ISME

www.nature.com/ismecomms

ARTICLE OPEN

Horizontal metaproteomics and CAZymes analysis of lignocellulolytic microbial consortia selectively enriched from cow rumen and termite gut

Emeline Auer[1], Adèle Lazuka[1], Bertrand Huguenin-Bizot[1], Nico Jehmlich[2], Sébastien Déjean[3], Vincent Lombard[4,5], Bernard Henrissat[4,5,6], Michael O'Donohue[1] and Guillermina Hernandez-Raquet[1]✉

Selectively enriched microbial consortia are potentially useful for the conversion of lignocellulose (LC) into biofuels and commodity chemicals. Consortia are also of interest to elucidate the roles of individual microorganisms and the dynamics of enzymes involved in LC deconstruction. Using metaproteomics, 16 S rRNA gene amplicon sequencing and multivariate discriminant analysis, we revealed the temporal dynamics of microbial species and their proteins during anaerobic conversion of LC by microbial consortia derived from cow rumen (RWS) and termite gut (TWS) microbiomes. Bacteroidetes (Bacteroidota), Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) phyla were dominant, irrespective the inoculum origin, displaying functional complementarities. We identified a large variety of carbohydrate-active enzymes, distributed in 94 CAZy families, involved in biomass deconstruction. Additionally, proteins involved in short chain fatty acids biosynthesis were detected. Multivariate analysis clearly differentiates RWS and TWS metaproteomes, with differences originating in the initial inoculates. Further supervised discriminant analysis of the temporal succession of CAZymes revealed that both consortia consume easily accessible oligosaccharides during the early stage of incubation, degrading more complex hemicellulose and cellulose fractions at later stages, an action that pursues throughout the incubation period. Our results provide new insights regarding the functional roles and complementarities existing in lignocellulolytic consortia and highlight their potential for biorefinery applications.

ISME Communications; https://doi.org/10.1038/s43705-023-00339-0

INTRODUCTION

Lignocellulose (LC) is the major component of plant cell walls and the primary source of renewable carbon on Earth available for a sustainable production of biofuels and commodity chemicals. However, biomass bioconversion is very challenging due to LC recalcitrance [1]. Therefore, to improve the economic feasibility of biomass biorefining, more effective strategies are required to break down LC.

In natural ecosystems, LC decomposition is performed by complex microbial communities producing large enzyme arsenals. These communities includes the ones involved in the composting processes [2, 3], cellulose and leaf-litter decomposition in soils [4–6], and the digestive systems of herbivores [7] and termites [8, 9]. In the case of herbivores, LC decomposition is the result of the symbiosis between the animal host and its digestive microbiome, the latter providing short-chain volatile fatty acids (VFA) and microbial proteins to the host. Therefore, to improve the economic feasibility of biomass biorefining, one strategy is to develop Nature-inspired solutions by harnessing the power of naturally occurring LC-degrading microbiomes. However, the challenge is to deploy such microbial ecosystems in controlled bioreactor environments.

LC deconstruction is a complex process that marshals a large diversity of enzymes rarely produced by a single microbial species [10]. The process involves carbohydrate-active enzymes (CAZymes) that break down, modify or assemble glycans [11]. CAZymes are classified 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) [11]. Other non-catalytic protein domains include cohesins (COH), dockerins (DOC) and S-layer homology (SLH) domains. These are often appended to CAZymes and provide the means to incorporate cellulosome complexes that link enzymes to the cell surfaces [12]. Microorganisms present in LC degrading environments such as the digestive tract of termites or cow rumen provide vast reservoirs of CAZymes, which are sources of new potent lignocellulolytic enzymes for biomass deconstruction [13].

To better understand the biological mechanisms that underpin LC degradation, meta-omics technologies are being used to

[1]Toulouse Biotechnology Institute - TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, F-31077 Toulouse, France. [2]Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research-UFZ, Permoserstrasse 15, 04318 Leipzig, Germany. [3]Toulouse Mathematics Institute, UMR 5219 CNRS, Toulouse University, Toulouse, France. [4]CNRS UMR 7257, Aix-Marseille University, F-13288 Marseille, France. [5]USC 1408 AFMB, INRAE, F-13288 Marseille, France. [6]Present address: Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia and DTU Bioengineering, Technical University of Denmark, DK-2800, Kgs Lyngby, Denmark.
✉email: hernandg@insa-toulouse.fr

Received: 7 July 2023 Revised: 15 November 2023 Accepted: 22 November 2023
Published online: 04 January 2024

Table 1. RWS and TWS sample characteristics

	Dry matter degradation (%)	VFA (mgC/L)	Acetate:Propionate:Butyrate (C molar ratio)	Xylanase (mUA/L)
RWS-1[a] T1	7.6 ± 7.1	552.20 ± 23.1	57:24:18	174.6 ± 22.8
RWS-2[a] 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

[a]1 and 2 indicate the biological replicate bioreactors. Data are means ± SD.

investigate how microbial and enzymatic synergies contribute to LC deconstruction [14]. 16 S rRNA gene amplicon sequencing and shotgun metagenomics have provided information on both the taxonomic and metabolic potential of various lignocellulolytic microbial communities [7, 10, 15]. Moreover, they have shed light on the synergies between the members of these communities [2, 4, 16]. 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. For this, metaproteomics is a precious tool since it provides information on the entire protein component of microbial communities, links proteins to specific microbial taxa and correlates their presence with metabolic activity [17]. This powerful approach thus provides useful information on metabolic networks and symbiotic microbial interactions.

Previous metaproteomics studies on rumen ecosystems have shown that the most abundant proteins are affiliated to Bacteroidetes (Bacteroidota), Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) [18], where LC degradation is associated with high redundancy of key enzyme activities. Regarding the termite gut microbiome, metaproteomics of *Nasutitermes corniger* showed that among 197 identified proteins with known functions, 48 proteins are directly related to glycan hydrolysis [19]. However, 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 dynamics of proteins involved in LC breakdown.

To gain insight into how lignocellulolytic ecosystems function, selective ecosystem enrichment is often used to reduce microbial complexity and hone community functions for use in bioprocesses [20–23]. However, most previous metaproteomic studies performed on simplified ecosystems only revealed a small number of proteins [23–27]. They also fail to fully capture the temporal dynamics of microbial species and CAZymes, although one previous study has provided insight into the protein expression dynamics of a subset of enzymes [28]. To further elucidate the mechanisms employed by microbial consortia to decompose LC, it is thus vital to fully capture the temporal dynamics of all active species and expressed proteins. Using LC as the sole carbon source, we postulate that inoculum-specific microbial species can be maintained, although we expect that time-dependent enzyme profiles to vary depending as a function of substrate modifications occurring during the degradation process. Nevertheless, we also expect all enzyme profiles will display genericity, regardless of the source of the inoculum.

In previous studies, we reported the selective enrichment of two anaerobic lignocellulolytic microbial consortia derived from cow rumen (RWS) and from the termite gut microbiome of *Nasutitermes ephratae* (TWS) [20, 21]. These naturally-occurring anaerobic microbiomes present significant and contrasting levels of diversity, and display great potential for LC degradation [7, 8, 13]. The enrichment of these microbiomes by a sequential batch reactor process, using wheat straw as sole carbon source, resulted in consortia displaying high LC-degradation activity and good ability to produce carboxylates (mainly VFAs with acetate, butyrate and propionate as main products). These products are valuable chemicals for producing bioplastics and liquid biofuels [29, 30]. The kinetic characteristics of these enriched consortia were determined by measuring LC-degradation rate, xylanase activity and carboxylate production. Herein, we expand on previous work, performing shotgun metaproteomic analysis over the reaction period of wheat straw hydrolysis by RWS and TWS consortia.

MATERIALS AND METHODS

Lignocellulose degradation by RWS and TWS microbial consortia

The kinetic behaviors of the enriched lignocellulolytic microbial consortia derived from cow rumen (RWS) and termite gut (TWS), summarized in Table 1 and Figure S1, have been described previously [20, 21]. For each consortium, two identical anaerobic bioreactors were carried out, using a mineral medium containing wheat straw as the sole carbon source (20 $g.L^{-1}$). Bioreactors were operated for 15 days at 35 °C under stirring (400 rpm) and pH control (6.15), as detailed in Supplementary Material and Methods.

The temporal dynamics of species and expressed proteins in RWS and TWS along LC degradation were assessed by 16 S rRNA gene sequencing and shotgun metaproteomics performed on four time points for each bioreactor. Time point selection was based on wheat straw degradation, VFA production and xylanase activity profiles (Table 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) correspond with the start and end of maximal xylanase activity and peak lignocellulose degradation rate. The fourth time point (T4) captures the final phase, when wheat straw degradation, VFA production and xylanase activity stagnate.

Microbial diversity analysis

Total DNA/RNA were extracted from the pellet fraction of 1.5 mL samples (13,000 x *g*, 5 min, 4 °C) using the PowerMicrobiome isolation kit (Qiagen, Courtaboeuf, France) according to the manufacturer's instructions. After DNA purification (AllPrep DNA/RNA MiniKit, Qiagen), the hypervariable V3-V4 region of the 16 S rRNA gene was amplified by Illumina MiSeq sequencing (GenoToul Genomics and Transcriptomics platform, Toulouse, France) using the conditions and primers previously described [31]. Sequencing data was analyzed using Find Rapidly OTUs with Galaxy solution (FROGS) pipeline [32], as detailed in Supplementary Information. R CRAN software (v4.0.0) was used for further analysis. Diversity metrics were obtained using R's Phyloseq package v1.32.0 [33]. Sequencing data were deposited to the Sequence Read Archive (SRA) under accession number PRJNA729464.

Metaproteomics analysis

Protein extraction, peptides digestion and mass spectrometry analysis. Protein extraction was carried out on 3 technical replicates per sample (3 mL), using a phenol buffer following the procedure for complex sediment samples [34] and separated by SDS-PAGE. After trypsin proteolysis (Promega, Fitchburg, WI, USA) and purification of peptides (ZipTips, C18, Merck, Millipore, Billerica, MA, USA), the dried samples were stored at −20 °C.

Mass spectrometry (MS) was performed on a Q Exactive HF MS (Thermo Fisher Scientific, Waltham, MA, USA) with a TriVersa NanoMate (Advion, Ltd., Harlow, UK) source in liquid chromatography chip coupling mode. Peptide separation settings were as in procedures previously described [35]. MS data have been deposited into the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the proteomics identification (PRIDE) partner repository [36]. Peptide identification was performed with the Thermo Proteome Discoverer software (v1.4; Thermo Fisher Scientific, Waltham, MA, USA) using Sequest HT against bacterial sequences only (plant and eukaryotic sequences excluded) of Uniprot-TrEMBL database (release date of 6th April 2016) [37]. Peptide identification considered a false discovery rate (FDR) below 1% calculated by Percolator [38], a minimum length of six amino acids, and a peptide rank of one. Protein matches were only accepted if they were identified by a minimum of one unique peptide and a high confidence. "Protein Groups" (hereafter referred to as proteins) were form with strict parsimony and using the highest scoring protein in the group as the confident master candidate protein. The detailed protocol is included in Supplementary Material and Methods.

Protein abundance, taxonomic and functional annotations. PROPHANE pipeline (www.prophane.de) [39] was used for protein identification and taxonomic annotation using, respectively, the highest sequence similarity to the UniProtKB/TrEMBL database and BlastP considering 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 [40]. 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) [41], considering the first hit for each protein (e-value ≤ 0.01), and the Hidden Markov Model profiles with HMMER3 algorithm, considering the first hit for each protein (gathering cut-off) [42]. Only identified proteins were retained for further analysis. For each sample, PROPHANE estimated protein abundance based on the normalized spectral abundance factor (NSAF) [43]. Replicate-to-replicate variation was assessed by Pearson correlation analysis using the cor R function, accepting a minimum value of 0.7. Proteins present in only one technical replicate were discarded; remaining proteins were expressed as mean values. Computational assignment of protein functions were carefully checked, completed, and manually curated.

CAZymes annotation. The identified proteins were assigned to CAZyme families following the day-to-day updates procedure of the CAZy database (accessed March 30, 2018) as described previously [44, 45]. Briefly, 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 ancillary) and a HMMER3 search against a collection of hidden Markov models based on each CAZy family [45]. Presence of signal peptides in CAZymes was predicted by SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/).

Data analysis and visualization. Hierarchical clustering (Ward method with Euclidean distance) was used to group the samples with *dist* and *hclust* R functions. Plots were constructed using ggplot2 package v3.3.2. Relative abundances of bacterial OTUs or proteins were aggregated at a target taxonomic level for stacked bar plot representation. Low abundance taxa (<1% Phylum and <2% Genus level), were gathered as "Other". A unique protein affiliated to *Chitinispirillum*, was excluded from the analysis.

Statistical significance of differences was determined using Wilcoxon tests with Benjamini-Hochberg p-value correction for pairwise comparisons [46]. For multivariate analysis, centered log-ratio (CLR) transformation of variables [47] was performed to take into account the compositional properties of our data. A correction factor equivalent to 70% of the lowest value of each variable was applied to eliminate zero values observed in the dataset [48, 49]. Microbial species and CAZymes that best discriminate consortia and incubation time were investigated by principal component analysis (PCA) and identified by multivariate integrative partial least squares discriminant analysis (MINT-PLS-DA) performed with factoextra v1.0.7 and mixOmics v6.12.2 packages [50, 51]. Discriminant factors were validated with permutational multivariate analysis of variance (PERMANOVA), using R's vegan package v2.5–6 to test the statistical significant differences [52].

RESULTS

Data overview

16 S rRNA gene sequencing of the lignocellulolytic consortia RWS and TWS generated 800,329 high quality reads (Supp. data 1); rarefying of samples to 15,000 sequences captured most of the microbial diversity (Fig. S2). Metaproteomic analysis yielded a total of 10,342 proteins (Supp. data 1), with high similarity between technical replicates (Pearson correlation >0.7; Table S1) and similar coverage, with an average of 2,792 proteins per sample (Table S2). 33.7% of total proteins were identified in both RWS and TWS consortia (proteins with same accession numbers), while 35.4% and 30.9% of proteins were specific to RWS and TWS, respectively. Hierarchical clustering of overall proteins of RWS and TWS revealed consortia-specific profiles (Fig. S3), evolving as a function of LC degradation.

Taxonomic and functional profiles of RWS and TWS throughout wheat straw degradation

Taxonomy deduced from 16 S rRNA gene sequencing data showed that TWS displayed a lower richness and Shannon diversity index. However, the Shannon diversity estimated from metaproteomics data [53] were very close for both consortia (Table S2). Indeed, a similar taxonomic composition for the two consortia was revealed by the taxonomic affiliation derived from both approaches, with a dominance of Bacteroidetes (Bacteroidota)(> 60%), Firmicutes (Bacillota) (about 20%) and Proteobacteria (Pseudomonadota) (about 10%) phyla (Fig. S4). A minor fraction of both communities belonged to Spirochaetes, while Fibrobacteres (Bacillota) was exclusively found in RWS.

Affiliation of proteins at the genus level revealed that the active populations, contributing most to protein production, belonged to *Bacteroides* in all samples of both consortia (Fig. 1A), these remaining abundant throughout the experiment. In second place was *Clostridium*, but its abundance decreased over time. Multivariate PCA analysis of the abundance of all taxa-affiliated proteins, clearly differentiated RWS and TWS communities (Fig. 1B). This difference was confirmed by a Wilcoxon test (Fig. S5) showing a higher abundance of proteins belonging to *Dysgonomomas, Fibrobacter* and *Enterococcus* in RWS while TWS showed a stronger abundance of proteins affiliated to *Clostridium, Lachnoclostridium* and *Bacteroides*, and to minor genera belonging to Firmicutes-

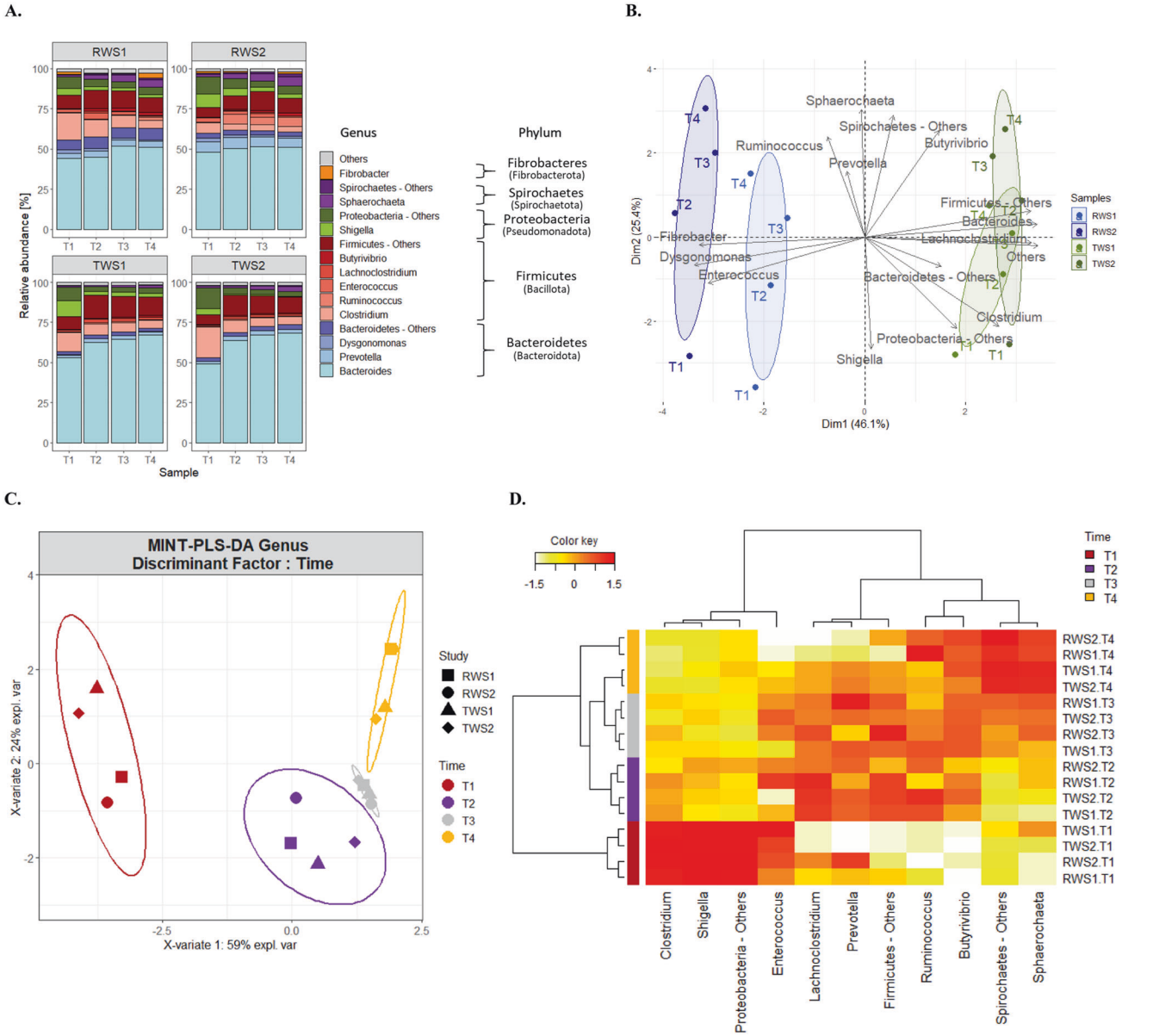

Fig. 1 Community composition of RWS and TWS and its temporal dynamics. A Taxonomic community composition at genus level deduced from the identified proteins (metaproteomics data) for RWS (top graphs) and TWS (bottom graphs) (1 and 2 indicate the biological duplicates). Relative abundance of proteins 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" gathers 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 (Bacteroidota) (blue), Firmicutes (Bacillota) (red); Proteobacteria (Pseudomonadota) (green), Spirochaetes (Spirochaetota) (purple), Fibrobacteres (Fibrobacterota) (orange). **B** Impact of the inoculum origin evidenced by Principal components analysis (PCA) ordination; PCA component 1 and 2 explained respectively 46% and 25.4% of the total variance and **C**. Impact of the incubation time in the community dynamic of RWS and TWS metaproteomes, assessed by Multivariate Integrative Partial Least Square Discriminant Analysis (MINT-PLS-DA) based on the abundance of taxa deduced from proteins (genus level, CLR-transformed data). MINT-PLS-DA component 1 and 2 explained 59% and 24% of the total variance. Ellipses at 95% confidence. **D** Clustered Image Map (CIM) represented the most discriminant genera of the different sampling times for RWS and TWS. CIM was built using the genera contributing the most to the two first MINT-PLS-DA dimensions. Hierarchical clustering was derived using the Euclidean distance and Ward methodology. Genera are represented in columns and samples in rows. The boxes on the left highlights the clusters discriminating the sampling points (yellow, gray, purple and dark-red). The abundance of proteins affiliated to each genus is indicated by the white-to-red color gradient (increasing values).

Others and Proteobacteria-Others. The temporal dynamics of the main taxa of both consortia was evidenced by MINT-PLS-DA analysis, using the incubation time as the discriminant factor. This revealed three clusters formed by T1, T2/T3 and T4 samples (Fig. 1C). T1 samples showed higher abundance of proteins affiliated to *Clostridium*, *Shigella, Enterococcus* and Proteobacteria-Others (Fig. 1D) whose abundances decreased with incubation time. Samples T2/T3 showed higher abundances of proteins affiliated to *Lachnoclostridium*, *Prevotella* and Firmicutes-Others while in samples T4, proteins affiliated to *Sphaerochaeta* and Spirochaetes showed higher abundance, which increased over the incubation period. Proteins affiliated to *Ruminococcus* and

Butyrivibrio were also abundant in T4 samples, but also in T2 and T3 samples. PERMANOVA analysis (Table S3) confirmed that inoculum source (R-squared value 0.6214) and incubation time (R-squared value 0.2022) were the key factors explaining the protein-taxonomic composition and dynamics for both consortia.

The status of lignocellulose-related functions of RWS and TWS, revealed by COGs prediction from the protein dataset, revealed that the cellular function "metabolism" was highly represented (about 45%) (Fig. S6A), with a strong representation of the subrole "carbohydrate transport and metabolism" (about 15%), irrespective of the sampling time and consortia. Bacteroidetes (Bacteroidota) made the greatest contribution to proteins involved in this function (Fig. S6B), while Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) provided a lesser contribution to this function. Nevertheless, the normalization of proteins abundance at the phylum level, revealed that the specific COG profile of Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) members was particularly focused on the "carbohydrate transport and metabolism" function (Fig. S6C), whereas Bacteroidetes (Bacteroidota) species were involved in a wider range of functions.

Carbohydrate-active enzymes in RWS and TWS consortia

RWS and TWS expressed a large diversity of CAZymes involved in plant cell wall degradation, with 423 proteins containing at least one CAZyme domain (Supp. data 2). *In silico* prediction of signal peptides showed 222 CAZymes with a signal peptide (Supplementary data 1). 174 CAZymes (41%), distributed in 69 families, where common to both consortia (Fig. S7A), accounting for about 80% of the CAZymes abundance while 127 and 122 proteins and 15 and 10 CAZy families were exclusively found in RWS or TWS, respectively. Most of these proteins contained GH domains, representing more than 70% of the CAZyme domains detected, followed by CE domains (12–20%), CBMs (9.5–12.5%), GT domains (5–8%) and PL domains (about 2%). These CAZyme domains showed a higher average abundance in TWS than in RWS (Fig. S7B). The main purveyor of CAZymes in both consortia were Bacteroidetes (Bacteroidota) members, producing 236 proteins corresponding to about 80% of the CAZyme abundance (Fig. S7C) followed by Firmicutes (Bacillota) (138 proteins, average 15%) while Proteobacteria (Pseudomonadota) expressed a minor fraction of them (24 proteins, <1.5%).

The most highly represented CAZymes are typically active on hemicellulose and mainly affiliated to Bacteroidetes (Bacteroidota) (Fig. 2A), except those classified GH11. These proteins are putative hemicellulases with endo-β−1,4-xylanase (GH8, GH10, GH11) and β-mannanase (GH26) activities (Supplementary data 2). Hemicellulose-oligosaccharide-degrading enzymes, with mainly β-xylosidase or α-L-arabinofuranosidase functions (GH43), and hemicellulose-debranching enzymes with α-L-arabinofuranosidase (GH51) and xylan α−1,2-glucuronidase (GH67, GH115) activities were also abundant. Cellulose-degrading enzymes were mainly annotated as β-glucosidases (GH3) of Bacteroidetes (Bacteroidota) origin, while those with endoglucanase activity mainly belonged to Firmicutes (Bacillota)(GH5, GH9, GH48) (Fig. 2A). Some of these proteins, were appended to CBMs or cellulosome (COH and DOC) domains. Among the most prevalent proteins in both consortia were Bacteroidetes (Bacteroidota)- and Firmicutes (Bacillota)- affiliated proteins related to starch degradation and including α-galactosidase, 1,4-α-glucan branching enzyme, α-glucosidase and α-amylase activities. According to our data, Proteobacteria (Pseudomonadota) were only minor CAZymes contributors, being mostly sources of acetyl-xylan esterase activity (CE1; Fig. 2B). Non-catalytic carbohydrate binding modules (CBM), SLH, COH and DOC (*i.e.*, cellulosome components) were mainly Firmicutes (Bacillota) origin (Fig. 2B).

CAZymes involved in lignocellulose degradation

Multivariate statistical analyses were used to investigate how both the inoculum origin and progression of LC degradation influenced the expressed CAZyme profile. The PCA clearly separated RWS and TWS samples in function of the inoculum (Fig. S8A, first component) while the initial time points (T1) were clearly separated from the others (Fig. S8B, second component). PERMANOVA analysis (Table S4) confirmed that the inoculum source and incubation time (R-square values of 0.253 and 0.284, respectively) were the key parameters determining the CAZyme expression profiles. The importance of incubation time was confirmed by Multigroup Supervised Partial Least Squares Discriminant Analysis (MINT-PLS-DA), using the incubation time as the discriminant factor. This analysis separated the initial point (T1) from the subsequent sampling points (T2/T3) and from the final point (T4) (Fig. 3A). T1 samples (Fig. 3B, dark blue box) clustered CAZymes with oligosaccharide-degrading activities linked to cellulose (GH1) and hemicelluloses (GH43), pectin methylesterase (CE8), and α-glucan degradation (GH4, GH13, GH31); they also included starch-specific CBMs (CBM20, CBM25, CBM26). Other CAZymes with hemicellulose-oligosaccharide degradation (GH2 and GH29) or polysaccharide lyase (PL9) activities, were also found in T1 samples (Fig. 3B, light-blue box). However, this enzyme cluster was also present (at a lower abundance) at the end of the incubation period (T4). The cluster formed by the sampling points T2 and T3 (Fig. 3B, purple box), where most of the LC degradation occurred, was characterized by high abundance of endoglucanases (GH5, GH9 and GH48) and a large panel of hemicellulolytic CAZymes, including endo-hemicellulases (GH11), deacetylating (CE1, CE11) and debranching (GH67, GH115) enzymes and enzymes active on hemicellulose-oligosaccharides (GH30, GH27). CAZymes related to pectin-oligosaccharide degradation (GH105) were also found, as well as xylan- or amorphous cellulose-specific CBMs (CBM3, CBM4, CBM13, CBM22, CBM36), dockerins typical of cellulosome structures, and glucosyl/galactosyltransferases (GT4, GT5). This cluster of CAZymes was also highly abundant at the latter stage of LC degradation (T4; Fig. 3B, dark-green box) along with a range of enzymes including endo-hemicellulases (GH10) and xylan deacetylating (CE6) or hemicellulose-oligosaccharides degrading enzymes (GH36, GH125) (Fig.3B, light-green box). CBMs specific for cellulose (CBM30), hemicellulose (CBM11) and starch (CBM48), and starch phosphorylases (GT20, GT28) completing this cluster. Similar results were observed when the analysis was performed for CAZymes with predicted signal peptide only (Fig. S9).

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

Lignocellulose degradation by RWS and TWS, like in other microbial consortia, is associated with VFA production [40, 54]. RWS and TWS produced mainly acetate, propionate and butyrate with a higher proportion of the latter in TWS (Table 1; average molar ratio of acetate:propionate:butyrate of 59:23:18 for RWS and 47:20:33 for TWS). Based on previous COG assignments [16, 55, 56], metaproteomic data revealed that the key enzymes involved in VFA biosynthesis constituted 2.8 ± 0.4% and 3.5 ± 0.2% of total protein abundance in RWS and TWS, respectively. According to the higher butyrate production measured in TWS, proteins related to butyrate biosynthesis (219 proteins) were more abundant in this consortium (Figure S8). In both consortia, this function was associated with a large variety of enzymes (numbers of proteins 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 (Bacillota) (138 proteins) and Bacteroidetes (Bacteroidota) (23 proteins), but with abundance levels similar for both phyla (Fig. 4). At the initial phase of incubation (T1), *Clostridium* species were the major contributors of butyrate biosynthesis enzymes in both

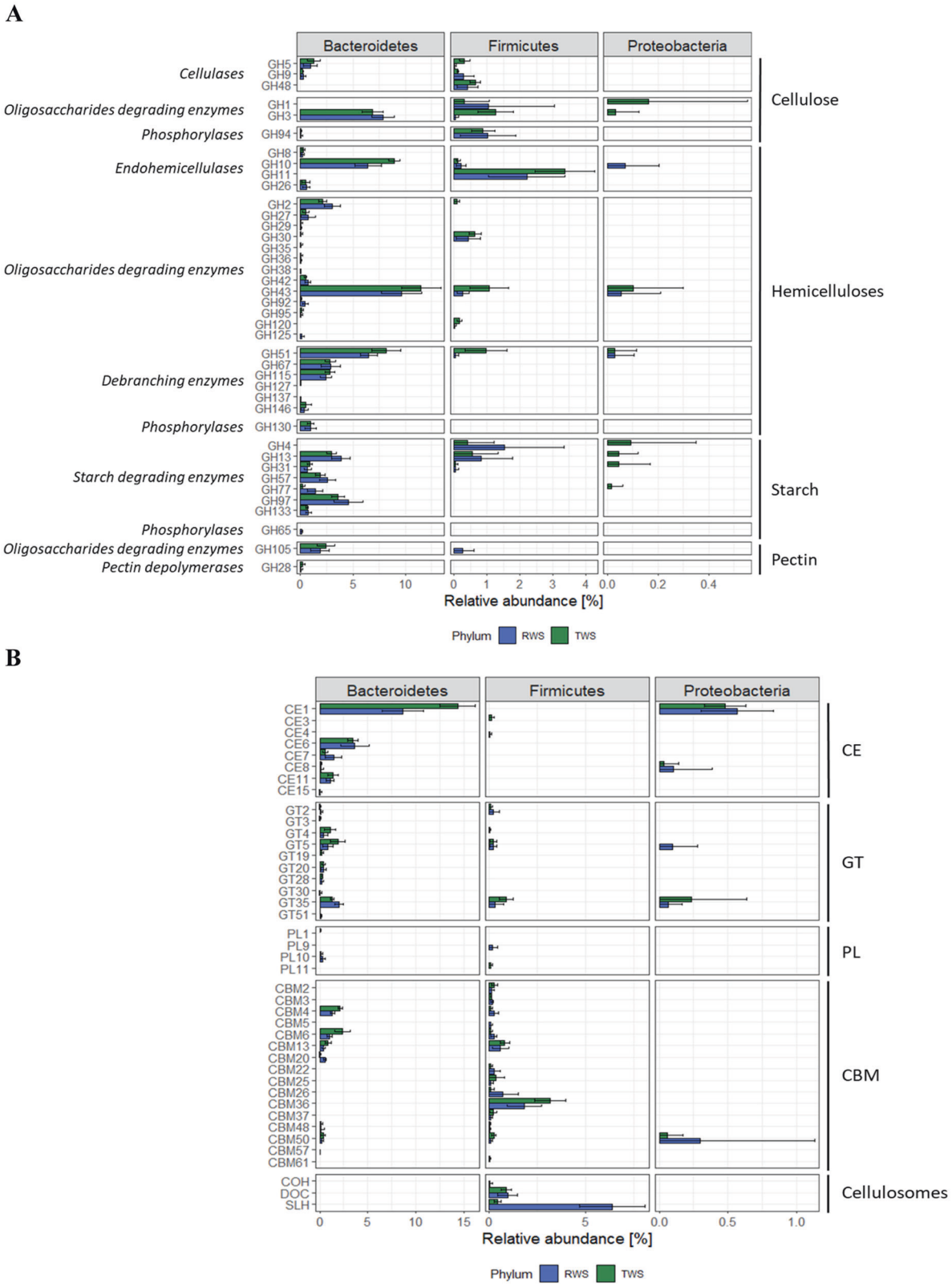

Fig. 2 Distribution of CAZyme families in the main phyla found in RWS and TWS. A Relative abundance of glycoside hydrolases (GH) in Bacteroidetes (Bacteroidota), Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) phyla of RWS and TWS metaproteomes. Abundances were normalized to total CAZymes expression. Only GH targeting cellulose, hemicelluloses, starch and pectin fractions are showed. Error bars represent the standard deviation. **B** Relative abundance of carbohydrate esterases (CE), glycosyl transferases (GT), polysaccharide lyases (PL), carbohydrate-binding modules (CBM) and cellulosomes domains (S-layer homology (SLH), dockerins (DOC), cohesins (COH)) present in Bacteroidetes (Bacteroidota), Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota) phyla in RWS and TWS metaproteomes. Abundances were normalized to total CAZymes expression. Error bars represent the standard deviation.

consortia (67 proteins, about 30% abundance), but 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.

Propionate production was similar in both consortia (Table 1) and appeared to be associated with the presence of 66 propionate biosynthesis-related proteins with three main functions: acetyl/propionyl-CoA carboxylase (34 proteins), methylmalonyl-CoA mutase (23 proteins) and epimerase (9 proteins), revealing that propionate was formed via the succinate pathway in both consortia. These proteins were mainly affiliated to Bacteroidetes (Bacteroidota), particularly to *Bacteroides* (37 from 66 proteins) (Fig. 4), while in Firmicutes (Bacillota), *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 the least abundant among the VFA-biosynthesis enzymes (61 proteins accounting for about 0.3% of the total protein abundance; Fig. S10). The presence of acetate kinases (26 proteins), phosphate acetyltransferases (22 proteins), and aldehyde dehydrogenases (13 proteins), expressed by all phyla (Fig. 4), suggests that acetate production in the consortia occurrs via the Wood-Ljungdahl pathway. Seven of these proteins affiliated to *Bacteroides* formed

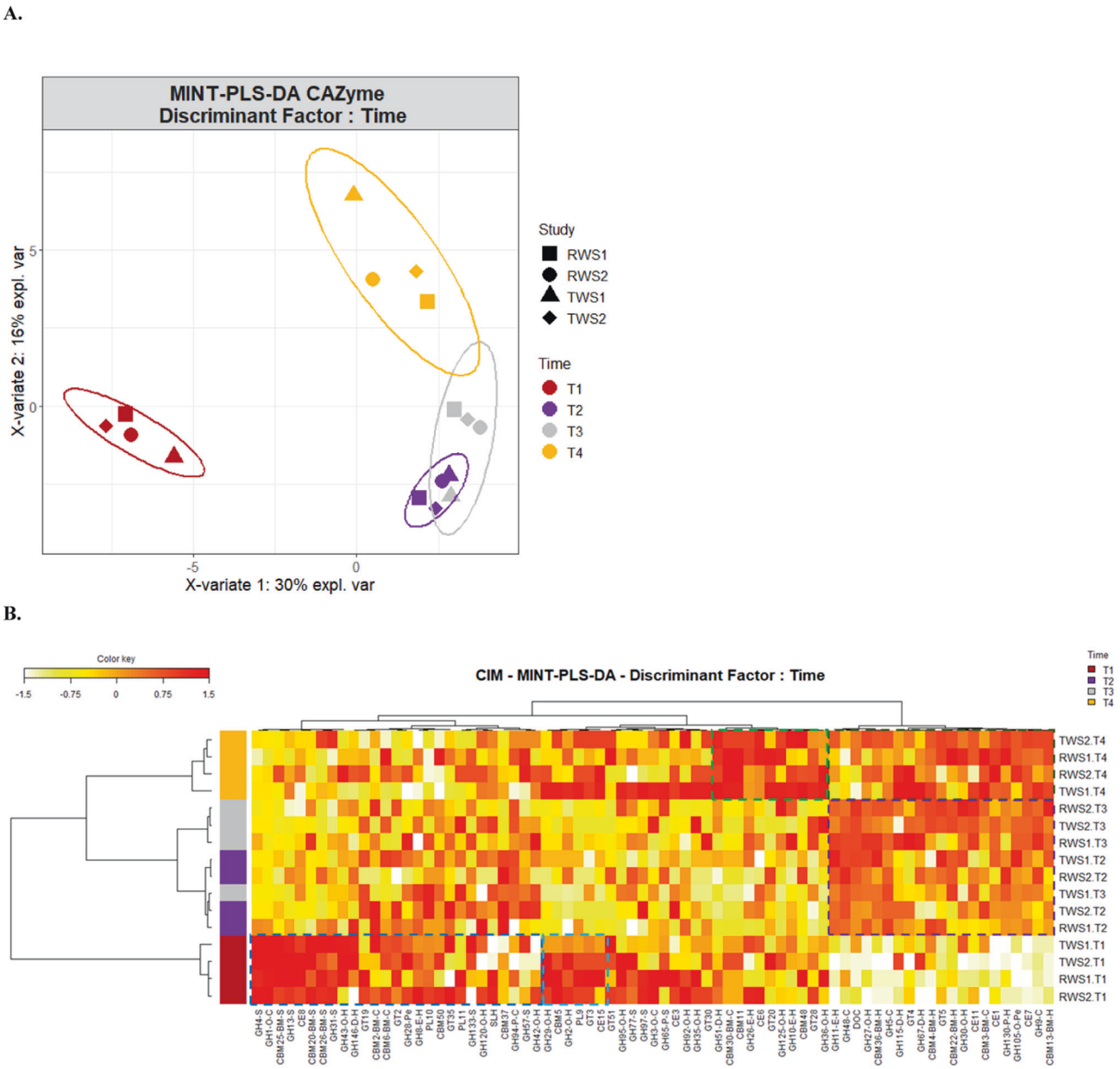

Fig. 3 Temporal dynamics of CAZymes in RWS and TWS metaproteomes. A Impact of the incubation time in the CAZyme dynamics of RWS and TWS metaproteomes, assessed by Multivariate Integrative Partial Least Square Discriminant Analysis (MINT-PLS-DA) based on CAZyme abundances (family level, CLR-transformed data). MINT-PLS-DA component 1 and 2 explained 30% and 16% of the total variance. Ellipses at 95% confidence. **B** Clustered Image Map (CIM) represented the most discriminant CAZyme families of the different sampling times in RWS and TWS. CIM was built using the main CAZyme families explaining the first two MINT-PLS-DA dimensions. Hierarchical clustering (Euclidean distance and Ward method) represents CAZymes in columns and samples (T1–T4) in rows. The boxes on the left highlights the clusters discriminating the sampling points (yellow, gray, purple and dark-red). The abundance of CAZy families is indicated by the white-to-red color gradient (increasing values).

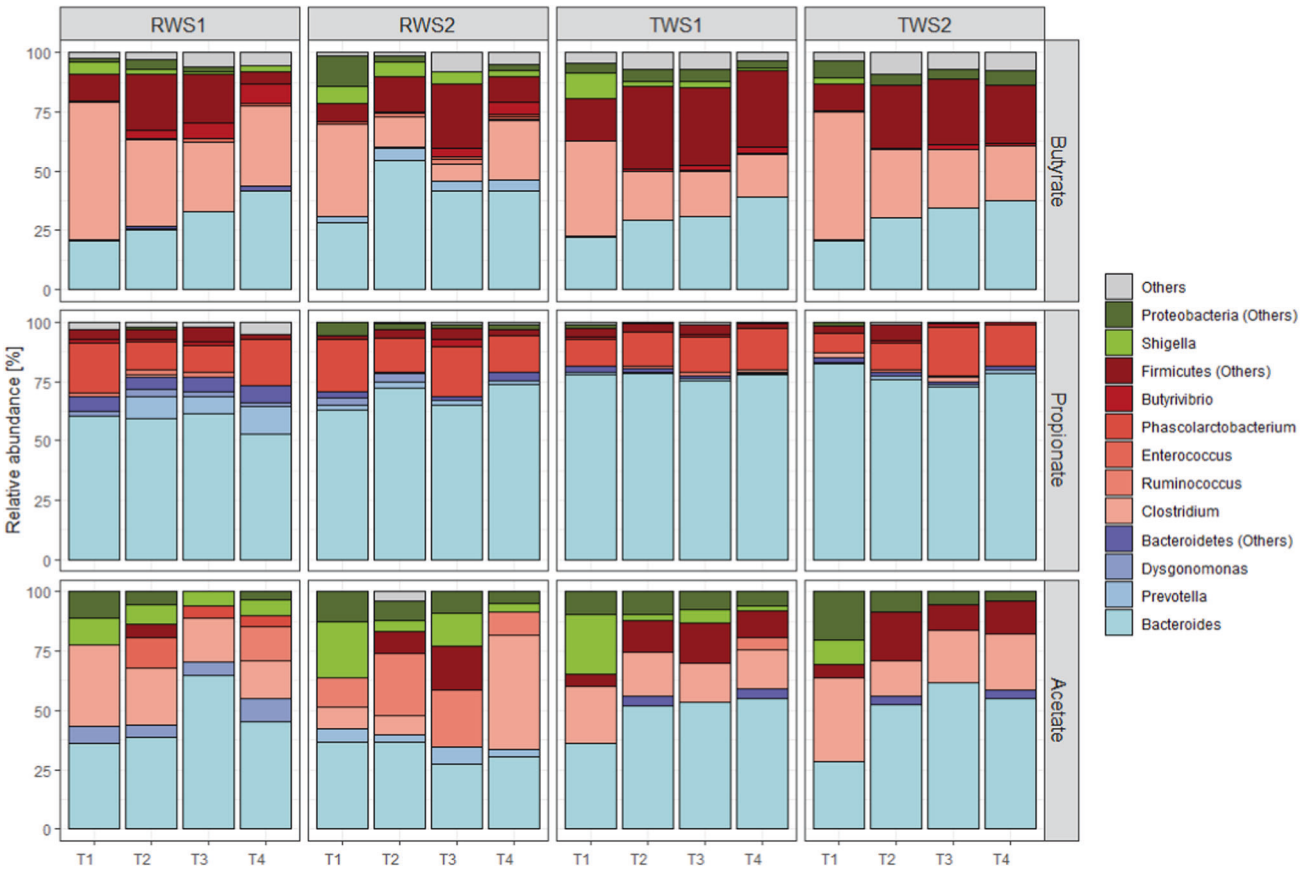

Fig. 4 Temporal profiles of major producers of enzymes involved in VFAs production. Relative abundance and phylogenetic origin of bacterial enzymes involved in butyrate, propionate and acetate production in the metaproteomes of RWS and TWS (1 and 2 indicate the biological duplicates). Taxonomic affiliation of proteins at the genus level. Proteins belonging to the same bacterial phylum were represented with the same color palette: Bacteroidetes (Bacteroidota) (blue), Firmicutes (Bacillota)(red) and Proteobacteria (Pseudomonadota) (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).

the most abundant group, representing 39.3% and 49.3% of enzymes involved in acetate biosynthesis in RWS and TWS metaproteomes, respectively, while those belonging to *Clostridium* accounted for about 20% (Fig. 4). A low abundance of proteins involved in carboxylate biosynthesis and belonging to Proteobacteria (Pseudomonadota) was also detected, in particular for acetate and butyrate production, at early stages of incubation.

DISCUSSION

To gain new knowledge pertaining to LC-degradation process mediated by microbial consortia, we studied two lignocellulolytic consortia, RWS and TWS, derived from the complex microbiomes of the cow rumen and termite gut, respectively. Compared to previous investigations, metaproteomic analysis of these consortia enabled the detection of a high number of non-redundant proteins (10,342) [23, 25, 28, 57], providing a complete repertoire of enzymes involved in LC degradation and VFA production.

Although RWS and TWS are derived from contrasting parental microbiomes, the taxonomic community composition deduced from 16 S rRNA gene sequencing and metaproteomic data showed that the active communities of both consortia were rather similar. They displayed a prevalence of Bacteroidetes (Bacteroidota)- and Firmicutes (Bacillota)-affiliated proteins, with the genera *Bacteroides* and *Clostridium* being the main representatives of these phyla, respectively. This similarity is probably the result of the selective pressure exerted by the wheat straw substrate during the enrichment process, as reported in previous studies [58–60]. *Bacteroides* and *Clostridium* are key players in LC degradation in the digestive microbiomes of herbivores, where they break down complex carbohydrates particularly cellulose, xylan and starch [61, 62]. 16 S rRNA and metaproteomic data confirmed that the most abundant genera were also the major source of plant cell wall degrading enzymes in RWS and TWS consortia, in agreement with previous observations [18]. This nevertheless contrasts with previous metatranscriptomics studies on the termite gut [63] and studies performed on enriched microbial consortium able to degrade rice and wheat straw [27], sugar cane bagasse [23] and a corn-stover degrading consortium [28]. Our findings reveal that the relationship between the abundance of microbial phyla and their functional impact on the community activity is not a simple one, and highlights the importance of metaproteomics to assign functional roles to different phyla.

The metabolic functions of RWS and TWS, assessed by COG analysis, revealed that proteins related to “translation, ribosomal structure and biogenesis”, “carbohydrate transport and metabolism” and “energy production and conversion” were dominant in both communities, as reported in previous studies [16, 28, 57]. These COG functions were expressed by the three main phyla

found in RWS and TWS, Bacteroidetes (Bacteroidota), Firmicutes (Bacillota) and Proteobacteria (Pseudomonadota), but phylum-normalized data showed that protein expression by the two latter was strongly directed towards carbohydrate metabolisms, while the former (the major phylum Bacteroidetes - Bacteroidota) was associated with larger spectrum of functions.

A large diversity of CAZymes related to LC degradation, including cellulases, hemicellulases and CBMs were detected in RWS and TWS metaproteomes. They belong to 94 families, illustrating the relative richness of these consortia when compared to the whole CAZy database ($n = 472$). CAZymes in RWS and TWS accounted for about 4% of total proteins expressed. Although most previous studies report the number of non-redundant CAZymes, rather than their abundance [16, 18, 28], the abundance reported here is comparable to that measured (3.5% and 5.9%) for other lignocellulolytic consortia [27]. This implies that dedicating less than 5% of their expressed proteins to LC degradation, RWS and TWS achieved high (>45% in our study) biomass degradation levels.

Taxonomic affiliation of CAZymes revealed that LC degradation by RWS and TWS results from the combined action of proteins affiliated to Bacteroidetes (Bacteroidota), the main purveyors of CAZymes, followed by Firmicutes (Bacillota). A closer inspection revealed that the Bacteroidetes (Bacteroidota) members mainly produce hemicellulose-debranching and oligosaccharide-degrading enzymes as well as enzymes involved in starch degradation. In contrast, Firmicutes (Bacillota) members produced enzymes specific for β-glucan and β-xylan (cellulases and hemicellulases). This differential expression of CAZymes offers interesting prospects for engineering synergy. Moreover, it earmarks these two phyla as the key players in wheat straw degradation. It is noteworthy that other studies do not systematically identify these phyla as the dominant purveyors of CAZymes, particularly when considering the termite gut microbiome [64]. To rationalize this observation, we suggest that our results are the consequence of the combined contributions of the original microbial consortium and the specific process conditions employed [23, 26–28]. The substrate, the enrichment process and the availability of oxygen all determine the relative development of obligate aerobes, facultative aerobes and strict anaerobes.

Although RWS and TWS consortia displayed taxonomic similarities, only a third of proteins were common to both consortia. Multivariate analysis highlighted the taxa and proteins linked to the parental inoculum source. Indeed, some features can be attributed to the initial cow rumen and termite gut microbiome. For example, the larger diversity of CAZymes and the ruminococcal and clostridial cellulosome components characteristic of RWS are also characteristic of the cow microbiome [65, 66]. Similarly, the high abundance of GH10, GH43, CE1 and especially GH11, typical of TWS, is also a feature of termite gut microbiome [67, 68].

Despite differences between RWS and TWS, multi-group MINT-PLS-DA supervised analysis highlighted in both consortia Bacteroidetes (Bacteroidota)- and Firmicutes (Bacillota)-affiliated enzymes, particularly ones produced by *Bacteroides, Clostridium* and *Enterococcus* genera, were marshalled. According to annotation, these enzymes target the minor starch fraction and holocellulose-derived oligosaccharides, meaning that the aforementioned genera forage ready available resources and in doing so facilitate further breakdown of holocellulose polymers. Similarly, in both consortia the occurrence of *Lachnoclostridium*, *Prevotella* and Other-Firmicute members was associated with the principal LC biomass degradation phase, which was also related with an increase in the abundance of CAZy families related to cellulose hydrolysis, deacetylation and cleavage of hemicelluloses and pectin depolymerization. In addition, during the last incubation stage, several CAZymes hydrolyzing and deconstructing hemicellulose and hemicellulose-oligosaccharides also increased, as well as various CBMs binding cellulose, hemicellulose and starch. Therefore, further biochemical characterization of these enzymes could be of interest to elucidate their specific role in LC-biomass degradation, in particular to identify enzymes able to degrade the most recalcitrant components.

Concerning VFA biosynthesis, our data showed that acetate biosynthesis was the result of Wood-Ljungdahl pathway in both our consortia, and highlighted the major role of acetogenic bacteria belonging to Bacteroidetes (Bacteroidota) and Firmicutes (Bacillota) phyla. Propionate production in RWS and TWS mostly resulted from the succinate pathway as evidenced by the detection of methylmalonyl CoA mutases and epimerases [69], with *Bacteroides* and *Phascolarctobacterium* species as the key propionate producers. This is consistent with previous data that showed that the succinate pathway is the main route reported for rumen [70]. Butyrate biosynthesis resulted 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 (Bacillota) 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 revealed that Bacteroidetes (Bacteroidota) are responsible for the majority of acetate and propionate production, while butyrate biosynthesis is mainly handled by Firmicutes (Bacillota) species [16, 71]. LC-hydrolysis remains the limiting step and thus VFA-biosynthesis cannot be used to improve LC conversion into VFA, but could be of interest to drive VFA production towards specific products (e.g. butyrate).

A remarkable finding in this work is 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 [20, 21]. Nevertheless, the fact that no lignin-degrading enzymes were evidenced does not imply that these were 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. A comparison of 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 (Table S5 [65, 72–79]) revealed that this study unmasked greater diversity of cell wall degrading CAZymes. Moreover, the comparison revealed that our study captured most CAZy families degrading the plant cell wall, 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 and its date of access [11] 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 and the further expansion of the CAZy database will surpass our study. Hopefully, these will provide an even deeper understanding of lignocellulolytic functions in microbial ecosystems and provide the means to identify proteins that are currently unclassified. Furthermore, our data showed that RWS and TWS consortia represent excellent simplified models to study the mechanisms governing the complex lignocellulose degradation process and to better

understand and exploit multispecies lignocellulolytic enzyme systems for biotechnological applications.

DATA AVAILABILITY

The data supporting the results of this study are available within the article and in its Supplementary Information. Sequencing data are available in the Sequence Read Archive (SRA) under accession number PRJNA729464. All other data supporting the findings of this study are available under request to the corresponding author.

REFERENCES

1. Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, et al. Lignocellulose degradation mechanisms across the Tree of Life. Curr Opin Chem Biol. 2015;29:108–19. https://doi.org/10.1016/j.cbpa.2015.10.018
2. Wang C, Dong D, Wang H, Müller K, Qin Y, Wang H, et al. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. Biotechnol Biofuels. 2016;9. https://doi.org/10.1186/s13068-016-0440-2.
3. Liu D, Li M, Xi B, Zhao Y, Wei Z, Song C, et al. Metaproteomics reveals major microbial players and their biodegradation functions in a large-scale aerobic composting plant. Microb Biotechnol. 2015;8:950–60. https://doi.org/10.1111/1751-7915.12290
4. Eichorst SA, Kuske CR. Identification of cellulose-responsive bacterial and fungal communities in geographically and edaphically different soils by using stable isotope probing. Appl Environ Microbiol. 2012;78:2316–27. https://doi.org/10.1128/AEM.07313-11
5. Schneider T, Keiblinger KM, Schmid E, Sterflinger-Gleixner K, Ellersdorfer G, Roschitzki B, et al. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. ISME J. 2012;6:1749–62. https://doi.org/10.1038/ismej.2012.11
6. López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P. Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. Sci Rep. 2016;6. https://doi.org/10.1038/srep25279
7. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from Cow Rumen. Science. 2011;331:463–7. https://doi.org/10.1126/science.1200387
8. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature. 2007;450:560–5. https://doi.org/10.1038/nature06269
9. Dumond L, Lam PY, van Erven G, Kabel M, Mounet F, Grima-Pettenati J, et al. Termite gut microbiota contribution to wheat straw delignification in anaerobic bioreactors. ACS Sustain Chem Eng. 2021;9:2191–202. https://doi.org/10.1021/acssuschemeng.0c07817
10. Singh KM, Reddy B, Patel D, Patel AK, Parmar N, Patel A, et al. High Potential Source for Biomass Degradation Enzyme Discovery and Environmental Aspects Revealed through Metagenomics of Indian Buffalo Rumen. BioMed Res Int. 2014. https://doi.org/10.1155/2014/267189.
11. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–5. https://doi.org/10.1093/nar/gkt1178
12. Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomics. Chem Rec NYN. 2008;8:364–77. https://doi.org/10.1002/tcr.20160
13. Scharf ME. Omic research in termites: an overview and a roadmap. Front Genet. 2015;6:76 https://doi.org/10.3389/fgene.2015.00076
14. Rosnow JJ, Anderson LN, Nair RN, Baker ES, Wright AT. Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. Crit Rev Biotechnol. 2017;37:626–40. https://doi.org/10.1080/07388551.2016.1209158
15. Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, et al. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc Natl Acad Sci USA. 2009;106:1948–53. https://doi.org/10.1073/pnas.0806191105
16. Deusch S, Camarinha-Silva A, Conrad J, Beifuss U, Rodehutscord M, Seifert J. A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. Front Microbiol. 2017;8. https://doi.org/10.3389/fmicb.2017.01605
17. Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol. 2004;6:911–20. https://doi.org/10.1111/j.1462-2920.2004.00687.x
18. Hart EH, Creevey CJ, Hitch T, Kingston-Smith AH. Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria. Sci Rep. 2018;8:10504 https://doi.org/10.1038/s41598-018-28827-7
19. Burnum KE, Callister SJ, Nicora CD, Purvine SO, Hugenholtz P, Warnecke F, et al. Proteome insights into the symbiotic relationship between a captive colony of Nasutitermes corniger and its hindgut microbiome. ISME J. 2011;5:161–4. https://doi.org/10.1038/ismej.2010.97
20. Lazuka A, Auer L, Bozonnet S, Morgavi DP, O'Donohue M, Hernandez-Raquet G. Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. Bioresour Technol. 2015;196:241–9. https://doi.org/10.1016/j.biortech.2015.07.084
21. Lazuka A, Auer L, O'Donohue M, Hernandez-Raquet G. Anaerobic lignocellulolytic microbial consortium derived from termite gut: enrichment, lignocellulose degradation and community dynamics. Biotechnol Biofuels. 2018;11:284 https://doi.org/10.1186/s13068-018-1282-x
22. Cortes-Tolalpa L, Jiménez DJ, de Lima Brossi MJ, Salles JF, van Elsas JD. Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. Appl Microbiol Biotechnol. 2016;100:7713–25. https://doi.org/10.1007/s00253-016-7516-6
23. Tomazetto G, Pimentel AC, Wibberg D, Dixon N, Squina FM. Multi-omic directed discovery of cellulosomes, polysaccharide utilization loci, and lignocellulases from an enriched rumen anaerobic consortium. Appl Environ Microbiol. 2020;86:e00199–20. https://doi.org/10.1128/AEM.00199-20
24. Jiménez DJ, de Lima Brossi MJ, Schückel J, Kračun SK, Willats WGT, van Elsas JD. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. Appl Microbiol Biotechnol. 2016;100:10463–77. https://doi.org/10.1007/s00253-016-7713-3
25. D'haeseleer P, Gladden JM, Allgaier M, Chain PSG, Tringe SG, Malfatti SA, et al. Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. PLoS ONE. 2013;8. https://doi.org/10.1371/journal.pone.0068465.
26. Jiménez DJ, Maruthamuthu M, van Elsas JD. Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. Biotechnol Biofuels. 2015;8. https://doi.org/10.1186/s13068-015-0387-8.
27. Alessi AM, Bird SM, Bennett JP, Oates NC, Li Y, Dowle AA, et al. Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. Sci Rep. 2017;7:2356 https://doi.org/10.1038/s41598-017-02506-5
28. Zhu N, Yang J, Ji L, Liu J, Yang Y, Yuan H. Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. Biotechnol Biofuels. 2016;9. https://doi.org/10.1186/s13068-016-0658-z.
29. Agler MT, Wrenn BA, Zinder SH, Angenent LT. Waste to bioproduct conversion with undefined mixed cultures: the carboxylate platform. Trends Biotechnol. 2011;29:70–8. https://doi.org/10.1016/j.tibtech.2010.11.006
30. Torella JP, Ford TJ, Kim SN, Chen AM, Way JC, Silver PA. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. Proc Natl Acad Sci USA. 2013;110:11290–5. https://doi.org/10.1073/pnas.1307129110
31. Auer L, Lazuka A, Sillam-Dussès D, Miambi E, O'Donohue M, Hernandez-Raquet G. Uncovering the potential of termite gut microbiome for lignocellulose bioconversion in anaerobic batch bioreactors. Front Microbiol. 2017;8:2623 https://doi.org/10.3389/fmicb.2017.02623
32. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: find, rapidly, OTUs with galaxy solution. Bioinformatics. 2018;34:1287–94. https://doi.org/10.1093/bioinformatics/btx791
33. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE. 2013;8:e61217 https://doi.org/10.1371/journal.pone.0061217
34. Lin Y-W, Tuan NN, Huang S-L. Metaproteomic analysis of the microbial community present in a thermophilic swine manure digester to allow functional characterization: A case study. Int Biodeterior Biodegrad. 2016;115:64–73. https://doi.org/10.1016/j.ibiod.2016.06.013
35. Starke R, Müller M, Gaspar M, Marz M, Küsel K, Totsche KU, et al. Candidate Brocadiales dominates C, N and S cycling in anoxic groundwater of a pristine limestone-fracture aquifer. J Proteomics. 2017;152:153–60. https://doi.org/10.1016/j.jprot.2016.11.003
36. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47:D442–50. https://doi.org/10.1093/nar/gky1106
37. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V, Deligios M, et al. The impact of sequence database choice on metaproteomic results in gut microbiota studies. Microbiome. 2016;4:51 https://doi.org/10.1186/s40168-016-0196-8
38. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007;4:923–5. https://doi.org/10.1038/nmeth1113
39. Schneider T, Schmid E, de Castro JV, Cardinale M, Eberl L, Grube M, et al. Structure and function of the symbiosis partners of the lung lichen (Lobaria pulmonaria L.

Hoffm.) analyzed by metaproteomics. Proteomics. 2011;11:2752–6. https://doi.org/10.1002/pmic.201000679
40. Brune A. Symbiotic digestion of lignocellulose in termite guts. Nat Rev Microbiol. 2014;12:168–80. https://doi.org/10.1038/nrmicro3182
41. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41 https://doi.org/10.1186/1471-2105-4-41
42. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Res. 2012;40:D290–301. https://doi.org/10.1093/nar/gkr1065
43. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res. 2006;5:2339–47. https://doi.org/10.1021/pr060161n
44. Svartström O, Alneberg J, Terrapon N, Lombard V, de Bruijn I, Malmsten J, et al. Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. ISME J. 2017;11:2538–51. https://doi.org/10.1038/ismej.2017.108
45. Delannoy-Bruno O, Desai C, Raman AS, Chen RY, Hibberd MC, Cheng J, et al. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. Nature. 2021;595:91–5. https://doi.org/10.1038/s41586-021-03671-4
46. Xia Y, Sun J. Hypothesis testing and statistical analysis of microbiome. Genes Dis. 2017;4:138–48. https://doi.org/10.1016/j.gendis.2017.06.001
47. Calle ML. Statistical Analysis of Metagenomics Data. Genomics Inform 2019;17. https://doi.org/10.5808/GI.2019.17.1.e6
48. Lubbe S, Filzmoser P, Templ M. Comparison of zero replacement strategies for compositional data with large numbers of zeros. Chemom Intell Lab Syst. 2021;210:104248 https://doi.org/10.1016/j.chemolab.2021.104248
49. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. Math Geol. 2003;35:253–78. https://doi.org/10.1023/A:1023866030544
50. Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput Biol. 2017;13:e1005752 https://doi.org/10.1371/journal.pcbi.1005752
51. Rohart F, Eslami A, Matigian N, Bougeard S, Lê Cao K-A. MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. BMC Bioinformatics. 2017;18:128 https://doi.org/10.1186/s12859-017-1553-8
52. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x
53. Mewis K, Lenfant N, Lombard V, Henrissat B. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. Appl Environ Microbiol. 2016;82:1686–92. https://doi.org/10.1128/AEM.03453-15
54. Dijkstra J. Production and absorption of volatile fatty acids in the rumen. Livest Prod Sci. 1994;39:61–9. https://doi.org/10.1016/0301-6226(94)90154-6
55. Polansky O, Sekelova Z, Faldynova M, Sebkova A, Sisak F, Rychlik I. Important metabolic pathways and biological processes expressed by chicken cecal microbiota. Appl Environ Microbiol. 2016;82:1569–76. https://doi.org/10.1128/AEM.03473-15
56. Tilocca B, Burbach K, Heyer CME, Hoelzle LE, Mosenthin R, Stefanski V, et al. Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome—from days to weeks. Microbiome. 2017;5. https://doi.org/10.1186/s40168-017-0362-7
57. Alessi AM, Bird SM, Oates NC, Li Y, Dowle AA, Novotny EH, et al. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. Biotechnol Biofuels. 2018;11:166 https://doi.org/10.1186/s13068-018-1164-2
58. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, Tangphatsornruang S, Eurwilaichitr L, Igarashi Y, et al. Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. Appl Microbiol Biotechnol. 2013;97:8941–54. https://doi.org/10.1007/s00253-013-4699-y
59. Eichorst SA, Joshua C, Sathitsuksanoh N, Singh S, Simmons BA, Singer SW. Substrate-specific development of thermophilic bacterial consortia by using chemically pretreated switchgrass. Appl Environ Microbiol. 2014;80:7423–32. https://doi.org/10.1128/AEM.02795-14
60. Simmons CW, Reddy AP, Simmons BA, Singer SW, VanderGheynst JS. Effect of inoculum source on the enrichment of microbial communities on two lignocellulosic bioenergy crops under thermophilic and high-solids conditions. J Appl Microbiol. 2014;117:1025–34. https://doi.org/10.1111/jam.12609
61. Nishiyama T, Ueki A, Kaku N, Watanabe K, Ueki K. Bacteroides graminisolvens sp. nov., a xylanolytic anaerobe isolated from a methanogenic reactor treating cattle waste. Int J Syst Evol Microbiol. 2009;59:1901–7. https://doi.org/10.1099/ijs.0.008268-0
62. Munir RI, Schellenberg J, Henrissat B, Verbeke TJ, Sparling R, Levin DB. Comparative analysis of carbohydrate active enzymes in clostridium termitidis CT1112 reveals complex carbohydrate degradation ability. PLoS ONE. 2014;9. https://doi.org/10.1371/journal.pone.0104260
63. Marynowska M, Goux X, Sillam-Dussès D, Rouland-Lefèvre C, Roisin Y, Delfosse P, et al. Optimization of a metatranscriptomic approach to study the lignocellulolytic potential of the higher termite gut microbiome. BMC Genomics. 2017;18. https://doi.org/10.1186/s12864-017-4076-9
64. Calusinska M, Marynowska M, Bertucci M, Untereiner B, Klimek D, Goux X, et al. Integrative omics analysis of the termite gut system adaptation to Miscanthus diet identifies lignocellulose degradation enzymes. Commun Biol. 2020;3:1–12. https://doi.org/10.1038/s42003-020-1004-3
65. Dai X, Tian Y, Li J, Su X, Wang X, Zhao S, et al. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the Cow Rumen. Appl Environ Microbiol. 2015;81:1375–86. https://doi.org/10.1128/AEM.03682-14
66. Lamed R, Setter E, Kenig R, Bayer EA. Cellulosome: a discrete cell surface organelle of Clostridium thermocellum which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. (Tel Aviv Univ., Ramat Aviv, Israel; 1983).
67. Tokuda G, Mikaelyan A, Fukui C, Matsuura Y, Watanabe H, Fujishima M, et al. Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites. Proc Natl Acad Sci USA. 2018;115:E11996–2004. https://doi.org/10.1073/pnas.1810550115
68. Marynowska M, Goux X, Sillam-Dussès D, Rouland-Lefèvre C, Halder R, Wilmes P, et al. Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites. Microbiome. 2020;8:96 https://doi.org/10.1186/s40168-020-00872-3
69. Watanabe Y, Nagai F, Morotomi M. Characterization of Phascolarctobacterium succinatutens sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces. Appl Environ Microbiol. 2012;78:511–8. https://doi.org/10.1128/AEM.06035-11
70. Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S, et al. Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. PLoS ONE. 2010;5. https://doi.org/10.1371/journal.pone.0008812
71. Abdallah Ismail N, Ragab SH, Abd ElBaky A, Shoeib ARS, Alhosary Y, Fekry D. Frequency of Firmicutes and Bacteroidetes in gut microbiota in obese and normal weight Egyptian children and adults. Arch Med Sci AMS. 2011;7:501–7. https://doi.org/10.5114/aoms.2011.23418
72. Wang L, Hatem A, Catalyurek UV, Morrison M, Yu Z. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the Rumen of Cows. PLoS ONE. 2013;8. https://doi.org/10.1371/journal.pone.0078507
73. Jose VL, Appoothy T, More RP, Arun AS. Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. AMB Expr. 2017;7. https://doi.org/10.1186/s13568-016-0310-0
74. Li J, Zhong H, Ramayo-Caldas Y, Terrapon N, Lombard V, Potocki-Veronese G, et al. A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. GigaScience. 2020;9:giaa057 https://doi.org/10.1093/gigascience/giaa057
75. Comtet-Marre S, Parisot N, Lepercq P, Chaucheyras-Durand F, Mosoni P, Peyretaillade E, et al. Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the Rumen of dairy cow fed a mixed diet. Front Microbiol. 2017;8. https://doi.org/10.3389/fmicb.2017.00067
76. Shinkai T, Mitsumori M, Sofyan A, Kanamori H, Sasaki H, Katayose Y, et al. Comprehensive detection of bacterial carbohydrate-active enzyme coding genes expressed in cow rumen. Anim Sci J Nihon Chikusan Gakkaiho. 2016;87:1363–70. https://doi.org/10.1111/asj.12585
77. He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, et al. Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. PLoS ONE. 2013;8. https://doi.org/10.1371/journal.pone.0061126
78. Grieco MB, Lopes FAC, Oliveira LS, Tschoeke DA, Popov CC, Thompson CC, et al. Metagenomic analysis of the whole gut microbiota in brazilian termitidae termites cornitermes cumulans, cyrilliotermes strictinasus, syntermes dirus, nasutitermes jaraguae, nasutitermes aquilinus, grigiotermes bequaerti, and orthognathotermes mirim. Curr Microbiol. 2019;76:687–97. https://doi.org/10.1007/s00284-019-01662-3
79. Romero Victorica M, Soria MA, Batista-García RA, Ceja-Navarro JA, Vikram S, Ortiz M, et al. Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes. Sci Rep. 2020;10:3864 https://doi.org/10.1038/s41598-020-60850-5

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.

AUTHOR CONTRIBUTIONS

EA carried out and analyzed metaproteomics and microbial diversity data under GHR supervision. AL realized the bioreactor experiments and prepared the samples for metaproteomics analysis under the supervision of GHR, MO and NJ. NJ performed and analyzed LC-MC-MS experiments. BH and VL participate on CAZy analysis. SD supervised the statistical data analysis realized by EA and BH-B. As project coordinator, GHR designed the study, participated on experimental design and contributed at all stages. The manuscript was written by EA under GHR and MO supervision and with important intellectual contributions from all authors. All authors contributed and have given approval to the final version of the manuscript. GHR and NJ participated in funding this research.

FUNDING

This study 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.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43705-023-00339-0.

Correspondence and requests for materials should be addressed to Guillermina Hernandez-Raquet.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.