Metaproteomics reveals enzymatic strategies deployed by anaerobic microbiomes to maintain lignocellulose deconstruction at high solids

Payal Chirania¹,²,³,⁶, Evert K. Holwerda³,⁴,⁶, Richard J. Giannone¹,³, Xiaoyu Liang⁴, Suresh Poudel¹, Joseph C. Ellis¹,³, Yannick J. Bomble³,⁵, Robert L. Hettich¹,³ & Lee R. Lynd³,⁴

Economically viable production of cellulosic biofuels requires operation at high solids loadings —on the order of 15 wt%. To this end we characterize Nature’s ability to deconstruct and utilize mid-season switchgrass at increasing solid loadings using an anaerobic methanogenic microbiome. This community exhibits undiminished fractional carbohydrate solubilization at loadings ranging from 30 g/L to 150 g/L. Metaproteomic interrogation reveals marked increases in the abundance of specific carbohydrate-active enzyme classes. Significant enrichment of auxiliary activity family 6 enzymes at higher solids suggests a role for Fenton chemistry. Stress-response proteins accompanying these reactions are similarly upregulated at higher solids, as are β-glucosidases, xylosidases, carbohydrate-debranching, and pectin-acting enzymes—all of which indicate that removal of deconstruction inhibitors is important for observed undiminished solubilization. Our work provides insights into the mechanisms by which natural microbiomes effectively deconstruct and utilize lignocellulose at high solids loadings, informing the future development of defined cultures for efficient bioconversion.
Biological production of liquid fuels from lignocellulosic feedstocks is of high interest as society navigates a transition away from fossil resources. However, the recalcitrant character of such feedstocks impedes biological conversion and represents a major cost barrier. One-step consolidated bioprocessing (CBP) without added enzymes, mediated by defined cultures of anaerobic lignocellulose-fermenting bacteria, is a promising strategy for converting lignocellulose to liquid fuels at low cost. Because substantial titers of liquid fuel products are achieved from cultures of anaerobic lignocellulose-fermenting bacteria, is a processing (CBP) without added enzymes, mediated by dehydrogenase and substrate-adhered phases. Liang et al. performed metagenomic and phylogenetic analysis of the resulting microbiome via 1D LC-MS/MS-based metaproteomics, where the residence time is fixed at 10 days, with increasing solids loading from 30 to 75, 120, and finally 150 g/L of the same feedstock.

Results

Microbiome maintains undiminished solubilization at higher solids. An anaerobic lignocellulosic thermophilic (55 °C) methanogenic microbiome was established by enriching an anaerobic digester inoculum on 30 g/L unpertreated June-harvested non-senescent switchgrass and incubating the culture for over 120 days before entering the data collection phase. The microbiome with a 1 L working volume was operated semi-continuously at a 10-day residence time by daily withdrawal of 100 mL of well-mixed bioreactor contents followed by addition of 100 mL total volume of switchgrass and growth media. The bioreactor was operated at solids loadings of 30, 75, 120, and 150 g/L, with the cultivation at each loading lasting for at least 50 days, leading to an overall microbiome cultivation time of 550 days (Fig. 1A, Supplementary Table 1, and Methods).

The fraction of carbohydrate solubilized at steady state—quantified by the loss of glucose, xylose, and arabinose moieties in the solids—remained relatively constant at ~66.8% regardless of solids loading over the range considered. The rate of carbohydrate solubilization increased in proportion to solids loading (Fig. 1B). The main products observed were methane (CH₄) and carbon dioxide (CO₂), whereas volatile fatty acids (VFA) and hydrogen (H₂) were not detected at steady state (Supplementary Fig. 1, Supplementary Table 2). Methane and CO₂ measured in effluent gas ranged from 48.7 to 49.9%, and 47.2 to 49.0% respectively. Nitrogen was also present at < 5%, likely introduced in the feeding and sampling process. Volumetric gas production and the methane production rate increased proportionally with increasing solids loading (Fig. 1C, Supplementary Table 2), indicating that product formation by the microbiome was not differentially affected up to 150 g/L solids.

Fractionation reveals spatial distribution of enzymatic categories. A metaproteomic analysis was undertaken at each solids loading to provide a detailed molecular view of the composition and function of the lignocellulose-fermenting microbiome. The overall amount of microbial biomass, as assessed by an initial screen of unfractionated microbiome via 1D LC-MS/MS-based metaproteomic analysis, did not increase proportionally (p-value > 0.05, two-tailed Welch’s t-test) with the 5-fold increase...
in substrate concentration; although significant abundance changes across a myriad of functional categories were observed with increasing solids (Fig. 1A, Supplementary Note 1, Supplementary Data 1). In order to further examine the changes in the community’s functional structure across solids loadings and deepen the overall analysis, a multidimensional metaproteomic measurement was conducted on fractionated microbiome samples. As follows, each sample was split into three fractions: supernatant (SNT), planktonic cells (PC), and substrate-bound (SB). Sample fractionation to segregate extracellular proteins from planktonic- and substrate-bound microbes and enzymes has been demonstrated previously, enabling the exploration of metabolic variability due to spatial localization of enzymes and microbes.  

A total of 16,644 microbial protein groups were quantified based on unique peptide evidence across the three fractions at each substrate loading (Supplementary Data 2–4). There were both complexity and proteomic differences between the fractions, as only a third of the proteins were common among all three (Supplementary Fig. 2A). However, within each fraction, a comparable number of proteins were quantified across all solids loadings (Supplementary Fig. 2B), suggesting complexity...
differences were driven by the type of fraction rather than the solids loading. When considering proteome changes within each fraction, replicates of each solids loading clustered together, while broader differences were observed between low (30 g/L) and high loadings (120, 150 g/L) (Supplementary Fig. 2C).

Since the methane production rate increased proportionally with increasing solids, we first examined methanogenesis pathway-related proteomic signatures across the three fractions and assessed whether the observed increase was corroborated by metaproteomic analysis (Supplementary Fig. 3–6, Supplementary Data 5, Supplementary Note 2). The PC and SB fractions, which both include whole cells, had a greater representation of enzymes mapping to the different methanogenic pathways than the SNT fraction (Supplementary Fig. 3A). All methanogenesis pathway enzymes exhibited a clear increasing trend in the PC fraction that correlated with increasing solids loadings. This fraction also had the greatest representation of enzymes involved in the methanogenesis process (Supplementary Figs. 3A, B and 5). Closer inspection revealed that the PC fraction contained the highest total abundance of proteins from *Euryarchaeota* such that this phylum contributed substantially to the PC metaproteome relative to the other fractions (Supplementary Fig. 3C). Among the archaeal proteins, both the aggregate abundance and diversity (protein count) of those mapping the hydrogenotrophic pathway correlated with increasing solids loadings. This fraction also had the greatest representation of enzymes involved in the methanogenesis process (Supplementary Figs. 3A, B and 5). Closer inspection revealed that the PC fraction contained the highest total abundance of proteins from *Euryarchaeota* such that this phylum contributed substantially to the PC metaproteome relative to the other fractions (Supplementary Fig. 3C). Among the archaeal proteins, both the aggregate abundance and diversity (protein count) of those mapping the hydrogenotrophic pathway markedly exceeded that of the acetoclastic pathway (Supplementary Figs. 3A and 7), suggesting this to be the major route for methanogenesis.

**A diverse Firmicutes-dominated CAZyme repertoire is expressed.** CAZymes are directly responsible for catalyzing the conversion of carbohydrates harbored in lignocellulose into a soluble form. To fully characterize this central process and detail how CAZymes from different organisms and/or phyla synergize to efficiently deconstruct lignocellulose at increased solids, the fractionated metaproteomes were analyzed in terms of CAZyme type, abundance, and location. Prior to this analysis, a survey of the entire metagenome was conducted using the dbCAN2 meta server to annotate important CAZyme families and their functional activities. Figure 2 shows the overview of CAZymes identified in each fraction and their general trends with solids loading (Supplementary Data 6). In total, 551 CAZyme protein groups were quantified (representing 1,246 proteins or ~14% of the total CAZymes annotated in the metagenome), with differences evident across the fractions (Fig. 2A). Notably, only 75 CAZymes (~14%) were present in all three fractions, whereas the PC and SB fractions harbored the greatest number of unique CAZymes (Fig. 2A). These differences demonstrate the compartmentalization of distinct CAZymes in the microbiome system and the importance of measuring spatially resolved fractions. About twenty percent (110 of 551) of the CAZymes harbored a carbohydrate binding domain (CBM) or a cellulosomal domain (cohesin or dockerin) (Supplementary Fig. 8, Supplementary Data 6). The proportion of these affinity-conferring CAZymes declined in the SNT fraction (from ~60% to ~20%) with increasing solids, while minimal changes were observed in the PC (~30%) and SB (~50%) fractions (Supplementary Fig. 8, 9), suggesting that free enzymes are somewhat excluded from binding substrate at lower solids loadings and remain in the SNT until surfaces become available.

Taxonomically, members of the phyla Firmicutes (*Clostridia* followed by *Bacillus*), known for their cellulosytic capabilities in thermophilic environments, were the dominant contributors of these measured CAZymes. However, fraction-specific differences in the taxonomic lineages of the CAZymes were observed (Fig. 2B). For example, CAZymes from the phylum Chloroflexi were more prevalent in the PC fraction while enzymes from the phylum *Thermotoga* were more represented in the SNT and SB fractions (Fig. 2B), suggesting the presence of functional niches. From a functional perspective, Glycoside Hydrolases (GHs) formed the most numerous and abundant CAZyme class irrespective of the fractions (Fig. 2B), in line with previous studies. Other CAZyme classes—polysaccharide lyases (PLs), carbohydrate esterases (CEs), and glycosyl transferases (GTs)—appeared to be dominant in the substrate- or cell-associated fractions while auxiliary activities (AAs) were more represented in the extracellular fraction (Fig. 2B), further indicating the enrichment of enzymatic activities based on the cellular location.

Quantitative assessment of CAZyme classes revealed trends in their abundances from low to high solids loadings in all three fractions (Fig. 2C, Supplementary Fig. 9C). AAs formed a small component of the CAZyme repertoire; however, they exhibited a significant increase in abundance in all fractions with increasing solids loading. Meanwhile, the abundances of PLs and CEs both decreased significantly in the SNT fraction at higher solids (120 and 150 g/L) but exhibited opposing trends in PC and SB fractions, especially in the case of PLs. The abundance of PLs is roughly similar across both SNT and SB fractions at low solids. However, as solids increase, the PLs decrease in the SNT while proportionally increasing in the SB, indicating that these enzymes may be overexpressed at low solids but only become associated with substrate fractions at higher solids when presumably more binding area is available (Fig. 2C).

**Oligosaccharide acting β-glucosidases increase at higher solids.** Multiple GH proteins across the three main cellulolytic categories (endoglucanases, exoglucanases, and β-glucosidases) were present in the cultures at varying levels. Both endoglucanases (mostly GH9 and GH5, in accordance with other studies) and β-glucosidases (primarily GH3) were highly represented, indicating the importance of these cellulolytic activities in the system (Fig. 3, Supplementary Data 7). Unexpectedly, at 150 g/L solids—representing a 5-fold increase in switchgrass—the collective abundance of endoglucanases, a main constituent of many cellulose degrading microbes, did not change in the SB fraction, decreased significantly in the SNT fraction, and only increased moderately (1.4-fold in 120 g/L) in the PC fraction relative to 30 g/L (Fig. 3, Supplementary Fig. 11, 12D). Also, exoglucanases—considered the main enzymes responsible for cellulose chain depolymerization, exhibited a similar abundance trend with increasing solids,
specifically with regard to SB and SNT fractions. In contrast, a significant increase in the cumulative abundance of β-glucosidases was observed especially in the SNT fraction (~12-fold increase) where small oligosaccharides such as cellulose might occur (Figs. 2D, 3, Supplementary Fig. 11, 12A). β-glucosidases are known to relieve end-product inhibition of cellulases, improve conversion yields and cellulose saccharification rates, and lower total enzyme requirements. These observations are consistent with a synergistic role of β-glucosidases in maintaining cellulose hydrolysis at higher solids loadings. Taxa-level dynamics of β-glucosidase expression were also observed (Supplementary Fig. 12). While members of multiple phyla produced β-glucosidases, the increase in abundance at high solids was driven by the phylum Dictyoglomi (Supplementary Fig. 12A). Distantly related to Caldicelluliruptor spp., members of this phylum are known to be extremely thermophilic, possess rigid cell membranes, and produce thermostable enzymes.

Hemicellulose debranching enzymes increase at higher solids. Multiple proteins across different GH and CE families were grouped into a variety of hemicellulolytic activities and their trends with increasing solids loading explored. These activities included: xylanases, xylosidases, arabinosidases, glucuronidases, fucosidases, galactosidases, mannosidases, and acetyl-xylan esterases (Fig. 3, Supplementary Fig. 11, Supplementary Data 7). Both the major xylan-hydrolyzing enzymatic classes, xylanases and xylosidases, were represented by numerous proteins and were most prevalent in the SB fraction (Fig. 3, Supplementary Fig. 12b, Supplementary Data 7). The total abundance of xylanases at 150 g/L solids loading decreased significantly in the SNT and PC fractions (~10-fold and ~2-fold, respectively) compared to 30 g/L, but remained relatively unchanged in the SB fraction (Fig. 3, Supplementary Fig. 12B). However, xylosidases exhibited a consistent increase at higher solids in the SB fraction (~3-fold in 120 g/L and 1.9-fold in 150 g/L) (Fig. 3, Supplementary Fig. 12B).
Since xylan breakdown was undiminished at high solids loading, the observations here suggest xylanases may be needed for maintaining overall xylan hydrolysis—similar to what was observed with β-glucosidases. Likewise, the importance of xylanolytic activity at higher solids was further substantiated by community dynamics, where the increase was driven by a protein group from Bacteroidetes (Supplementary Fig. 12B), members of whom are key xylanolytic organisms in the rumen and human gut ecosystems.

Along with xylan-acting enzymes, numerous hemicellulose-acting and debranching enzymes were expressed by the microbial community (Fig. 3, Supplementary Data 7). In particular, proteins with arabinosidase activity were highly represented (61 proteins) and demonstrated increased abundance with increasing solids loadings in all fractions (Fig. 3, Supplementary Fig. 12E). Even though arabinose makes up only ~4% of switchgrass carbohydrate content, the disproportionately high number of proteins dedicated to cleaving branching arabinose residues suggests a need for arabinosidases at higher solids. Similarly, despite galactose constituting <2% of the sugar content in switchgrass, 47 proteins with galactosidase activity were detected. This group is composed primarily of family GH2 and contains CAZymes with both α- and β-galactosidase activities that cleave the branching galactose residues in mannan and xylan. Their abundance significantly increased in the PC fraction at higher solids (Fig. 3), a trend driven by the phylum Chloroflexi along with Firmicutes-Clostridia (Supplementary Fig. 12F). The filamentous Chloroflexi are enriched in the PC and SB fractions and are known to be involved in hemicellulose and pectin breakdown.

**Fig. 3** Analysis of CAZyme functional categories at higher solids loadings. (left side) Enumeration of quantified unique CAZyme proteins across different CAZyme families (GH, CE, PL) after organizing them based on functional annotation as described in Supplementary Data 3. (right side) Heatmap depicting the change in aggregate abundance of CAZymes in each functional category across solids loadings for each fraction. Log2 differences from 30 g/L solids in the respective fraction are shown. A two-tailed Welch’s t-test was performed with Benjamini Hochberg FDR correction for each solids loading versus the 30 g/L loading. + means adjusted (adj.) p-value ≤ 0.05 and * means adj. p-value ≤ 0.05 with absolute fold change ≥2x vs. respective 30 g/L condition. Exact p-values for each comparison are listed in Supplementary Data 6. SNT: supernatant, PC: planktonic cells, and SB: substrate bound fraction. Source data are provided as a Source Data file.
carbohydrate fermenters. They are likely involved in the degradation of polymeric organic compounds facilitating the growth of other bacteria to maintain overall solubilization. Conversely, the abundance of other enzymes considered important for solubilization, such as acetyl-xylan esterases (34 protein groups) and C_6 phosphorylases (44 protein groups including cellobiose phosphorylases), decreased in abundance or remained relatively unchanged at higher solids loadings although they were well-represented (Fig. 3, Supplementary Data 7).

Pectin-acting enzymes increase in substrate-associated fractions. Pectin represents a small fraction of switchgrass on a mass basis (~2%), but it provides structural tenacity and contributes to its recalcitrance. Numerous pectin-acting proteins representing different CAZyme families (PL, GH, and CE) were expressed by the microbiome, which can be roughly grouped into four functional categories: galacturonan lyase, rhamnogalacturonan lyase, rhamnosidase, and pectinesterase (Fig. 3, Supplementary Fig. 11, Supplementary Data 7). While pectinesterases and rhamnosidases had decreasing abundances at higher solids, significant increases in the two pectin lyases, driven by the phyla Firmicutes, were observed in PC and/or SB fractions at high solids. These pectin lyases can be particularly important because they cleave the alpha-1,4- linkages in esterified pectin (main chain) without requiring accessory enzymes or water molecules, as is the case with PLs.

Bacterial AA6 proteins increase dramatically at higher solids. Auxiliary activities (AAs) are a recently constituted class of CAZymes that act on lignin or are involved in modification of polysaccharides such as cellulose in the plant cell wall. A significant increase in the abundance of AA enzymes was observed across all three fractions as solids loadings increased (Fig. 2C). In the current study, 3 AA families were identified (Fig. 4A). AA2 (peroxidase) and AA3 (glucose-methanol-choline (GMC)
oxido reductase) were identified only in the PC fraction at relatively low abundance. The cumulative abundance of AA2 proteins had a decreasing trend with increasing solids loadings (Supplementary Fig. 13A). AA3 was an order of magnitude more abundant than AA2 and represented by multiple proteins (11/14 AA proteins mapping to family AA6) across the three fractions. AA6s exhibited significantly increased abundance at higher solids (~23-fold in SNT at 150 g/L); a trend that was consistently observed across all three sample fractions (Fig. 4A, B). Members of the phyla Firmicutes-Clostridia and Firmicutes-Bacilli were the primary contributors to the AA6 protein group, but the increase in abundance at higher solids was driven by the phylum Firmicutes-Bacilli (Fig. 4B). AA6 family enzymes are characterized as 1,4-benzoquinone reductases and catalyze the NADPH-mediated conversion β-benzoquinone to hydroquinone61 (Fig. 4C). Although the switchgrass was not pretreated in the current study, β-benzoquinone is a prevalent and lignin-derived growth/enzyme inhibitor that is difficult to remove, suggesting that AA6-like enzymes could be involved in biomass hydrolysis detoxification52,63. While recent studies have identified an oxido reductase gene, ZMO1116, that can convert benzoquinone to non-toxic hydroquinone52,63, multiple sequence alignment analysis of the AA6 proteins identified here only revealed local sequence similarities to ZMO1116 (Supplementary Fig. 13A), suggesting a potentially different mode of action in this system.

Similar to biochemical mechanisms observed in brown-rot fungi, AA6 in the bacterial community may also be involved in the production of extracellular oxyradicals for lignin modification. AA6, or quinone reductase in brown-rot fungi, is involved in the generation of extracellular Fenton reagents via redox cycling61,64. Unlike white-rot fungi, which degrade and utilize lignin, cellulolytic brown-rot fungi modify lignin using non-enzymatic, energetically inexpensive Fenton chemistry to break the lignin barrier and enable enzyme access to the underlying polysaccharides.65,66. Fenton reactions make use of peroxide (H2O2) and iron (Fe2+ to Fe3+) to produce reactive oxygen species (ROS), such as highly reactive hydroxyl radicals, which can non-specifically cleave lignin66. Since AA6 in the CAZy database (all annotated AA6 are from fungi) are suspected to produce Fenton reagents51, the bacterial community may follow a route similar to that of brown-rot fungi to effectively solubilize increasing concentrations of switchgrass (Fig. 4C). Although proteins annotated for direct production of H2O2 were not identified in the samples, multiple oxidases were observed (Supplementary Data 8). Additionally, hallmarks of oxidative stress such as superoxide dismutases67,68 (SOD2) and superoxide reductases68 (SOD2) were also present in the samples, both of which increased significantly in abundance with increasing solids in one (SOR, in SB fraction) or all (SOD2) fractions (Fig. 4C, D, Supplementary Fig. 13C, D, Supplementary Data 9). Furthermore, bacterioferritins from multiple taxa were also detected and significantly increased in all three fractions (Fig. 4d, Supplementary Fig. 13C). Bacterioferritins are enzymes with ferroxidase activity; ferroxidases are upregulated in brown-rot fungi during Fenton reactions.69. Additionally, as is the case in brown-rot fungi, limited proteins for lignin metabolism/assimilation were observed in the samples, indicating lignin cleavage may enhance enzyme accessibility rather than be used for energy (Supplementary Fig. 15–20). Notably and in contrast to fungal AA6s which are known to act aerobically, these bacterial AA6s appear to have an anoxic mode of operation, which has also been observed in a recent study70. Specifically, these results suggest that in bacterial communities, AA6 enzymes (here produced by Firmicutes) could be involved in the generation of Fenton reagents and subsequent cleavage of the lignin barrier to increase the accessibility of lignocellulose to attack by CAZymes.

**Discussion**

The thermophilic, lignocellulose-fermenting, methanogenic anaerobic microbiome reported here exhibits a key feature desired for an industrial process: undiminished fractional carbohydrate solubilization with increasing substrate loading (Fig. 1). In search of clues as to how this is achieved, this study reports the first metaproteomic characterization of a microbiome as a function of substrate loading across different fractions, focusing on salient CAZyme and methanogenesis features. Comparing relative protein abundances at 150 g/L and 30 g/L substrate loadings in Fig. 5A, a greater than 2-fold increase was observed in one or more fractions for β-glucosidases, hemicellulose-debranching and xylosidase enzymes, pectinases, and AA6 enzymes, as well as all hydrogenotrophic methanogenesis pathway enzymes. The latter corroborates the observed increase in methane production as solids increased (Fig. 1C), and as described before in anaerobic thermophilic microbiomes,71 is the preferred methanogenesis route under thermophilic temperatures72,73. Coupled with syntrophic acetate oxidation for acetate degradation, this metabolic process is performed by members of Euryarchaeota, who may not need to be proximal to lignocellulosic substrate. As can be seen from Fig. 5B, this process primarily occurred in the cell containing fractions, PC and SB, with the bulk of the activity in the PC fraction for 150 g/L.

Considering enzymatic deconstruction of lignocellulose, the largest increases were seen for β-glucosidases in the SNT, pectinases in the PC and SB fractions, and for AA6 family, debranching, and xylosidase enzymes in all fractions. The observation that the microbiome produces these particular enzymes instead of increasing amounts of backbone depolymerizing enzymes (such as endoglucanases and xylanases) suggests their importance for undiminished hydrolysis at high solids—potentially through removal of accumulated inhibitive (hemi) cellulose solubilization products (β-glucosidases and xylosidases) and enabling enzymatic access to the preferred carbohydrate polymers. For example, debranching of xylan “decorations” alters its interaction with other cell-wall polymers (cellulose and lignin) effectively decreasing the recalcitrance of feedstock74. Pectinases/polygalacturonases synergistically improve the hydrolytic efficiency of cellulosates by removing pectin, and thereby improve access to cellulose75–77. Finally, the marked increase in AA6 enzymes—purported to utilize Fenton chemistry for deconstruction and permeation of the lignin barrier—fits this paradigm. These AA6 family enzymes are also particularly notable since oxidative reactions are not generally associated with anaerobic lignocellulose fermentation. However, Schalk et al.33 recently documented a role for fungal driven AA6-mediated Fenton chemistry in the termite gut. Although further investigation is needed in the mechanism proposed here, the observations suggest that AA6 plays an important synergistic role with other CAZyme categories for high solids deconstruction under anaerobic thermophilic conditions. Our results also provide insight into the spatial location of bacterial enzymes, their phylogenetic origin(s), and how these are impacted by substrate loading (Fig. 5B). At the substrate loading of 30 g/L, β-glucosidases and AA6 enzymes are relatively more prominent in the substrate-bound fraction, but at 150 g/L become most prominent in the supernatant. In contrast, hemicellulose-debranching enzymes and pectinases are most prominent in the supernatant at 30 g/L, but in the substrate-bound fraction at 150 g/L. The clear difference in functional distribution across fractions highlights the importance of this
Fig. 5 Microbiome functionally and spatially orchestrates enzyme expression to continue solubilization at high solids. A Average fold changes (ratios) in aggregate protein abundance of the major enzymatic categories at 150 g/L solids loading relative to 30 g/L solids loading compared across the three fractions. B The relative proportion of the respective fraction metaproteome that each of these enzymatic categories make up is shown here. This qualitative representation depicts the spatial distribution of these enzymatic categories among the three fractions and changes in their distribution from 30 g/L to 150 g/L. SNT: supernatant, PC: planktonic cells, and SB: substrate bound fraction. Source data are provided as a Source Data file.

A decrease in the collective protein synthesis effort of the microbiome towards enzymes that act on compounds known to be inhibitors of deconstruction - cellobiose, hemicellulose, pectin, and lignin, with roughly constant effort devoted to the CAZymes that mediate mainline (hemi)cellulolytic deconstruction. These observations provided by metaproteomics provide an unprecedented level of molecular detail which will be critical to inform development of defined cultures for conversion of cellulosic biomass to products other than methane, which thus far exhibit decreasing fractional carbohydrate solubilization with increasing substrate loading.

Methods

Feedstocks and microbiome origin. Mid-season switchgrass (Panicum virgatum L, Cave in Rock) harvested in June 2012 at Rock Springs Research Farm (Spring Mills, PA, 40.71040290971864°N, −77.94974506560891 °W) was used as the substrate in this study. Characterization, preparation and storage information is described in detail in Liang et al.44. An existing microbial community was used in this study, as described in Liang et al. The inoculum for the initial enrichment was obtained by sampling the anaerobic digester from Vermont Technical College (Randolph, Vermont). The enrichment was matured as a microbiome for over 120 days before experiments with different residence times were initiated, the results of which are described in Liang et al.44. Reactor HR1 (formerly named R2 in Liang et al.44) ran at various residence times (RT) and was cultivated for a total 214 days at 30 g/L solids loading before it was transitioned to 75 g/L to generate the additional higher solids data for the study described here. A detailed characterization of the microbiome can be found in Liang et al.44. Potential microbial contributions via the addition of unsterilized feedstock to the microbiome were investigated by Liang et al.44, and found to be small.

Cultivation medium. Mixed enrichment (ME) medium was used in this study44, except that the amount of Wolfe’s vitamin solution was doubled for 75 g/L and up compared to 30 g/L. The following stock solutions were used: Wolfe’s modified elixir stock solution89 (50×), Wolfe’s vitamin solution89 (50×), Ammonium & Phosphate stock solution (50×) and Ferrous iron stock solution (1000×) are described in Liang et al.44. The solutions were either autoclaved (Ferrous iron stock solution) or filter-sterilized (all other stock solutions). Switchgrass was added (without sterilization) to the reactor to a concentration of 30, 75, 120, or 150 g/L (as-is basis with 6.05% moisture) according to the solids loading per feeding period. The pH of the medium was adjusted with 1 N Sodium Hydroxide solution after feeding to maintain a value above 6.0 as needed.

Cultivation system. The bioreactor set-up used for this study was a Qplus multi-bioreactor system (Sartorius Stedim, Bohemia NY) which includes Module Operator Service Program (MFCs) data collection software (3.0, level 43, 2008).
Cultivation start-up and operation. As described in Liang et al., prior to this study the bioreactor was operated semi-continuously for 214 days, fed with medium containing 30 g/L switchgrass at various residence times (20–3.3 days). For an RT of 10 days and 30 g/L solids loading the total cultivation time was 50 days of operation as described in Liang et al., and fundamental data and samples for subsequent analysis for that solids loading were incorporated into the results described here. For the current study the reactors were fed with medium containing increasing loadings of switchgrass starting at 75 g/L, 120 g/L and 150 g/L. At the start of this study the solids loading was increased from 30 to 75 g/L and the RT fixed at 10 days (from 3.3 days). The frequency of slurry removal and feeding was maintained at 10 times per RT resulting in 1 feeding sampling event each day for the 10-day residence time and the length of the overall cultivation time.

The microbiome was monitored for 506 days after transition to 75 g/L and RT = 10 days. It was operated for 204 days at 75 g/L switchgrass loading in the feed and then fed 120 g/L switchgrass for 85 days. Finally, the solids loading was increased to 150 g/L and maintained at this level for 217 days. The reactor went through medium component optimization during the first 150 days at 75 g/L solids loading and RT = 10 days and was fed with the medium described earlier from 151th day onwards. This condition was considered to start the steady-state regime. For all the other conditions, the reactor was considered to have reached steady state after 3 RTs. Total reactor operating length and steady state length at each condition is summarized in Supplementary Table 1 and in Supplementary Fig. 1. Slurry was withdrawn via a 50 mL pipette by removing a plug from the head-plate as described in Liang et al., and the slurry was collected at 20-min intervals at normal operating conditions. All the measured liquid-solids-volume indication was adjusted accordingly (total volume equaling 50 mL).

Biogas measurement. Measurement of biogas production rate was conducted using a wet tip gas meter (www.wettipgasmeter.com) filled with acidified water (pH < 2). At each feeding, data was recorded either manually or via a data logger (HOBO Pendant Event, 64 K, Onset Computer Corporation, Pocasset, MA). The concentrations of CH4, CO2, H2 and N2 were measured by a Model 310 Education Gas Chromatograph (SRI Instruments, Torrance, CA) and a 1.8 m x 3 mm stainless steel Haysep D packed column at 50 °C with a thermal conductivity detector (150 °C) using helium (20 mL/min) for CH4, CO2 and N2 and nitrogen (13.5 mL/min) for H2 as carrier gas. For data see Supplementary Table 2 and Supplementary Fig. 1.

Fractional carbohydrate solubilization. Measurement of fractional carbohydrate solubilization, FCS, is based on the amount of glucose, xylose and arabinose in the solids before and after fermentation. Sugar content is determined by acid hydrolysis via the quantitative saccharification protocol. Whole broth samples were centrifuged for 10–11 min at 2,800 × g (30–120 g/L) or 12,000 × g (150 g/L). Pelleted solids were dried overnight in a 60 °C oven (30–120 g/L) or by lyophilization (150 g/L), weighed and then subjected to quantitative saccharification. The hydrolysis products monomeric glucose, xylose and arabinose were then quantified by HPLC (Waters, Milford, MA) with an Amex HPLC-878 column (Bio-Rad, Hercules, CA) at 60 °C and detected by refractive index. HPLC eluent was 5 mM sulfuric acid with a flow rate of 0.6 mL/min. FCS equals to the monomeric reducing sugars of the final carbohydrates (g/L) as measured by the initial carbohydrate. Steady state FCS data as shown in Fig. 1b is based on averages of two or more samples after at least 3 residence times following a change in solids loading. For data see Supplementary Table 2 and Supplementary Fig. 1.

Quantification of volatile fatty acids. Measurement of volatile fatty acids (VFA) was done by analyzing filtered broth samples for formic acid, acetic acid, propionic acid, butyric acid, iso-butyric acid and valeric acid. All samples were performed in duplicate against a known standard (Valeric-acid-free acid mix standard, 46975-U SUPELCO, Sigma-Aldrich). Analysis was by HPLC with an Amex HPLC-878 column as described for Fractional Carbohydrate Solubilization. For data see Supplementary Table 2 and Supplementary Fig. 1.

Assembly of the bioreactor gene catalog for metaproteomics analysis. The metagenomic data generated in Liang et al., which described the development of a stable switchgrass-fermenting microbiome at various residence times and temperatures, was examined for functional gene profile diversity. Metagenomic sequence reads of 15 samples were downloaded from the US Department of Energy’s Joint Genome Institute website under proposal ID 502980 (https://www.jgi.doe.gov/portal/cms/PortfolioPage/18392769545.html). All metagenomic sequence files were combined into a single file in a conglom- erative format to capture as many prokaryote coding regions as possible. Quality filtering and trimming was performed using Atropos with a minimum sequence length of 50 nucleotides and a minimum Phred score of 30 to ensure only the highest quality segments were included. The high-quality sequences were then assembled into 506,732 contigs (a total of 777,641,587 bp using MEGAHIT and prokaryote genes identified using Prodigal). The minimum contig size was 200 bp and the maximum contig was 868,950 bp with an average of 1,533 bp in length and 87.4% coverage by Prodigal's prokaryote gene finder. All these contigs. The genes were then assembled into a bioreactor gene catalog in FASTA format for use in metaproteomics analysis. The bioreactor gene set comprised of 485,820 full length prokaryote genes and 590,333 fragmented prokaryote genes. For metaproteomics analysis, only the full-length genes were kept for inclusion in the bioreactor gene atlas to ensure high quality results.

Sample preparation, fractionation, preparation, and measurements for 2D-LC/MS/MS-based metaproteomics. Switchgrass fermentation whole-broth samples for each of the four solids loadings (30 g/L, n = 4; 75 g/L, n = 3; 120 g/L, n = 4; and 150 g/L, n = 5) at steady-state conditions were collected and stored at –80 °C for analysis. For metaproteome measurements, samples were thawed under cold running water to keep the temperature at 4 °C. Each sample was then fractionated into three phases by centrifugation: (a) the protein fraction enriched for the microbial proteome and enzymes adhering to the microbial substrate (or substrate bound fraction- “SB”), (b) the planktonic microbes (or planktonic cells fraction- “PC”), and proteins which were secreted (supernatant fraction- “SNT”) as described. For centrifugation-mediated enrichment, each sample was centrifuged at 200 × g for 10 min to pellet the residual solids substrates and bound microbial matter (SB fraction), and the resulting pre-supernatant primarily containing the free cells and enzymes was transferred into a new tube. This pre-supernatant was further centrifuged at 1,000 × g for 20 min. The resulting pellet was the PC fraction, enriched for unbound microbial cells and the resulting supernatant was the SNT fraction, enriched for free proteins. Each segregated fraction was then analyzed separately for metaproteomics.

Each sample combination of loading and fraction was lyophilized by bead beating with 0.15 mm zirconium oxide beads in Tris-HCl (100 mM at pH 8.0) containing 4% SDS (sodium dodecyl sulfate, Sigma) and 10 mM DL-Dithiothreitol (Sigma). Resulting lysates were precleared by centrifugation at 21,000 × g for 10 min, incubated at 90 °C for 10 min to denature proteins, and adjusted to 30 mM IAA (iodoacetamide, Sigma) for a 20-minute incubation prior to proteinase K digestion. Samples were filtered through the concentrated supernatant were removed and collected tryptic peptides. The resulting peptide solution was acidified with formic acid (FA; LC/MS grade) to a final concentration of 1% to precipitate remaining SDC, followed by additional removal of the precipitate using water-saturated ethyl-acetate. Peptides were then concentrated by drying under a SpeedVac, resuspended in 0.5% (v/v) FA/acetonitrile (FACN)/1% (v/v) FA, and analyzed by 2D-LC/MS/MS using a Vanquish UHPLC system (Thermo Scientific) with autosampler coupled to an Orbitrap Q Exactive Plus mass spectrometer (Thermo Scientific).

Fixed amounts of peptides depending on the complexity of the fraction: 10 μg for peptides for SB, 14 μg for PC, and 6 μg for SNT fractions, respectively, were loaded into an in-house built triphasic MUDPIF back column (100 μM ID packed with RP-SCX; RP– reversed-phase C18 resin, 5 μm Kinetex, Phenomenex; SCX–strong cation exchange, 5 μm Luna) coupled to an in-house pulled nanospray emitter (75 μm ID) packed with 30 cm of 5 μm RP C18 resin for online 2D HPLC separation. Peptides from each sample type were then trapped, desalted, separated, and analyzed over successive salt cuts of ammonium acetate and formic acid (100–500 mM), each followed by an organic gradient to elute peptides. The eluting peptides were measured and sequenced by the mass spectrometer (operated via Xcalibur v.4.2.4.7, Thermo Scientific) in data-dependent mode.

Metaproteomics data analysis. The acquired peptide fragmentation spectra were searched against the metaproteome database generated from the 30 g/L samples’ metagenomes (as described above) appended with common contaminant proteins, and the switchgrass (Panicum virgatum) proteome employing a target decoy approach using the MS Amdanda algorithm (v2.0) integrated in Proteome Discoverer software (version 2.3.0.523, Thermo Scientific). The resulting peptide spectrum matches (PSMs) were required to be at least 5 amino acids long, fully tryptic with a maximum 2 missed cleavages, contain a static modification of 57.0214 Da on cysteine (carbamidomethylation) and a maximum mass modification of 15.9949 Da on methionine (oxidized) residues. False-discovery rates (FDRs) were controlled at 1% (IMP-Elutator node in Proteome Discoverer) at both
packed with 15 cm of 1.7-micron Kinetex C_{18} reversed-phase resin (Phenomenex). Eluting peptides were measured and sequenced by data-dependent acquisition. Peptide fragmentation spectra were searched against the concatenated databases and quantified using Proteome Discoverer software as described above. A total of 14,386 peptides from 26,127 detected peptides were considered quantifiable after removal of low confidence hits and contaminants and were used for further analyses (Supplementary Data 1). Functional and taxonomic annotations of the peptides were derived by mapping back to the corresponding proteins. Peptide abundances for peptides mapping to a taxonomic or functional category were summed to identify trends of the category with solids loadings. In-house scripts were used to ascertain the taxonomic source of a peptide if it mapped to >1 GO term with at least taxonomic origins based on GhostKOALA assignment. Peptides were qualified as microbial (Bacteria or Archaea), plant (Switchgrass or other plant sequences), or mixed origin (indistinguishable). Peptides from cellulolytic organisms were determined if they mapped to proteins from CAZymes as determined by dbCAN2. Peptides from docken were determined from InterProScan. Peptides from methanogenic organisms were determined if they belonged to Euryarchaeota proteins. Pairwise comparisons for each solids loadings against the 30 g/L condition were conducted using Welch’s t-test and a cutoff of p-value ≤ 0.05 was used. Python and R libraries were used to generate figures.

**Data availability**

The data on solubilization and microbiome performance generated in this study is available in the supplementary information files. All proteomics raw mass spectra for protein quantification in this study are available at the ProteomeXchange Consortium via the MassIVE repository (MassIVE accession: MSV000088319 [https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=0b51547b0c1045795125da310d3a6a]; ProteomeXchange accession: PXD029682). All proteome abundance data along with the master annotations are available as Supplementary Data files. Source data are provided with this article.

Received: 4 January 2022; Accepted: 16 June 2022; Published online: 05 July 2022

**References**

1. Lynd, L. R. The grand challenge of cellulosic biofuels. *Nat. Biotechnol.* 35, 912–915 (2017).

2. Lynd, L. R., Wyman, C. E. & Gerngross, T. U. Biocommodity engineering. *Biotechnol. Prog.* 15, 777–793 (1999).

3. Himmel, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807 (2007).

4. Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. Microbial cellulose utilization: Fundamentals and biotechnology (vol 66, pg 506, 2002). *Microb. Mol. Biol. R.* 66, 739–739 (2002).

5. Modenbach, A. A. & Nokes, S. E. Enzymatic hydrolysis of biomass at high-solids loadings: a review. *Biomass. Bioenergy* 56, 526–544 (2013).

6. Chen, X. W. et al. DMR (deacetylation and mechanical refining) processing of corn stover achieves high monomeric sugar concentrations (230 g L\(^{-1}\)) during enzymatic hydrolysis and high ethanol concentrations (>10% v/v) during fermentation without hydrolysate purification or concentration. *Energy Environ. Sci.* 9, 1237–1245 (2016).

7. Lynd, L. R. et al. Toward low-cost biological and hybrid biological/catalytic conversion of cellulosic biomass to fuels. *Energy Environ. Sci.* 15, 938–990 (2022).

8. Jørgensen, H., Vibe-Pedersen, J., Larsen, J. & Helby, C. Liquefaction of lignocellulose at high-solids concentrations. *Biotechnol. Bioeng.* 96, 862–870 (2007).

9. Hodge, D. B., Karim, M. N., Schell, D. J. & McMillan, J. D. Model-based fed-batch for high-solids enzymatic cellulose hydrolysis. *Appl. Biochem. Biotechnol.* 152, 88–107 (2009).

10. Du, J. et al. Enzymatic liquefaction and saccharification of pretreated corn stover at high-solids concentrations in a horizontal rotating bioreactor. *Bioprocess Biosyst. Eng.* 37, 173–181 (2014).

11. Liotta, F. et al. Effect of total solids content on methane and volatile fatty acid production in anaerobic digestion of food waste. *Waste Manag Res* 32, 947–953 (2014).

12. Liotta, F. et al. Modified Anaerobic Digestion Model No.1 for dry and semi-dry anaerobic digestion of solid organic waste. *Environ. Technol.* 36, 870–880 (2015).
ABBASSI-GUENDOUZ, A. et al. Total solids content drives high solid anaerobic digestion via mass transfer limitation. Bioreour. Technol. 111, 55–61 (2012).

Wang, H. et al. Establishing practical strategies to run high loading corn stover anaerobic digestion: methane production performance and microbial communities. Bioresour. Technol. 130, 1, http://dx.doi.org/10.1016/j.biortech.2013.10.025 (2014).

Motte, J. C. et al. Total solids content: a key parameter of metabolic pathways in dry anaerobic digestion. Biotechnol. Biofuels 6, Art. 164 10.1186/s41164-1754-6834-6-164 (2013).

Du, J. et al. Identifying and overcoming the effect of mass transfer limitation on feedstock solubilization and biofuel production by Clostridium thermocellum. Biotechnol. Biofuels 11, 1–8 (2018).

Li, C., Mortelmaier, C., Winter, J. & Gallert, C. Effect of moisture of municipal biowaste on start-up and efficiency of mesophilic and thermophilic dry anaerobic digestion. Bioreour. Technol. 168, 23–32 (2014).

Verbeke, T. J., Garcia, G. M. & Elkins, J. G. The effect of switchgrass loadings on seedling solubilization and biofuel production by Clostridium thermocellum. Biotechnol. Biofuels 10, 1–9 (2017).

Shao, X. J., Murphy, S. J. & Lynd, L. R. Characterization of reduced carbohydrate solubilization during Clostridium thermocellum fermentation with high switchgrass concentrations. Biomass Bioenergy 139, ARTN 105623, https://doi.org/10.1016/j.biombioe.2020.105623 (2020).

Holwerda, E. K. et al. Metabolic and evolutionary responses of Clostridium thermocellum to genetic interventions aimed at improving ethanol production. Biotechnol. Biofuels 13, 40 (2020).

Holwerda, E. K. et al. The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading. Biotechnol. Biofuels 7, 155 (2014).

Bayer, E. A., Lamed, R. & Himmel, M. E. The potential of cellulosomes and cellulosomes for cellulose waste management. Curr. Opin. Biotech. 18, 237–245 (2007).

Caparel, R. L. et al. The Carbohydrate-Active Enzymes database (CAZY): an expert resource for Glycogenomics. Nucleic Acids Res 37, D233–D238 (2009).

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZY) in 2013. Nucleic Acids Res 42, D490–D495 (2014).

Nelson, M. C., Morrison, M. & Yu, Z. A. meta-analysis of the microbial diversity observed in anaerobic digesters. Bioresour. Technol. 102, 3730–3739 (2011).

Sundberg, C. et al. 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. FEMS Microbiol. Ecol. 85, 612–626 (2013).

Ma, S. et al. A microbial gene catalog of anaerobic digestion from full-scale biogas plants. Gigascience 10, https://doi.org/10.1093/gigascience/giaa164 (2021).

Allgayer, M. et al. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. PLoS ONE 5, e8812 (2010).

D’Haeseleer, P. et al. Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. PLoS ONE 8, e68465 (2013).

Lillington, S. P., Leggieri, P. A., Heorn, K. A. & O’Malley, M. A. Nature’s recyclers: anaerobic microbial communities drive crude biomass deconstruction. Curr. Opin. Biotech. 62, 38–47 (2020).

Lim, J. W., Park, T., Tong, Y. W. & Yu, Z. The microbiome driving anaerobic digestion and microbial analysis. Adv. Bioenergy 5, 1–61 (2020).

Kougias, P. G. et al. Spatial distribution and diverse metabolic functions of lignocellulose-degrading uncultured bacteria as revealed by genome-centric metagenomics. Appl. Environ. Microb. 84, https://doi.org/10.1128/AEM.01244-18 (2018).

Campanaro, S. et al. Metagenomic analysis and functional characterization of the biogas microbe using high throughput shotgun sequencing and a novel binning strategy. Biotechnol. Biofuels 9, 26 (2016).

van der Leelie, D. et al. The metagenome of an anaerobic microbial community decomposing poplar wood chips. PLoS ONE 7, e36740 (2012).

Contet-Marre, S. et al. Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the ruminants of cow fed a mixed diet. ISME J. 11, 2538–2535 (2017).

Svartström, O. et al. Ninety-nine de novo assembled genomes from the moose (Alces alces) rumin panoramic images provide new insights into microbial plant biomass degradation. ISME J. 11, 2538–2535 (2017).

Campanaro, S. et al. New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. Biotechnol. Biofuels 13, 25 (2020).
65. Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28**, 1883–1896 (2011).

66. Arantes, V., Jellison, J. & Goodell, B. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Appl Microbiol. Biot.* **94**, 323–338 (2012).

67. Volf, P., Schmidt-Weber, B. C. & Pfefferle, S. Enzymatic decolorization of delignified birch pulp: -O- and -OH radical scavenging. *Biochem. Biophys. Acta Gen. Substr.* **1865**, 472–480 (2016).

68. Slesak, I., Slesak, H. & Kruk, J. Oxygen and hydrogen peroxide in the early stages of cottonwood xylem. *Front. Microbiol.* **7**, 1518 https://doi.org/10.3389/fmicb.2016.01518 (2016).

69. McGivern, B. B. et al. Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat. Commun.* **12**, 12466 (2021).

70. Hagen, L. H. et al. Quantitative Metaproteomics Highlight the Metabolic Contribution of Uncultured Phytoparasitic Fungi in a Thermophilic Anaerobic Digestor. *Appl. Environ. Microbiol.* **83**, https://doi.org/10.1128/AEM.01955-16 (2017).

71. Dyksma, S., Figueiredo, R., Jelitto, J. & Layont, J. Enzymatic degradation of lignin by aerobic fungi and the role of ligninolytic enzymes. *Appl. Environ. Microbiol.* **85**, 3161–3170 (2019).

72. Timmers, P. A. H. A. et al. Metabolism and occurrence of methanogenic and sulfate-reducing syntrophic acetate oxidizing communities in halokline environments. *Front. Microbiol.* **9**, 3039 (2018).

73. Lopes, A. M., Ferreira, E. X. & Moreira, L. R. S. An update on enzymatic cocktails for lignocellulose breakdown. *J. Appl. Microbiol.* **125**, 632–645 (2018).

74. Pakarinen, A., Hämäläinen, J., Räihä, K. & Vepsäläinen, M. Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol. Bioeng.* **86**, 503–511 (2004).

75. Hallwell, G. & Griffin, M. Nature and mode of action of cellulolytic component C1 of Trichoderma koningii on native cellulose. *Biochem. J.* **135**, 587–594 (1973).

76. Chen, M. et al. Strategies to reduce end-product inhibition in family 48 glycoside hydrolases. *Proteins* **84**, 295–304 (2016).

77. Kumar, R. & Wyman, C. E. Strong cellulase inhibition by Mannan polysaccharides in cellulose conversion to sugars. *Biotechnol. Bioeng.* **111**, 1341–1353 (2014).

78. Qiu, Y., Li, X., Wang, Y., Wang, X. & Wu, X. Hydrolysis of cellulose by mixed cellulases from *Rhizopus oligosporus*. *Appl. Biochem. Biotechnol.* **161**, 3217–3227 (2010).

79. Zhu, X. et al. Deletion of a gene cluster encoding pectin degrading enzymes in *Caldericellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. *Biotechnology for Biofuels* **11**, 284 (2018).

80. Qin, L. et al. Inhibition of lignin-derived phenolic compounds to cellulase. *Biotechnology for Biofuels* **9**, ARTN 70 https://doi.org/10.1186/s13068-016-0485-2 (2016).

81. Li, X. et al. Inhibitory effects of lignin on enzymatic hydrolysis: the role of lignin chemistry and molecular weight. *Renew. Energ.* **123**, 664–674 (2018).

82. Rahikainen, J. L. J. et al. Inhibitory effect of lignin during cellulose bioconversion: The effect of lignin chemistry on non-productive enzyme adsorption. *Biotechnol. Bioeng.* **133**, 270–278 (2013).

83. Kubis, M. R., Holwerda, E. K. & Lynd, L. R. Declining carbohydrate accessibility of lignin during cellulose bioconversion by *Clostridium thermocellum*: documentation and diagnostic tests. *Biotechnol Biofuels* **15**, ARTN 12 https://doi.org/10.1186/s13068-012-02110-4 (2022).

84. Xiao, C. W. & Anderson, C. T. Roles of pectin in biomass yield and processing. *Front. Plant Sci.* **4**, ARTN 67 https://doi.org/10.3389/fpls.2013.00067 (2013).

85. Qin, L. et al. Inhibition of lignin-derived phenolic compounds to cellulase. *Biotechnology for Biofuels* **9**, ARTN 70 https://doi.org/10.1186/s13068-016-0485-2 (2016).

86. Li, X. et al. Inhibitory effects of lignin on enzymatic hydrolysis: the role of lignin chemistry and molecular weight. *Renew. Energ.* **123**, 664–674 (2018).

87. Kubis, M. R., Holwerda, E. K. & Lynd, L. R. Declining carbohydrate accessibility of lignin during cellulose bioconversion: The effect of lignin chemistry on non-productive enzyme adsorption. *Biotechnol. Bioeng.* **133**, 270–278 (2013).

88. Rahikainen, J. L. J. et al. Inhibitory effect of lignin during cellulose bioconversion: The effect of lignin chemistry on non-productive enzyme adsorption. *Biotechnol. Bioeng.* **133**, 270–278 (2013).

89. Lovley, D. R., Greening, R. C. & Ferry, J. G. Rapidly growing rumen methanogenic organism that synthesizes coenzyme M and has a high affinity for formate. *Appl. Environ. Microbiol.* **48**, 81–87 (1984).

90. Seaman, J. F., Bubl, J. L. & Harris, E. E. Quantitative Sacrification of Wood and Cellulose. *Ind. Eng. Chem. Anal. Ed.* **17**, 35–37 (1945).

91. Sluiter, A. & National Renewable Energy Laboratory (U.S.). Determination of structural carbohydrates and lignin in biomass: laboratory analytical procedure (LAP); issue date, 4/25/2008, https://purl.fdlp.gov/GPO/LPS94089.

92. Walker, C., Ryu, S., Giannone, R. J., Garcia, S. & Trinh, C. T. Understanding and Eliminating the Detrimental Effect of Thiamine Deficiency on the Oleaginous Yeast Yarrowia lipolytica. *Appl. Environ. Microbiol.* **86**, https://doi.org/10.1128/AEM.02299-19 (2020).

93. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

94. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

95. Baath, T. S. et al. Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell Proteom.* **18**, 1027–1035 (2019).

Acknowledgements

Oak Ridge National Laboratory: Jason Witham (bioinformatics), Dawn M. Klingeman (sample preparation), Steven D. Brown (coordination) and James Elkins (microbial community insights). The Pennsylvania State University: Tom Richard (microbial enrichment insights), Dartmouth College: Xiaojun Shao (assistance in experimental and method design), Sean Murphy, Jules Wheaton, Lion Hertford, Liang Tian and Anela Ari (assistance with sample analysis and microbiome operation). Funding was provided by the BioEnergy Science Center and the Center for Bioenergy Innovation, both at the U.S. Department of Energy (DOE) Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Author contributions

E.K.H., L.R.L., P.C., R.J.G., and R.L.H. designed the study, X.L. and E.K.H. performed and analyzed the fermentation experiments. P.C., S.P., and R.J.G. conducted the metaproteomics measurements. J.C.E. co-assembled the metagenomes and generated the metaproteome database. P.C., R.J.G. and R.L.H performed all the metaproteomics data analyses. E.K.H. and Y.J.B. helped in the interpretation of CAZyme results, Y.J.B. co-wrote the paper. P.C., E.K.H., R.J.G., X.L., R.R.L and L.R.L. wrote the paper. All authors edited and reviewed the paper.

Competing interests

L.R.L. is a shareholder in a startup company focusing on cellulosic biofuel production. All other authors declare no competing interests.

Additional information

**Supplementary information** The online version contains supplementary material at https://doi.org/10.1038/s41467-022-31433-x.

**Correspondence** The Pennsylvania State University: E. Keith Hettich (ehettich@psu.edu) and Leo R. Lynd (lynd@psu.edu).

**Peer review information** *Nature Communications* thanks Jagroop Pandhal and Fabio Squina for their contribution to the peer review of this work. Peer reviewer reports are available.

**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.