2005). Optimization of dilute-acid pretreatment of corn stover using a high-solids percolation reactor. *Appl. Biochem. Biotechnol.* *121–124*, 1045–1054.
- Zou, Y., Xue, W., Lin, X., Hu, T., Liu, S.-W., Sun, C.-H., Luo, G., Lv, M., Dai, Y., Kristiansen, K., et al. (2021). Taxonomic Description and Genome Sequence of *Christensenella intestinhominis* sp. nov., a Novel Cholesterol-Lowering Bacterium Isolated From Human Gut. *Front. Microbiol.* *12*, 632361.
- Zybailov, B., Mosley, A.L., Sardi, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* *5*, 2339–2347.

