# Supplementary Information

# Material and Methods

**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* [1] and cow rumen [2]were used as inocula for a sequential batch bioreactor enrichment. The initial termite gut inoculum (500 dissected guts) was provided by IRD (Institute for Research and Development, Bondy, France). The initial rumen inoculum (2% w/v) was provided by INRAE (French National Institute for Agriculture, Food and Environment) Center of Theix (St-Gènes-Champanelle, France).

**Anaerobic enrichment and lignocellulose degradation bioreactors of RWS and TWS**

The lignocellulolytic communities from termite gut and cow rumen microbiomes were selectively enriched by successive cultivation using wheat straw as sole carbon source and mineral media (MM) [2], supplemented with 250 μL of V7 vitamin solution [3] and 1 mL of sterilized (0.2 μm filtration) trace elements solution [2]. The mineral medium (MM) contained, per liter of distilled water: KH2PO4, 0.45 g, K2HPO4, 0.45 g; NH4Cl, 0.4 g; NaCl, 0.9 g; MgCl2·6H2O, 0.15 g; CaCl2·2H2O, 0.09 g. MM was supplemented with 250 μL of V7 vitamin solution, and 1 mL trace elements solution, containing per liter of distilled water: H3BO3, 300 mg; FeSO4·7H2O, 1.1 g; CoCl2·6H2O, 190 mg; MnCl2·4H2O, 50 mg; ZnCl2, 42 mg; NiCl2·6H2O, 24 mg; NaMoO4·2H2O, 18 mg; CuCl2·2H2O, 2 mg; sterilized by filtration (0.2 μm). Each successive bioreactor was inoculated with 10% v/v of the previous bioreactor. All bioreactor experiments were carried out in 2L 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) [2, 4].

Lignocellulose degradation capacity of RWS and TWS was assessed in two replicate anaerobic bioreactors for each inoculum (4 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 enzymatic activities. Samples were also taken for biodiversity and metaproteomics analysis.

**Dry matter degradation**

Wheat straw concentration was determined along the incubation by measuring the dry matter. Total solids were determined using 10 mL samples that were first centrifuged (7,197 × 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.

**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) [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 [2].

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 µL sample (or an appropriate dilution thereof) were added to 400 µL of substrate solution and incubated at 35°C for 1h (xylanase). Reactions were stopped by adding 600 µL of DNS solution and heating for 5 minutes at 100°C. The amount of reducing sugars released was 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 µmol of reducing sugars per minute.

## Microbial diversity analysis

A minimum of 25,000 high quality joint-pair reads generated for each sample were simultaneously imported in Find Rapidly OTUs with Galaxy solution (FROGS) pipeline. Sequencing data preprocessing in FROGS 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 [6]. OTUs with a total abundance lower than 0.005% of the overall dataset were discarded [7] and chimeras were removed using VSEARCH [8] using default FROGS parameters. Taxonomic assignation of OTUs was performed by Blast using Silva 132 16S with a pintail of 100 as reference database [9]. A phylogenetic tree was produced using the FROGS dedicated tool and default parameters. This produce a final OTUs abundance table with the taxonomic affiliation.

## Metaproteomics analysis

### **Protein extraction, peptides digestion and mass spectrometry analysis**

### Protein extraction and separation

Protein extraction was carried out using a phenol buffer following the procedure for complex sediment samples [10]. 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 heated 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 mm²) and transferred into 0.5 mL Eppendorf tubes.

### Proteolytic digestion with trypsin and purification

After protein extraction and separation by SDS-PAGE, 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.

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-precolumn, 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).

### Liquid chromatography-tandem mass spectrometry analysis

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 3×106 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 2×105 with an injection time of 120 ms using the underfill ratio of 1% [11].

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 Uniprot-TrEMBL database (release date of 6th 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 and restricted to the taxa identified by 16S rRNA analysis [12, 13]; this exclude plant or eukaryotic proteins. 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 [14], 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).

**References**

1. Auer L, Lazuka A, Sillam-Dusses D, Miambi E, O'Donohue M *et al*: Uncovering the potential of termite gut microbiome for lignocellulose bioconversion in anaerobic batch bioreactors. Front Microbiol. 2017. 8:2623.

2. Lazuka A, Auer L, Bozonnet S, Morgavi DP, O'Donohue M *et al*: Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. Bioresour Technol. 2015. 196:241-249.

3. Pfennig N, Trüper HG: The family Chromatiaceae. In: *The Prokaryotes.* Edited by Balows A, Trüper HG, Dworkin M *et al*. New York: Springer; 1992: 3200-3221.

4. 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.

5. Cavaille L, Grousseau E, Pocquet M, Lepeuple AS, Uribelarrea JL *et al*: Polyhydroxybutyrate production by direct use of waste activated sludge in phosphorus-limited fed-batch culture. Bioresour Technol. 2013. 149:301-309.

6. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M: Swarm: robust and fast clustering method for amplicon-based studies. PeerJ. 2014. 2:e593.

7. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI *et al*: Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013. 10(1):57-59.

8. Rognes T, Flouri T, Nichols B, Quince C, Mahé F: VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016. 4:e2584.

9. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T *et al*: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013. 41(Database issue):D590-596.

10. 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. International Biodeterioration & Biodegradation. 2016. 115:64-73.

11. Starke R, Müller M, Gaspar M, Marz M, Küsel K *et al*: Candidate Brocadiales dominates C, N and S cycling in anoxic groundwater of a pristine limestone-fracture aquifer. J Proteomics. 2017. 152:153-160.

12. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V *et al*: The impact of sequence database choice on metaproteomic results in gut microbiota studies. Microbiome. 2016. 4, 51.

13. Jouffret, V., Miotello, G., Culotta, K. et al. Increasing the power of interpretation for soil metaproteomics data. Microbiome. 2021. 9, 195.

14. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ: Semisupervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods. 2007. 4, 923–925