
Article:
Mello, Bruno, Alessi, Anna, Riaño-Pachón, Diego et al. (6 more authors) (2017) Targeted metatranscriptomics of compost derived consortia reveals a GH11 exerting an unusual exo-1,4-β-xylanase activity. Biotechnology for biofuels. 254. ISSN 1754-6834

https://doi.org/10.1186/s13068-017-0944-4

Targeted metatranscriptomics of compost derived consortia reveals a GH11 exerting an unusual exo-1,4-β-xylanase activity

Bruno L. Mello[1], Anna M. Alessi[2], Diego M. Riaño-Pachón[3,4], Eduardo R. deAzevedo[1], Francisco E. G. Guimarães[1], Melissa C. Espirito Santo[1], Simon McQueen-Mason[2], Neil C. Bruce[2], Igor Polikarpov[1]

[1]Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, São Carlos, SP, 13560-970, Brazil, e-mails: bruno.luan.mello@usp.br; azevedo@ifsc.usp.br; guimarae@ifsc.usp.br; melissaespiritosanto@yahoo.com.br; ipolikarpov@ifsc.usp.br

[2]Department of Biology, University of York, Wentworth Way, York, UK, YO10 5DD, emails: anna.szczepanska@york.ac.uk; simon.mcqueenmason@york.ac.uk; neil.bruce@york.ac.uk

[3]Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Rua Giuseppe Máximo Scalfaro 10000, Campinas, SP, 13083-100, Brazil.

[4]Laboratório de Biologia de Sistemas Regulatórios, Departamento de Química, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP, 05508-000, Brazil, email: diriano@gmail.com

Corresponding author Igor Polikarpov, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, São Carlos, SP, 13560-970, Brazil

Phone +55 16 3373-8088, fax +55 16 3373-9881, email: ipolikapov@ifsc.usp.br

Abstract

Background: Using globally abundant crop residues as a carbon source for energy generation and renewable chemicals production stands out as a promising solution to reduce current dependency on fossil fuels. In nature, such as in compost habitats, microbial communities efficiently degrade the available plant biomass using a diverse set of synergistic enzymes. However, deconstruction of lignocellulose remains a challenge for industry due to recalcitrant nature of the substrate and the inefficiency of the enzyme systems available, making the economic production of lignocellulosic biofuels difficult. Metatranscriptomic studies of microbial communities can unveil the metabolic functions employed by lignocellulolytic consortia and identify new biocatalysts that could improve industrial lignocellulose conversion.

Results: In this study, a microbial community from compost was grown in minimal medium with sugarcane bagasse sugarcane bagasse as the sole carbon source. Solid-state nuclear magnetic resonance was used to monitor lignocellulose degradation; analysis of metatranscriptomic data led to the selection and functional characterization of several target genes, revealing the first glycoside hydrolase from Carbohydrate Active Enzyme family 11 with exo-1,4-β-xylanase activity. The xylanase crystal structure was resolved at 1.76 Å revealing the structural basis of exo-xylanase activity. Supplementation of a commercial cellulolytic enzyme cocktail with the xylanase showed improvement in Avicel hydrolysis in the presence of inhibitory xylooligomers.

Conclusions: This study demonstrated that composting microbiomes continue to be an excellent source of biotechnologically important enzymes by unveiling the diversity of enzymes involved in *in situ* lignocellulose degradation.

Keywords

Metatranscriptomics; xylanase; lignocellulose; compost; microbial community; bioethanol.

Background

The accelerated rate of fossil fuel depletion and concerns over global warming has triggered the search for renewable energy sources. Lignocellulose is the basic component of plant cell walls and one of the most abundant sources of carbon in the biosphere. Therefore, its bioconversion into liquid fuels represents a promising solution for energy generation [*1-3*].

In recent years, direct DNA extraction techniques from microbial communities coupled with next generation sequencing of metagenomes has given an unprecedented insight into microbial taxonomic groups and their interactions [*4*]. Metagenomic libraries also represent a vast resource for the discovery of enzymes with industrial applications.

Although many free-living organisms deconstruct plant biomass by enzyme-driven oxidation and hydrolysis [*5*], this bioprocess remains a formidable challenge for industry. One of the main obstacles to industrial-scale production of second-generation biofuel lies in the inefficient deconstruction of plant material, due to the recalcitrant nature of the substrate and relatively low activity of currently available enzymes [*6*].

Seeking to overcome these challenges, previous studies have sequenced and functionally characterized microbial communities from different biomass-degrading

environments. Examples include microbial communities from compost [*7-9*], bovine rumen [*6, 10*], guts of animals [*11-16*], soil [*17, 18*] and river water [*19*]. These studies have revealed the lignocellulolytic capabilities of microbial communities present in diverse ecosystems and the highly complex and cooperative interactions between multiple microbial species and their enzymes to achieve lignocellulose breakdown.

In this study, a compost-derived microbial community was grown in minimal medium supplemented with sugarcane bagasse as a sole carbon source aiming to enrich lignocellulose-degrading microorganisms. We monitored deconstruction of sugarcane bagasse using scanning electron microscopy, solid state nuclear magnetic resonance (ssNMR) spectroscopy and confocal microscopy. To obtain information on the community response to this submerged *in-vitro* environment, metatranscriptomic analysis was performed. A number of predicted genes that showed similarity to carbohydrate active enzymes (CAZymes) were selected for expression leading to the discovery of the first exo-1,4-β-xylanase from glycoside hydrolase family 11 (GH11). This enzyme was able to degrade xylooligomers, which are known inhibitors to commercially available cellulase cocktails [*20*], as well as xylan, yielding xylobiose as the only reaction product.

Results

Compositional and morphological changes in sugarcane bagasse

The compositional analysis of sugarcane bagasse collected weekly from *in vitro* composting cultures was investigated using ssNMR. In order to obtain ^{13}C quantitative spectra in an achievable measuring time, the spectra were acquired using the multiple cross polarization pulse sequence (Multi-CP) under fast (14 kHz) Magic

100 Angle Spinning [*21*]. Using Multi-CP, the integral of each signal in the NMR spectra
101 is proportional to the amount of the corresponding chemical group in the sample.
102 Therefore, quantitative information on sample relative composition can be obtained if
103 a reliable identification of the NMR lines is available [*21, 22*]. The complete
104 assignment of the signals can be found in references [*23-31*].

105 Fig. 1a shows the sugarcane bagasse ^{13}C Multi-CP spectra with spectral
106 regions specifically assigned to three major lignocellulose components (cellulose,
107 hemicellulose and lignin) highlighted. After normalizing the spectra by the total area,
108 we used the integrals over the specified regions to estimate the cellulose,
109 hemicellulose and lignin fractions in the sample [*32*]. The plot of this relative
110 percentage is depicted as a function of the growth weeks. Due to microbial growth
111 and enzymatic activities of the composting cultures, a gradual reduction in the relative
112 amount of cellulose was observed. The relative percentage of hemicellulose remained
113 mostly constant, with some fluctuation attributed to experimental uncertainties.
114 Consistently, the relative amount of lignin increased at the same rate that cellulose
115 decreased.

116 Lignin concentration and arrangement before and after microbial growth was
117 investigated by confocal imaging microscopy using two-photon excitation [*33*]. The
118 analysis showed no change in the emission spectra after five weeks of microbial
119 growth (Fig. 1b-f). This observation supports the ssNMR results that our compost
120 microbial communities promoted insignificant lignin degradation. It also suggests that
121 the structure or organization of the lignin residues remained mostly unmodified.

122 Further we obtained scanning electron microscopy images of the sugarcane
123 bagasse before and after five weeks of microbial growth. Control biomass sample (no
124 inoculum) showed a smooth, continuous surface with cohesive, well-defined

lignocellulose fibers (Fig. 2). In contrast, five weeks of microbial growth caused a complete loss in the biomass integrity, with separation of the fibers and decrease of particles' size.

Functional and phylogenetic characteristics of the sugarcane bagasse degrading microbial community

To examine the transcriptional responses of the sugarcane bagasse degrading microbial community, we performed RNA-seq metatranscriptomic analysis on weekly cultures grown for up to five weeks. Although it yielded 66 million paired-end reads (Table 1), the rarefaction analysis showed that the sequencing did not reach saturation (Fig. S1). The rarefaction analysis also revealed that week 1 cultures were more diverse than week 5 cultures since a larger proportion of new reads was obtained at the same sequencing depth. Resulting high quality, non-ribosomal RNA reads (63%) were *de novo* assembled into 302,961 transcripts and used to predict biochemical capabilities of the microbial community by mapping putative transcripts to the KEGG orthology. The distribution of genes classified to KEGG functions presented a similar profile for all time points (Fig. 3). Transcripts assigned to translation processes (mean = 8.5%) showed the highest relative abundance, followed by genes involved in energy, carbohydrate and amino acids metabolism and signal transduction (mean from 6 to 8%). We observed that the microbial community was more actively growing and breaking down the lignocellulose in initial stages of culture since a number of transcripts assigned to energy and carbohydrate metabolism were higher in week 1 compared to week 5 cultures.

Next, the metatranscriptome libraries were screened using HMMER alignment tool [*34*] and the dbCAN database [*35*] for genes encoding putative CAZymes

involved in lignocellulose degradation. From the 283,356 predicted open reading frames (ORFs), 2.5% (number of sequences = 7,196) showed homology to CAZymes. The CAZymes were distributed between carbohydrate binding modules (34.4%), glycoside hydrolases (34.3%), glycosyl transfereses (19.3%), carbohydrate esterases (15.9%), auxiliary activities (3.3%), polysaccharide lyases (2.2%) and cohesin and dockerin modules (0.8%). The expression level for the majority of CAZyme classes was higher in week 1 cultures (Fig. 4). KEGG assigned transcripts involved in carbohydrate metabolism presented a similar profile. Transcripts predicted as glycosyl transferases showed contrasting behavior with the highest expression level at week 5.

Since the most expressed CAZy-assigned transcripts were predicted as glycoside hydrolases (GHs), we analyzed the distribution and phylogenetic origin of these ORFs in details (Fig. 5). Endoglucanases (GH5, GH74) acting on the cellulose backbone and β-glucanases (GH3) involved in cellobiose hydrolysis, showed high expression at week 1, followed by a gradual decline over time (Fig. 5a). Similar profile was observed for endoxylanases (GH10, GH11) and hemicellulose debranching enzymes such as arabinofuranosidases (GH43). On the contrary, annotated lysozymes and chitinases from GH19 and GH25 families displayed higher expression in the later time points, indicating that these cell wall lytic enzymes might be associated with an increase of competitive interactions between microbial species in the later stages of culture. Phylogenetic origin of predicted GHs was also investigated (Fig. 5b). Proteobacteria expressed the majority of GH5 endoglucanases, GH11 endoxylanases and GH19 lysozymes, whereas Bacteroidetes members were major producers of GH13 amylases, GH23 lysozymes, GH43 hemicellulose debranching enzymes and GH109 α-N-acetylgalactosaminidase. The GH3-assigned transcripts were mostly expressed by species of Verrucomicrobia phylum. Some GH

families were also predicted to derive from eukaryotes. Specifically, starch and glycogen degrading enzymes of GH13 family were highly expressed by eukaryotes from the Animalia kingdom. Whereas, majority of GH25 lysozymes were not assigned beyond Domain level. The distribution of prokaryotic and eukaryotic origin of CAZymes (all classes) was further examined by the relative expression level of predicted enzymes assigned to specific phylogenetic level. Among the 7,196 predicted CAZyme genes, 75% were taxonomically assigned to phylum level. It revealed that the microorganisms most actively involved in carbohydrate modification belonged to Bacteroidetes and Proteobacteria (Fig. S2). Those lineages accounted for more than 65% of the bacterial diversity over all time points. The expression of CAZymes affiliated to Bacteria dropped from 20,171 to 6,465 fragments per kilobase of transcript per million (FPKM) over the five-week time course. During this period genes encoding putative CAZymes of eukaryotic origin showed a dramatic increase by 30-fold. At week 5, 82% of genes expressed by Eukaroytes were assigned to the kingdom Animalia with only 3% to be predicted as fungal genes.

In order to quantify the relative percentage of bacteria to fungi in the sugarcane bagasse degrading community, real time PCR was performed. Note two differences with the results described in the above paragraph: the phylogeny is regarding the entire community, not only CAZymes; the primers used to capture the Eukaryotic component of the community are specific to fungi. Overall Bacteria dominated the composting community (Table S1) but the fungal/bacterial ratio in weekly samples gradually increased from 5 to 20% by the third week and stabilized in older cultures.

Heterologous expression and characterization of putative CAZymes

Following, metatranscriptome analysis we selected 27 GH assigned transcripts with predicted cellobiohydrolase, endoglucanase and xylanase activities and higher expression levels for functional characterization (Table S2). Following initial recombinant expression screening, seven proteins (26% efficiency) were obtained in the soluble fraction of *E. coli* transformed cultures of which three proteins named compost7_GH6, compost13_GH10 and compost21_GH11 showed an enzymatic activity after assaying against a variety of polysaccharide substrates (Fig. S3).

BLASTP results against the NCBI-nr database showed that the protein compost7_GH6 has 49% identity to a GH6 β-1,4-glucan cellobiohydrolase from *Sorangium cellulosum*. Substrate screening showed that compost7_GH6 had activity against β-glucan and lichenan. No enzymatic activity was detected towards filter paper and carboxymethyl cellulose (CMC). compost7_GH6 displayed highest activity towards β-glucan at pH 10.0 and was able to retain 70% or more activity until the pH dropped to 4.0 (Fig. S3). Thermal shift assays (see method, Fig. S4) confirmed the alkaliphilic behavior of the enzyme with the highest thermostability at pH 6 to 9. The optimal temperature was assayed at both pH 6.0, where the enzyme was found to maintain >70% maximum activity, and 10.0. The highest activity at pH 6.0 and 10.0 was found at 50 and 45 °C, respectively. Next, the optimal pH and temperature were used to test enzyme specificity. The results showed that compost7_GH6 had highest specific activity towards β-glucan (2.0 U/mg) and lichenan (1.5 U/mg) amongst the substrates tested (Fig. S3c). We also determined that compost7_GH6 maintained 50% of its initial activity after 24 h incubation at 45 °C at pH 6.0 and 100% activity after 96 h incubation at 40 °C at pH 10.0.

The protein C13 was predicted as a member of GH10 family and had 91% and 89% identity to an endoglucanase and endo-β-1,4-xylanase from *S. cellulosum*,

respectively. C13 showed endo-xylanase activity against xylan, which was subsequently used as a substrate to determine the enzyme's optimal pH and temperature. C13 displayed highest activity at pH 6.0 and retained >50% of its activity over a broad pH range (pH 3.0 - 10.0) (Fig. S3c). The optimum temperature for enzyme activity was 65 °C at pH 6.0. These conditions were subsequently applied to test the enzyme specificity. The highest specific activity was found for xylan (25 U/mg) and arabinoxylan (11 U/mg). No activity was found against CMC. The residual activity study performed at 50 °C, pH 6.0 revealed that the enzyme retains more than 60% of its initial activity up to 96 h of incubation, demonstrating considerable thermal stability (Fig. S4).

The protein compost21_GH11 was predicted as a GH11 family member and shared 77% identity with a non-characterized GH from *Marinimicrobium agarilyticum*. The fully characterized homologue of compost21_GH11, was a β-1,4-xylanase from *S. cellulosum*, with 40% identity. compost21_GH11 was found to be an exo-1,4-β-xylanase with highest activity against xylan at pH 6.0 at 35 °C. The enzyme retained more than 60% activity for all tested pHs, but the observed activity quickly dropped at temperatures higher than 40 °C. However, at 35 °C and pH 6.0, compost21_GH11 retained 90% activity for up to 96 h. The activity screen against a number of polysaccharides revealed that compost21_GH11 was active towards xylan only. Using this substrate, compost21_GH11 showed a high specific activity of 320 U/mg even at the relatively low reaction temperature (35 °C).

Characterization and structure of compost21_GH11

Since compost21_GH11 presented no activity against substrates with a xylan backbone such as AZCL-linked xylan, a substrate specific for endo-xylanases due to

its cross-linked structure and dye labels [*36*], or arabinoxylan, this protein was chosen as a target for further characterization. To investigate the mode of action of compost21_GH11, we analyzed the reaction products by thin layer chromatography (TLC) and Dionex HPLC (Fig. 6). TLC results showed that compost21_GH11 was acting on xylan and xylooligosaccharides liberating xylobiose as the only reaction product. Dionex HPLC confirmed this hydrolysis pattern. Testing the enzyme with 4-nitrophenyl-β-D-xylopyranoside displayed no activity, confirming that the enzyme was not able to hydrolyze xylobiose.

In order to understand the molecular basis of exo-1,4-β-xylanase activity, the crystal structure of compost21_GH11 was solved at 1.76 Å resolution. The data collection and refinement statistics are summarized in Table 2. The final model was constructed from the first to last residue of the crystallized protein, which was cloned without the first 13 residues that were predicted as a disordered region and reported as a signal peptide [*37*]. Despite its low identity to the closest homologue in PDB (ID 1XNK, 36% identity), compost21_GH11 presents a typical GH11 fold. It consists of one α-helix and 15 β-sheets, labeled A1 to A6 and B1 to B9 (Fig. 7a). The curvature of the β-sheets B1 to B9 forms a cleft where the catalytic residues (nucleophile Glu98 in β-sheet B6 and proton donor Glu200 in β-sheet B4) are located. Two extra loops (EL1 and EL2) are found in the compost21_GH11 structure. They are created by additional residues, which stretch between β-sheets B5 and B6 and in β-sheet B4, respectively. To gain insight into substrate binding, 1XNK, which was co-crystallized with a modified xylotrioside, was aligned to compost21_GH11 (root mean square deviation of 0.898 Å when EL1 and 2 are ignored). EL2 blocks substrate interaction at subsite -3, which accommodates the non-reducing end of xylan (Fig. 7d). This loop is stabilized by a number of hydrogen bonds within the loop main/side chain residues

and van der Walls interactions, especially between Pro192 and Phe186 side chains. Hydrogen bonds are also established with EL1 and with the turn that connects β-sheets B7 and B8. The presence of EL1 seems essential to maintain EL2 in correct position by steric hindrance.

compost21_GH11 activity improves performance of a commercial enzyme mixture

To examine the effect of xylooligomers on cellulose digestion, we monitored the hydrolysis of 2% (w/v) Avicel by a commercial enzymatic cocktail in a presence or absence of oligosaccharides. The reactions improvements were further tested by supplementing the reactions with compost21_GH11 protein. Fig. 8a shows that xylooligomers strongly inhibited Accellerase activity, especially at the initial time points. The addition of xylooligomers resulted in activity decrease of 96.5% at 1h; after 96h, the activity decrease was 45%. This indicates that the enzymes present in the commercial cocktail were able to degrade, to some extent, the xylooligomers, reducing their inhibitory effect. When Accellerase was supplemented with compost21_GH11, the addition of xylooligomers decreased the activity by 83% at 1h; after 96h, the activity decrease was 53%. Hence, inhibition was much lower, particularly at initial stages (where a 380 times difference is seen between compost21_GH11 supplemented and non-supplemented reactions). Dionex HPLC demonstrated that xylooligomers with polymerization degree higher than 4 were depleted from compost21_GH11 non-supplemented reactions after 24h whereas it took only 4h for the xylooligomers to be depleted when supplemented with compost21_GH11. Since the xylooligomer concentration was reduced in both reactions, they reached approximately the same conversion after 96h.

Discussion

Plant cell walls are effectively degraded in various natural ecosystems by the action of microorganisms that act cooperatively by secreting an array of lignocellulolytic enzymes. In recent years, metatranscriptomic analysis applied to these ecosystems has begun to provide an insight into how lignocellulose breakdown is accomplished *in situ* [*2, 38, 39*].

Here we investigated the time course degradation of sugarcane bagasse by a microbial community derived from compost. Based on sugarcane bagasse biomass analysis we showed that the lignin component remained mostly unchanged and was not significantly modified by microbial activities. Our analysis was in agreement to a previous study [*1*] showing that biomass loss is mostly attributed to cellulose and hemicellulose degradation. Despite this limitation, composting community remained metabolically active during the experiment as surveyed by RNA sequencing.

Next, we explored the metatranscriptome-assembled library, by focusing on screening the resulting database for CAZymes. Although the predicted GHs accounted for a small fraction (0.87%) in our composting community metatranscriptome, this was similarly observed by others who investigated various lignocellulytic communities such as rice-straw enriched compost (0.97%) [*1*], soil-contacting sugarcane bagasse (0.97%) [*3*], termite lumen (0.78%) [*11*], bovine rumen (0.78%) [*10*] and macropod foregut (0.71%) [*14*]. Amongst GHs, oligosaccharide-degrading enzymes from GH3 family were highly expressed in our study. These enzymes are fundamental in lignocellulolytic processes [*5*] and were abundant in other lignocellulolytic environments [*40*]. Mhuantong *et al.* explored the metagenome of a microbial community extracted from soil-contacting sugarcane bagasse [*3*]. Six

out of the 10 most abundant GH families in the reported metagenome are amongst the most expressed GHs in each week of our metatranscriptome. Therefore, despite the different environments and techniques used, these communities have a reasonable level of similarity. Enzymes from auxiliary activity families, attributed to lignin oxidative modification and lytic polysaccharides degradation, accounted for a very small fraction (3.3%) of all predicted CAZymes in our metatranscriptome, supporting the lack of sugarcane bagasse lignin removal or structural changes in this polymer. This could be associated with a low relative abundance of fungi in our composting cultures, especially in early stages of the time course. Experimental design that uses liquid culturing of compost inoculum could have an effect on fungal growth and hence ligninolytic enzymes expression [*41*]. Recent studies showed that the composting conditions without liquid phase were preferable for CAZymes enrichment [*42*]. Other factors such as medium composition [*43*], temperature, agitation and inoculum source could also play critical role for suppressing fungal growth.

Majority of CAZymes predicted in our studies had bacterial origin, similar in composition and structure to other studies [*2, 3, 17, 40, 44*]. Our community was dominated by a metabolically diverse Proteobacteria and Bacteroidetes. As observed previously, Proteobacteria dominates oxygenated habitats [*3*] and Bacteroidetes are known for their contribution to the largest reservoir of CAZymes in various environments [*3, 40*].

Interestingly, in the later stages of composting process, CAZymes expression shifted towards Eukaryotes and Animal kingdom. Representatives of nematodes, protists and other groups will be present in a composting spot, but it is unlikely that they would survive weeks in the submerged cultures. One explanation can be that the algorithm LCA did not assign phylogeny correctly. Also de novo assembly of

350 metatranscriptomics reads and their mapping without reference genome can produce
351 errors. However, in recent years, an increasing evidence of Eukaryotic invertebrates
352 showed their critical role in the hydrolysis of plant cell wall [*45-48*] and this aspect of
353 our work should be further investigated.

354 Our comprehensive analysis led to identification of potentially, novel CAZy
355 proteins. The recombinant expression efficiency in this work demonstrates the
356 challenge that remains in characterizing novel genes derived from culture-
357 independent approaches using heterologous systems. The solubility was confirmed for
358 three target proteins (26%) but was lower than the 53% rate usually obtained in our
359 laboratory using the same expression system [*49*]. A β-1,3-(4)-glucanase with
360 specificity towards substrates with higher β-1,3 to β-1,4 ratio [*50, 51*] and no activity
361 for CMC was found in our study. compost7_GH6 was highly tolerant to an alkaline
362 environment, an essential characteristic for application in detergent industry [*52*].
363 compost13_GH10 presented substrate specificity and hydrolysis profile of a typical
364 endoxylanase [*53-55*]. In contrast, the enzyme compost21_GH11 presented a
365 hydrolysis profile of a typical exo-enzyme, releasing xylobiose from xylan and
366 xylooligosaccharides. The structure of compost21_GH11 (Fig. 7b-d) shows a typical
367 GH11 fold of a β-jelly-roll [*56-60*]. The architecture of other GH11 members show
368 the same pattern with little variation in the secondary structures lengths [*56*]. Despite
369 28 structures from GH11 members have already been solved, the compost21_GH11
370 structure reveals two extra loops previously unseen in the other family members.
371 However, multiple alignment analysis revealed that there are many other proteins that
372 might have these extra loops. Here we show that loop EL2 blocks one side of
373 compost21_GH11 active site, transforming this enzyme into an exo-1,4-β-xylanase
374 that acts from the non-reducing end. To our knowledge, the present study describes

375 the first example of an exo-xylanase from the GH11 family. compost21_GH11 has
376 high activity on insoluble polymeric xylan, in contrast to GH8 exo-oligoxylanases
377 that show preference for soluble xylooligosaccharides [*61, 62*].

378 It has been reported that xylooligosaccharides are strong cellulase inhibitors,
379 whereas xylose and xylobiose have a smaller inhibitory effect [*63*]. As commercial
380 enzymatic cocktails might have insufficient xylanase activity, a significant amount of
381 xylooligomers accumulates in the reaction [*20*]. Hence, supplementation of enzyme
382 cocktails with compost21_GH11 proved to increase their performance when there are
383 xylooligomers in the reaction mixture. Therefore, in biomass treatment processes
384 where xylooligomers accumulate [*20*], supplementing cocktails with
385 compost21_GH11 will improve enzyme performance.

386

387 ## Conclusions

388 In summary, our results indicate the ability of sugarcane bagasse adapted
389 microbial community in deconstructing lignocellulosic biomass by removing the
390 cellulose and hemicellulose fractions. The taxonomic binning and expression profile
391 of GHs illustrate the degradation of lignocellulosic biomass complexity. Phylogenetic
392 analysis also suggested a growing participation of eukaryotic microorganisms in this
393 process, indicating that the organisms studied up to now may not represent the major
394 organisms that degrade plant biomass in nature. Expression of genes selected from the
395 metatranscriptome library revealed challenging. However, considering the industrially
396 appealing features of proteins described here, we proved the importance of this line of
397 study. The isolated enzymes warrant further study to characterize their structure and
398 verify their ability to enhance commercially available cocktails, as have been
399 proposed.

Methods

Sample collection and culture

Composting samples were collected from the São Paulo University Recycling Project (São Carlos campus) during the final mesophilic phase at locations 30 cm below the surface. A 1% (w/v) homogenized composting sample was used to inoculate minimal medium [*64*]. Cultures were supplemented with 3% (w/v) sugarcane bagasse and incubated at 30 °C with 150 rpm agitation for up to five weeks. Sugarcane bagasse was kindly provided by the Cosan Group (Ibaté, São Paulo, Brazil) and prior to use, it was washed and dried at 50 °C. Weekly sampling was performed on three biological replicates. Sugarcane bagasse and microbial biomass were separated from culture supernatant by centrifugation at 3,000 x g for 5 minutes at room temperature and used for nucleic acid extraction and biomass analysis.

Biomass analysis

Prior analysis, sugarcane bagasse obtained from composting cultures was washed, dried and ground to a fine powder using ball milling (TissueLyser II, Qiagen, Hilden, Germany) for 60 sec at 30 Hz. The raw sugarcane bagasse was used as a control.

NMR experiments were performed using a Bruker Avance 400 spectrometer, equipped with a Bruker 4-mm magical angle spinning double-resonance probe, at ^{13}C and ^{1}H frequencies of 100.5 and 400.0 MHz, respectively. The spinning frequency at 14 kHz was controlled by a pneumatic system that ensures a rotation stability higher than ~1 Hz. Typical $\pi/2$ pulse lengths of 4 and 3.5 μs were applied for ^{13}C and ^{1}H, respectively. Proton decoupling field strength of $\gamma B_1/2\pi$ = 100 kHz was used. ^{13}C quantitative spectra were measured by using the Multi-CP excitation method

described by Johnson and Schmidt-Rohr [*21*]. A total of nine cross-polarization blocks were implemented with 1 ms and RF amplitude increment (90-100%), while the cross-polarization before acquisition was executed with 0.8 ms and the same amplitude increment. The recycle delay was 2 s and the duration of the repolarization period t_z was 0.9 s [*22*]. To obtain the fraction of cellulose, hemicellulose and lignin, the spectra were normalized with respect to their area and integrated over the specific regions for cellulose (1 and 1′), lignin (2 and 2′) and hemicellulose (3 and 3′). The relative percentage of each component was obtained by dividing the calculated value for the biomass after microbial growth with the value obtained in the control sample. Finally, this fraction was multiplied by the initial percentage of the component in the raw sugarcane bagasse, as reported by Lima *et al.* [*32*]. Chemical shifts were assigned based on published studies [*23-31*]. Approximately 4,000 scans were measured to acquire each spectrum. Chemical shift was assigned based on published studies [*23-31*].

Confocal microscopy was performed using a Zeiss LSM 780 confocal inverted microscope with a Coherent Chameleon laser (Ti:sapphire) as source for two-photons (2P) excitation at 800 nm. Ground sugarcane bagasse was hydrated for 24 h and observed with a C-Apochromat objective lens (20X, numerical aperture 0.8); the images were taken in the opposite side of the cover slip. The images were obtained by the average of 2 scans and no appreciate variation was observed. The spatial resolution was approximately 350 nm (considering the numerical aperture and the wavelength of excitation).

Scanning electron microscopy was performed using a scanning electron microscope model JSM-6390 LV (Jeol, Tokyo, Japan) operating with a 5 kV accelerating voltage. Ground sugarcane bagasse was hydrated for 24 h prior analysis;

a drop was directly applied to the sample pedestal and dried at room temperature for 12 h. After drying, samples were gold coated using a metalizer model MED 020 (Baltec, Liechtenstein). Images were obtained under vacuum. At least 10 images per sample were acquired from different areas to certify the reproducibility of the results.

Nucleic acid extraction from sugarcane composting cultures

A culture pellet (0.5 g of sugarcane bagasse and microbial cells) was used for cell lysis and nucleic acid extraction following a protocol modified from Griffiths *et al.* [*65*]. Briefly, 0.5 g pellet was added to 2 mL screw-cap tubes containing 0.5 g of acid washed 0.1 mm glass and 0.5 mm silica beads (each). 500 µL of CTAB extraction buffer (10% w/v hexadecyltrimethylammonium bromide in 700 mM NaCl mixed with an equal volume of 240 mM potassium phosphate buffer, pH 8.0) and 500 µL of phenol:chloroform:isoamyl alcohol (25:24:1) (pH 8.0) was added. Samples were lysed in a Bead Ruptor 24 (Omni, Kennesaw, GA, USA) for 30 s at 5.5 m/s and centrifuged at 17,000 x g for 5 min at 4 °C. The top aqueous phase was transferred to a new tube and extracted with an equal volume of chloroform:isoamyl alcohol (24:1) followed by centrifugation at 17,000 x g for 5 min at 4 °C. Total nucleic acid was precipitated with two volumes of PEG solution (30% w/v polyethelene glycol 6000 with 1.6 M NaCl) for 2 h at room temperature. Pellet was obtained by centrifuging the solution at 17,000 x g for 20 min at 4 °C and washed twice with 70% ethanol. The nucleic acids were suspended in 50 µL of water and stored at -80 °C until use. Its quality was determined with a LabChip GXII (PerkinElmer, Waltham, MA, USA).

All solutions and glassware were treated with 0.1% DEPC overnight at 37 °C under homogenization and autoclaved to create and RNase-free environment. Only certified RNase- and DNase-free plasticware was used.

475

476 **Real-time PCR of isolated genomic DNA**

477 The nucleic acid extracted from composting cultures was diluted to 200 ng/μL
478 and treated with 1:100 (v/v) RNase A:nucleic acid (Thermo Fisher Scientific,
479 Waltham, MA, USA) for 15 min at 37 °C. The metagenomic DNA was extracted with
480 phenol:chloroform:isoamyl alcohol and precipitated with 1/10 volume of 3M sodium
481 acetate pH 5.2 and 3 volumes of ethanol. Samples were incubated at 4 °C for 30 min
482 and centrifuged at 17,000xg for 30 min at 4 °C. Supernatant was discarded. The
483 isolated metagenomic DNA was washed with 70% ethanol and suspended with water.

484 Real-time PCR was performed using the metagenomic DNA and universal
485 primer sets for bacterial (515F and 806R) [*66*] and fungal (ITS1 and 5.8S) [*67*]
486 rDNA. PCR reactions contained 20-μL mixture of the following: 2.5 ng DNA, 300
487 nM of each forward and reverse primer and 10 μL of KAPA SYBR® FAST qPCR
488 Master Mix (KAPA Biosystems, Wilmington, MA, USA), which contained all the
489 nucleotide, polymerase, reaction buffer and SYBR green dye. The thermocycling
490 conditions were as follows: an initial hold at 95 °C for 5 min followed by 35 cycles of
491 95 °C for 30 s and 60 °C for 45 s, according to KAPA Biosystems recommendation.
492 Measurements were done using a CFX96 Real-Time System (Bio-Rad, Hercules, CA,
493 USA). All reactions were performed in triplicate during two independent experiments.

494

495 **cDNA library synthesis and sequencing**

496 The nucleic acid extracted previously was diluted to 200 ng/μL and treated
497 with DNase I (Invitrogen, Waltham, MA, USA) according to manufacturer's
498 recommendation. Equimolar volumes of the extracted RNA from biological replicates
499 were combined and the Prokaryotic ribosomal RNA (rRNA) was depleted with

RiboZero Magnetic Kit Bacteria (Epicentre). The remaining RNA was purified using the RNA Clean & Concentrator-5 kit (ZymoResearch, Irvine, CA, USA). TruSeq Stranded Total RNA Sample Preparation kit (Illumina, San Diego, CA, USA) was used to deplete Eukaryotic rRNA and to synthesize a ~450 bp cDNA library. Sequencing of each time-point cDNA library was performed on a MiSeq with a 500-cycles Reagent kit v2 (Illumina, San Diego, CA, USA).

Metatranscriptomics assembly and annotation

Sequenced reads were preprocessed with Trimmomatic [*68*] to remove adaptors, low quality and short sequences. SortMeRNA [*69*] was used to merge and remove contaminant ribosomal RNA sequences, which were identified using Silva [*70*] and Rfam [*71*] reference databases with an e-value cutoff of 1×10^{-5}. Non-ribosomal RNA reads were *de novo* assembled with Trinity [*72*] and genes were predicted using TransGeneScan [*73*]. Expression levels were calculated with eXpress [*74*] and Bowtie2 [*75*]. The phylogenetic origin of predicted ORFs was analysed using MEGAN v6 [*76*] and the Lowest Common Ancestor algorithm at default values. HTSeq [*77*] software was used to subsample the sequenced reads. Rarefaction curves were plotted using the software BBMap (sourceforge.net/projects/bbmap/) for sequencing depth analysis. Functional annotation was performed with HMMER alignment tool [*34*] against the dbCAN database [*35*]. Kyoto Encyclopedia of Genes and the Genomes (KEGG) [*78*] orthology classification was performed using the online tool GhostKOALA [*79*] and default values.

Target genes cloning, expression and purification

Twenty seven predicted CAZymes were selected for expression studies. The genes were codon optimized for *Escherichia coli* expression (https://www.idtdna.com/CodonOpt) and synthesized (GenScript, Jiangsu, China) after the predicted signal peptide and transmembrane helix was removed. Additionally, adapters were added to the 5' (CAGGGCGCCATG) and 3' (TAACCGCGTCGGGTC) sequence ends to allow cloning using ligation independent cloning (LIC) [*80*]. Standard molecular biology techniques were applied [*81*]. The gene fragments were cloned to pETTRXA-1a/LIC plasmid [*49*] and transformed into *E. coli* Rosetta (DE3) pLys cells (Merck, Darmstadt, Germany). Small-scale protein expression and solubility assays were performed as described previously [*49*]. Recombinant cells were stored at -80 °C in presence of 20% (v/v) glycerol.

Cells were grown overnight in the LB medium in presence of kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL) at 150 rpm shaking at 37 °C. 1 L of LB medium was inoculated with overnight culture (1% v/v) and incubated under 150 rpm shaking at 37 °C until the optical density at 600 nm reached 0.8. Incubation temperature was decreased to 17 °C and expression was carried out for 16 hours after induction with 0.5 mM IPTG. The cells were harvested at 9,000 x g for 20 minutes and resulting pellet was resuspended in 20 mL of lysis buffer (20 mM Tris-HCl, 300 mM NaCl, 5 mM imidazole, 5% (v/v) glycerol, 10 mM β-mercaptoethanol (β-ME), 1 mM phenylmethylsulfonylfluoride (PMSF), 0.2 mg/mL lysozyme, pH 8.0) with or without 0.6% (w/v) sarkosyl. Cells suspension was incubated on ice for 1 hour, sonicated for 6 minutes and centrifuged at 23,000 x g for 30 minutes. Supernatant was loaded on a column with 2 mL of nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen, Hilden, Germany) previously equilibrated with 10 volumes of lysis buffer. The column was washed with 4 volumes of wash buffer (20 mM Tris-HCl, 300 mM NaCl,

549 5 mM imidazole, 5% (v/v) glycerol, 10 mM β-ME, 1 mM PMSF pH 8.0) and in-
550 column digestion was performed by adding cleavage buffer (20 mM Tris-HCl, 300
551 mM NaCl, 5% (v/v) glycerol, 10 mM β-ME, 1 mM PMSF, pH 8.0) and 1:50
552 TEV:protein (measured by 280 nm absorbance). After overnight incubation at 10 °C
553 with homogenization, the recombinant protein was eluted in the flow-through. The
554 column was further washed with 4 volumes of elution buffer (20 mM Tris-HCl, 300
555 mM NaCl, 300 mM imidazole, 5% (v/v) glycerol, 10 mM β-ME, 1 mM PMSF, pH
556 8.0). The protein was further purified using Superdex™ 75 16/60 (GE Healthcare
557 Biosciences Corporation, Picataway, USA) column previously equilibrated with 20
558 mM Tris-HCl, 200 mM NaCl, pH 8.0. The protein purity was determined by sodium
559 dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie
560 blue staining [*82*].
561
562 **Sequence analysis and enzyme characterization**
563 Multiple alignment of amino acid sequence was performed with Clustal
564 Omega (http://www.ebi.ac.uk/Tools/msa/clustalo) [*83*]. Enzyme activity was
565 determined by the amount of reducing sugars released from polysaccharide
566 (Megazyme, Ireland; Sigma-Aldrich, St. Louis, MO, USA) using the DNS method
567 [*84*]. Xylan and other substrates (see Fig. S3c) were used at 1% and 0.5% (w/v) final
568 concentration, respectively. Glucose was used as a standard. All assays were
569 performed in quadruplicate. Initial activity assays were performed at mild conditions
570 using an array of substrates. Further enzyme assay was run at pHs ranging from 2 to
571 10 in 20 mM ABF buffer (20 mM of each sodium acetate, sodium borate and sodium
572 phosphate dibasic; pH adjusted with HCl/NaOH) using optimal substrate. The
573 reaction temperature was screened from 30 to 85 °C using the optimal substrate and

574 pH. Finally, activity was screened against different polysaccharides at enzyme
575 optimal pH and temperature conditions. Residual activity was tested by incubating the
576 enzyme at optimal buffer pH for up to 48 h. Aliquots were removed and activity
577 assays at optimal conditions were performed.
578
579 **Thermal stability analysis using ThermoFluor**
580 To investigate the effect of pH on thermal stability, the protein was mixed
581 with Sypro Orange (Invitrogen, Waltham, MA, USA), a reporter dye that binds
582 nonspecifically to hydrophobic regions of the protein. Because water quenches the
583 fluorescence of this dye, the fluorescence signal increases after the protein unfolds,
584 allowing to monitor the melting curve. The experiment was performed on a CFX96
585 Real-Time System (Bio-Rad, Hercules, CA, USA) with excitation and emission
586 wavelengths of 490 and 530 nm respectively. 20 µL reactions were prepared with 0.2
587 mg/ml enzyme in different buffer solutions and 2,000 times diluted dye. This mixture
588 was added to a 96-well thin wall PCR plate (Bio-Rad, Hercules, CA, USA) and sealed
589 with optical-quality sealing tape (Bio-Rad, Hercules, CA, USA). All buffers used in
590 the analysis were prepared at 50 mM. The temperature scan was from 25 to 90 °C,
591 with stepwise increments of 1 °C per minute. The melting temperature determination
592 and analysis were performed using GraphPad Prism software v5.0 (GraphPad
593 Software, La Jolla, CA, USA).
594
595 **Identification of enzymatic product on TLC and Dionex HPLC**
596 The enzymatic reaction products were analyzed by TLC on silica gel 60 F254
597 (Merck, Darmstadt, Germany) with n-butanol:acetic acid:water (2:1:1, v/v) as eluent.
598 The plates were developed with exposure to 10% (v/v) sulphuric acid in ethanol

599 followed by charring. A mixture of xylooligosaccharides with 2 to 6 xylose residues
600 (X2-X6) and xylose (X1) was used as standard.

601 Reaction products were also analyzed on a High-Performance Anion-
602 Exchange chromatography with Pulsed Amperometric Detection (HPAE-PAD). The
603 experiment was performed on a Dionex ICS-5000 Ion Chromatography system
604 equipped with an electrochemical detector, a CarboPac PA1 (4x250 mm) anion
605 exchange column and guard cartridge (Thermo Fisher Scientific, Waltham, MA,
606 USA). The following program was used: flow 1 mL/min, 30 °C, isocratic 100 mM
607 NaOH, [segment 1] 5-20' from 0 to 20 mM CH_3COONa, [segment 2] 20-24' up to
608 100 mM CH_3COONa, [segment 3] 24-30' isocratic 100 mM CH_3COONa.

609

610 **Xylooligosaccharide production and quantification**

611 Xylooligosaccharides were produced following a protocol modified from Qing
612 *et al.* [*63*]. Briefly, 5% (w/v) birchwood xylan was sealed in a 0.15 L stainless steel
613 reactor. The reactor was transferred to a sand bath at 330 °C. After the temperature
614 reached equilibrium at 200 °C, it was incubated for 10 min and quickly cooled in ice
615 water. Solids were removed by centrifugation at 4,000 x g for 10 min and filtration
616 through a 0.45 μm hydrophilic polyvinylidene fluoride (PVDF) filter (Merck,
617 Kenilworth, NJ, USA). Dionex HPLC was employed to verify the xylooligomer
618 distribution. The xylooligosaccharides were hydrolyzed in 4% (v/v) sulfuric acid for 1
619 h at 121 °C based on standard protocol from the National Renewable Energy
620 Laboratory (NREL) [*85*]. Total oligomer concentration was determined using a HPLC
621 (Shimadzu LC-20AT, Kyoto, Japan) equipped with refractive index and UV-VIS
622 detectors and an aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA).
623 Xylose standards were treated in parallel and used to calculate the sample

concentration. The program used was: flow rate 0.6 mL/min, 65 ºC, isocratic 5 mM H_2SO_4.

Xylooligosaccharide hydrolysis inhibition

2% (w/v) Avicel hydrolysis by Accellerase 1500 (DuPont, Wilmington, DE, USA) was performed in 50 mM sodium citrate pH 5.0 to access xylooligosaccharides inhibition (at 8 g/L) and activity improvement by compost21_GH11 addition (at 0.1 g/L). 0.02% (w/v) sodium azide was used to prevent microbial growth. Reactions were started by addition of Accellerase 1500 (DuPont, Wilmington, DE, USA) and Novozyme 188 (Sigma-Aldrich, St. Louis, MO, USA) diluted 25 and 250 times to give 5 FPU/g and 10 CBU/g respectively. The flasks were incubated at 35 °C with 150 rpm agitation. Substrate blanks without enzyme and enzyme blanks without substrate were also set. 0.5 mL aliquots taken after 1, 4, 7, 24, 48, 72 and 96 h of hydrolysis were immediately boiled for 10 min to inactivate enzymes, centrifuged at 10,000 x g for 1 min, filtered through a 0.45 μm PVDF filter and frozen at -20 °C. Two independent experiments were performed. The cellulose conversion yields were analyzed with HPLC, as discussed before, using glucose standards. This experimental set up was based on NREL standard protocol [*86*].

Crystallization and data collection

Crystallization conditions were screened for crystal growth using a HoneyBee crystallization robot 931 (Genomic Solutions, Ann Arbor, MI, USA) and commercial available screens. Crystals were obtained for protein compost21_GH11 in multiple conditions. Crystals grown at 18 °C in 0.1 M BIS-TRIS propane pH 7.5, 20% (w/v) PEG 3350, 0.2 M NaI were briefly soaked in a cryoprotective solution (crystallization

649 solution with 15% (v/v) ethylene glycol added) and flash-cooled in a gaseous nitrogen
650 steam at 100 K. The diffraction data was collected at the MX2 beamline [*87*] of the
651 Brazilian National Synchrotron Laboratory (LNLS, Campinas, Brazil) using
652 synchrotron radiation with wavelength set to 1.459 Å, PILATUS2M detector (Dectris,
653 Taefernweg, Switzerland) and an oscillation of 0.5° per frame. Diffraction data was
654 reduced and integrated with XDS [*88*].
655
656 **Molecular replacement, model building and structure refinement**
657 The molecular replacement, structure model building, refinement and
658 validation were performed using PHASER [*89*], PHENIX [*90*], Coot [*91*] and
659 MolProbity program [*92*]. PDB ID 1XNK was used as template. PyMOL (the
660 PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) was used for
661 structure representations.
662

663 ## List of abbreviations

664 CAZyme, carbohydrate active enzyme; CMC, carboxymethyl cellulose; EL, extra
665 loop; FPKM, fragments per kilobase of transcript per million; GH, glycoside
666 hydrolase; KEGG, Kyoto Encyclopedia of Genes and the Genomes; Multi-CP,
667 multiple cross polarization pulse sequence; ORF, open reading frame; ssNMR,
668 solid-state nuclear magnetic resonance; TLC, thin layer chromatography.
669

670 ## Declarations

671 **Ethics approval and consent to participate**
672 Not applicable.
673

Consent for publication

Not applicable.

Availability of data and materials

The protein structure supporting the conclusions of this article is available in the Protein Data Bank repository, 5VQJ. The underlying short sequencing reads have been deposited at NCBI, Bioproject PRJNA398086. Assembled fragments were deposited to the Metagenomics RAST server, project ID MGP18213.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Grants 10/52362-5 and 15/13684-0 to Igor Polikarpov and 11/21608-1 to Bruno L. Mello) and the Biotechnology and Biological Sciences Research Council (BB/I018492/1 to Neil C. Bruce and Simon McQueen-Mason). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Authors' contributions

BLM and DMRP assembled the metatranscriptome; BLM, ERA and MCES performed the nuclear magnetic resonance experiment; BLM and FEGG performed the microscopy experiments; BLM and MCES performed the high performance liquid chromatography experiments; BLM conducted the other experiments; AMA and

BLM wrote the manuscript; IP, NCB and SMM supervised the study and provided valuable input. All authors critically reviewed the manuscript and approved the final version.

Acknowledgments

The authors would like to thank Embrapa for nuclear magnetic resonance spectrometer time.

References

1. Wang C, Dong D, Wang H, Müller K, Qin Y, Wu W. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. Biotechnol Biofuels. 2016;9:22.
2. Simmons CW, Reddy AP, D'haeseleer P, Khudyakov J, Billis K, Pati A, et al. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. Biotechnol Biofuels. 2014;7:495.
3. Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsornruang S, Champreda V. Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. Biotechnol Biofuels. 2015;8:16.
4. Jiang Y, Xiong X, Danska J, Parkinson J. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. Microbiome. 2016;4:2.
5. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol. 2013;79:1545-54.
6. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. 2011;331:463-7.
7. Martins LF, Antunes LP, Pascon RC, de Oliveira JC, Digiampietri LA, Barbosa D, et al. Metagenomic analysis of a tropical composting operation at the são paulo zoo park reveals diversity of biomass degradation functions and organisms. PLoS One. 2013;8:e61928.
8. 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:e8812.
9. Evans R, Alessi AM, Bird S, McQueen-Mason SJ, Bruce NC, Brockhurst MA. Defining the functional traits that drive bacterial decomposer community productivity. ISME J. 2017, doi:10.1038/ismej.2017.22.
10. Brulc JM, Antonopoulos DA, 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 U S A. 2009;106:1948-53.
11. 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.
12. Zhu L, Wu Q, Dai J, Zhang S, Wei F. Evidence of cellulose metabolism by the giant panda gut microbiome. Proc Natl Acad Sci U S A. 2011;108:17714-9.
13. Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 2012;109:11002-7.
14. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA, et al. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. Proc Natl Acad Sci U S A. 2010;107:14793-8.
15. Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB. Comparative fecal metagenomics unveils unique functional capacity of the swine gut. BMC Microbiol. 2011;11:103.

753 16. Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG, Scharf ME. Parallel
754 metatranscriptome analyses of host and symbiont gene expression in the gut of
755 the termite Reticulitermes flavipes. Biotechnol Biofuels. 2009;2:25.
756 17. Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous
757 assessment of soil microbial community structure and function through analysis
758 of the meta-transcriptome. PLoS One. 2008;3:e2527.
759 18. Hollister EB, Forrest AK, Wilkinson HH, Ebbole DJ, Tringe SG, Malfatti SA,
760 et al. Mesophilic and thermophilic conditions select for unique but highly parallel
761 microbial communities to perform carboxylate platform biomass conversion.
762 PLoS One. 2012;7:e39689.
763 19. Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Oliveira
764 CST, et al. Metagenomics of the water column in the pristine upper course of the
765 Amazon river. PLoS One. 2011;6:e23785.
766 20. Kumar R, Wyman CE. Effect of xylanase supplementation of cellulase on
767 digestion of corn stover solids prepared by leading pretreatment technologies.
768 Bioresour Technol. 2009;100:4203-13.
769 21. Johnson RL, Schmidt-Rohr K. Quantitative solid-state C-13 NMR with
770 signal enhancement by multiple cross polarization. J Magn Reson. 2014;239:44-
771 9.
772 22. Bernardinelli OD, Lima MA, Rezende CA, Polikarpov I, deAzevedo ER.
773 Quantitative C-13 MultiCP solid-state NMR as a tool for evaluation of cellulose
774 crystallinity index measured directly inside sugarcane biomass. Biotechnol
775 Biofuels. 2015;8:110.
776 23. Wickholm K, Larsson PT, Iversen T. Assignment of non-crystalline forms
777 in cellulose I by CP/MAS 13C NMR spectroscopy. Carbohydrate Research.
778 1998;312:123-9.
779 24. Templeton DW, Scarlata CJ, Sluiter JB, Wolfrum EJ. Compositional analysis
780 of lignocellulosic feedstocks. 2. Method uncertainties. J Agric Food Chem.
781 2010;58:9054-62.
782 25. Rezende CA, de Lima MA, Maziero P, Deazevedo ER, Garcia W, Polikarpov
783 I. Chemical and morphological characterization of sugarcane bagasse submitted
784 to a delignification process for enhanced enzymatic digestibility. Biotechnol
785 Biofuels. 2011;4:54.
786 26. Focher B, Marzetti A, Cattaneo M, Beltrame PL, Carniti P. Effects of
787 structural features of cotton cellulose on enzymatic hydrolysis. Journal of
788 Applied Polymer Science. 1981;26:1989-99.
789 27. Hallac BB, Sannigrahi P, Pu Y, Ray M, Murphy RJ, Ragauskas AJ. Biomass
790 characterization of Buddleja davidii: a potential feedstock for biofuel production.
791 J Agric Food Chem. 2009;57:1275-81.
792 28. El Hage R, Brosse N, Sannigrahi P, Ragauskas A. Effects of process severity
793 on the chemical structure of Miscanthus ethanol organosolv lignin. Polymer
794 Degradation and Stability. 2010;95:997-1003.
795 29. Sannigrahi P, Miller SJ, Ragauskas AJ. Effects of organosolv pretreatment
796 and enzymatic hydrolysis on cellulose structure and crystallinity in Loblolly pine.
797 Carbohydr Res. 2010;345:965-70.
798 30. Foston MB, Hubbell CA, Ragauskas AJ. Cellulose Isolation Methodology for
799 NMR Analysis of Cellulose Ultrastructure. Materials. 2011;4:1985-2002.
800 31. Martínez AT, González AE, Valmaseda M, Dale BE, Lambregts MJ, Haw JF.
801 Solid-State NMR Studies of Lignin and Plant Polysaccharide Degradation by

802 Fungi. Holzforschung - International Journal of the Biology, Chemistry, Physics
803 and Technology of Wood. 1991;45:49-54.
804 32. Lima MA, Gomez LD, Steele-King CG, Simister R, Bernardinelli OD,
805 Carvalho MA, et al. Evaluating the composition and processing potential of novel
806 sources of Brazilian biomass for sustainable biorenewables production.
807 Biotechnol Biofuels. 2014;7:10.
808 33. Coletta VC, Rezende CA, Conceição FR, Polikarpov I, Guimarães FE.
809 Mapping the lignin distribution in pretreated sugarcane bagasse by confocal and
810 fluorescence lifetime imaging microscopy. Biotechnol Biofuels. 2013;6: 43.
811 34. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence
812 similarity searching. Nucleic Acids Research. 2011;39:W29-37.
813 35. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for
814 automated carbohydrate-active enzyme annotation. Nucleic Acids Research.
815 2012;40:W445-51.
816 36. McCleary BV. Measurement of polysaccharide degrading enzymes using
817 chromogenic and colorimetric substrates. Chemistry in Australia. 1991;58:398-
818 401.
819 37. Watanabe M, Inoue H, Inoue B, Yoshimi M, Fujii T, Ishikawa K. Xylanase
820 (GH11) from Acremonium cellulolyticus: homologous expression and
821 characterization. AMB Express. 2014;4:27.
822 38. Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB, Selinger LB,
823 Butler G, Tsang A, McAllister TA, Forster RJ. Snapshot of the eukaryotic gene
824 expression in muskoxen rumen--a metatranscriptomic approach. PLoS One.
825 2011;6:e20521.
826 39. Dai X, Tian Y, Li J, Luo Y, Liu D, Zheng H, et al. Metatranscriptomic
827 analyses of plant cell wall polysaccharide degradation by microorganisms in the
828 cow rumen. Appl Environ Microbiol. 2015;81:1375-86.
829 40. Jiménez DJ, Chaves-Moreno D, van Elsas JD. Unveiling the metabolic
830 potential of two soil-derived microbial consortia selected on wheat straw. Sci
831 Rep. 2015;5:13845.
832 41. Alessi AM, Bird SM, Bennett JP, Oates NC, Li Y, Dowle AA, et al. Revealing
833 the insoluble metasecretome of lignocellulose-degrading microbial communities.
834 Sci Rep. 2017;7:2356.
835 42. Heiss-Blanquet S, Fayolle-Guichard F, Lombard V, Hébert A, Coutinho PM,
836 Groppi A, et al. Composting-Like Conditions Are More Efficient for Enrichment
837 and Diversity of Organisms Containing Cellulase-Encoding Genes than
838 Submerged Cultures. PLoS One. 2016;11:e0167216.
839 43. Mello BL, Alessi AM, McQueen-Mason S, Bruce NC, Polikarpov I. Nutrient
840 availability shapes the microbial community structure in sugarcane bagasse
841 compost-derived consortia. Scientific Reports. 2016;6:38781.
842 44. Yu K, Zhang T. Metagenomic and metatranscriptomic analysis of microbial
843 community structure and gene expression of activated sludge. PLoS One.
844 2012;7:e38183.
845 45. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH. Diversity of
846 beetle genes encoding novel plant cell wall degrading enzymes. PLoS One.
847 2010;5:e15635.
848 46. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH,
849 et al. The ecoresponsive genome of Daphnia pulex. Science. 2011;331:555-61.

47. Danchin EG, Rosso MN, Vieira P, Almeida-Engler J, Coutinho PM, Henrissat B, et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Natl Acad Sci USA. 2010;107:17651-6.
48. Kern M, McGeehan JE, Streeter SD, Martin RN, Besser K, Elias L, et al. Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. Proc Natl Acad Sci USA. 2013;110:10189-94.
49. Camilo CM, Polikarpov I. High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). Protein Expr Purif. 2014;99:35-42.
50. Bacic A, Fincher GB, Stone BA. Chemistry, biochemistry, and biology of (1-3)-beta-glucans and related polysaccharides. 1st ed. New York: Elsevier; 2009.
51. Wood PJ, Weisz J, Blackwell BA. Structural studies of (1- 3),(1- 4)-beta-d-glucans by c(13)-nuclear magnetic-resonance spectroscopy and by rapid analysis of cellulose-like regions using high-performance anion-exchange chromatography of oligosaccharides released by lichenase. Cereal Chemistry. 1994;71:301-7.
52. Sarethy IP, Saxena Y, Kapoor A, Sharma M, Sharma SK, Gupta V, et al. Alkaliphilic bacteria: applications in industrial biotechnology. J Ind Microbiol Biotechnol. 2011;38:769-90.
53. Knob A, Carmona EC. Purification and characterization of two extracellular xylanases from Penicillium sclerotiorum: a novel acidophilic xylanase. Appl Biochem Biotechnol. 2010;162:429-43.
54. Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. Appl Microbiol Biotechnol. 2011;90:1933-42.
55. Amel BD, Nawel B, Khelifa B, Mohammed G, Manon J, Salima KG, et al. Characterization of a purified thermostable xylanase from Caldicoprobacter algeriensis sp. nov. strain TH7C1(T). Carbohydr Res. 2016;419:60-8.
56. Kataoka M, Akita F, Maeno Y, Inoue B, Inoue H, Ishikawa K. Crystal structure of Talaromyces cellulolyticus (formerly known as Acremonium cellulolyticus) GH family 11 xylanase. Appl Biochem Biotechnol. 2014;174:1599-612.
57. Jänis J, Hakanpää J, Hakulinen N, Ibatullin FM, Hoxha A, Derrick PJ, et al. Determination of thioxylo-oligosaccharide binding to family 11 xylanases using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and X-ray crystallography. FEBS J. 2005;272:2317-33.
58. Vardakou M, Dumon C, Murray JW, Christakopoulos P, Weiner DP, Juge N, et al. Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases. J Mol Biol. 2008;375:1293-305.
59. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M. Mutational and crystallographic analyses of the active site residues of the Bacillus circulans xylanase. Protein Sci. 1994;3:467-75.
60. Paës G, Berrin JG, Beaugrand J. GH11 xylanases: Structure/function/properties relationships and applications. Biotechnol Adv. 2012;30:564-92.
61. Valenzuela SV, Lopez S, Biely P, Sanz-Aparicio J, Pastor FI. The Glycoside Hydrolase Family 8 Reducing-End Xylose-Releasing Exo-oligoxylanase Rex8A

898 from Paenibacillus barcinonensis BP-23 Is Active on Branched
899 Xylooligosaccharides. Appl Environ Microbiol. 2016;82:5116-24.
900 62. Juturu V, Wu JC. Microbial xylanases: engineering, production and
901 industrial applications. Biotechnol Adv. 2012;30:1219-27.
902 63. Qing Q, Yang B, Wyman CE. Xylooligomers are strong inhibitors of
903 cellulose hydrolysis by enzymes. Bioresour Technol. 2010;101:9624-30.
904 64. Bennet JW, Lasure LL. Growth Media. In: Bennett JW, Lasure LL, editors.
905 More gene manipulation in fungi. New York: Elsevier; 1991. p. 441-57.
906 65. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. Rapid method for
907 coextraction of DNA and RNA from natural environments for analysis of
908 ribosomal DNA- and rRNA-based microbial community composition. Applied and
909 Environmental Microbiology. 2000;66:5488-91.
910 66. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA,
911 Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions
912 of sequences per sample. Proc Natl Acad Sci U S A. 2011;108:4516-22.
913 67. Fierer N, Jackson JA, Vilgalys R, Jackson RB. Assessment of soil microbial
914 community structure by use of taxon-specific quantitative PCR assays. Appl
915 Environ Microbiol. 2005;71:4117-20.
916 68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for
917 Illumina sequence data. Bioinformatics. 2014;30:2114-20.
918 69. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of
919 ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211-7.
920 70. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA:
921 a comprehensive online resource for quality checked and aligned ribosomal RNA
922 sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188-96.
923 71. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, et al.
924 Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res.
925 2011;39:D141-5.
926 72. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al.
927 Full-length transcriptome assembly from RNA-Seq data without a reference
928 genome. Nature Biotechnology. 2011;29:644-52.
929 73. Ismail WM, Ye Y, Tang H. Gene finding in metatranscriptomic sequences.
930 BMC Bioinformatics. 2014;15 Suppl 9:S8.
931 74. Roberts A, Pachter L. Streaming fragment assignment for real-time
932 analysis of sequencing experiments. Nat Methods. 2013;10:71-3.
933 75. Langmead B, Salzberg SL.Fast gapped-read alignment with Bowtie 2. Nat
934 Methods. 2012;9:357-9.
935 76. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic
936 data. Genome Res. 2007;17:377-86.
937 77. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with
938 high-throughput sequencing data. Bioinformatics. 2015;31:166-9.
939 78. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes.
940 Nucleic Acids Res. 2000;28:27-30.
941 79. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG
942 Tools for Functional Characterization of Genome and Metagenome Sequences. J
943 Mol Biol. 2016;428:726-31.
944 80. Aslanidis C, Dejong PJ. Ligation-independent cloning of PCR products
945 (LIC-PCR). Nucleic Acids Research. 1990;18:6069-74.

81. Michael RG, Joseph S. Molecular cloning: a laboratory manual. 4th ed. New York: Cold Spring Harbor; 2012.
82. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227:680-5.
83. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.
84. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 1959;31:426-8.
85. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D. In: Determination of Sugars, Byproducts, and Degradation Products in Liquid Fraction Process Samples. NREL/TP-510-42623. 2008. http://www.nrel.gov/docs/gen/fy08/42623.pdf. Accessed 28 Apr 2016.
86. Selig M, Weiss N, Ji Y. In: Enzymatic Saccharification of Lignocellulosic Biomass. NREL/TP-510-42629. 2008. http://www.nrel.gov/docs/gen/fy08/42629.pdf. Accessed 28 Apr 2016.
87. Guimarães BG, Sanfelici L, Neuenschwander RT, Rodrigues F, Grizolli WC, Raulik MA, et al. The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. J Synchrotron Radiat. 2009;16:69-75.
88. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66:125-32.
89. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658-74.
90. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66:213-21.
91. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60:2126-32.
92. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66:12-21.
93. R. Berlemont, Allison SD, Weihe C, Lu Y, Brodie EL, Martiny JBH, *et al.* Cellulolytic potential under environmental changes in microbial communities from grassland litter. Front Microbiol. 2014;5:639.
94. Wang Y, Ghaffari N, Johnson CD, Braga-Neto UM, Wang H, Chen R, et al. Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. BMC Bioinformatics. 2011;12 Suppl 10:S5.

Figure Legends

Fig. 1

Analysis of sugarcane bagasse composition. **a** Solid state nuclear magnetic resonance spectra of sugarcane bagasse prior and after microbial growth. Regions 1 and 1´ are assigned to the C1 carbon of cellulose (103-106 ppm) and to C4 carbon of crystalline cellulose (86-92 ppm). Regions 2 and 2´ are assigned to C1, C2 and C4 aromatic carbons of lignin (123-142 ppm) and to aryl methoxyl carbons of lignin (50-56 ppm). Regions 3 and 3´ are assigned to C1 carbon of hemicellulose and to CH_3 in acetyl groups of hemicelluloses. The relative abundance of cellulose, hemicellulose and lignin was estimated from regions 1, 1', 2, 2', 3 and 3'. **b-e** Confocal images of sugarcane bagasse lignin prior and after microbial growth. **f** Corresponding average emission spectra.

Fig. 2

Scanning electron microscopy imaging of the sugarcane bagasse prior **a-b** and after 5-weeks **c-d** of microbial growth analyzed in 2x magnifications.

Fig. 3

Functional profile of KEGG-assigned genes in sugarcane bagasse composting community metatranscriptome. Relative abundance of predicted open reading frames in terms of the KEGG function was assigned. Value for a functional profiles are normalized by the sum of all functions for each time point.

Fig. 4

Expression of putative CAZymes in the microbial community metatranscriptome. AA: auxiliary activities, CE: carbohydrate esterases, CBM: carbohydrate binding modules, GH: glycoside hydrolases, GT: glycosyl transferases, PL: polysaccharide lyases.

Fig. 5

Differential expression and phylogenetic distribution of glycoside hydrolase (GH) families identified in sugarcane bagasse composting community metatranscriptome. **a** Heat map representation of the GHs expression. Columns represent time when sample was collected. Rows depict different GHs families identified in the metatranscriptome dataset. The color key for GH 3 expression at week 1 is out of range with expression level of 2494. GH families were grouped based on substrate preferences, as reported elsewhere [*93*]. **b** Phylogenetic assignment of reads belonging to the most expressed families using the Lowest Common Ancestor algorithm.

Fig. 6

Experimental evidence that compost21_GH11 is an exo-1,4-β-xylanase. **a** Thin layer chromatography (TLC) of the products formed by: lanes 1 and 3: compost21_GH11 and compost13_GH10 action on xylan, respectively; lane 2: reaction blank; lanes 4 to 8: compost21_GH11 action on X2 to X6; GAc: glucuronic acid; M: X1-X6 standard. **b** HPAEC-PAD of the products formed by compost21_GH11 action on xylan.

Fig. 7

Structural evidence that compost21_GH11 is an exo-1,4-β-xylanase. **a** Amino acid sequence multiple alignment of compost21_GH11 with closest homologues selected

based on searches in the NCBI-nr and PDB databases. The residues responsible for formation of extra loops 1 and 2 (EL1 and EL2) are shown in a red box while the catalytic residues are marked in a blue box. The extra loops are present in other proteins for which structure has not been solved. **b** Overall secondary structure of compost21_GH11 colored from blue to red (N- to C-terminal). **c** Superposition of all available GH11 structures (in gray) with compost21_GH11 (in magenta). The orientation of the structures are rotated by 90° in relation to the structure presented on item B. The non-aligned domains from proteins with non-common domains were hidden from representation. **d** Surface representation of compost21_GH11 (in magenta) aligned to closest PDB homologue 1XNK (in gray). Note that the non-reducing end of the ligand methyl 4,4^{II}-dithio-α-xylotrioside present in 1XNK structure is sterically hindered by EL2 in compost21_GH11 -3 subsite.

Fig. 8

Impact of xylooligomers on commercial cocktail performance. **a** Hydrolysis of Avicel by Accellerase® in presence of xylooligomers and compost21_GH11 during 96h time course. **b** HPLC analysis showing xylooligomer degradation by Accellerase® and **c** Accellerase® supplemented with compost21_GH11, as assessed by detector response in nanocoulombs (nC). Note that in Fig. 3c xylobiose concentration increases after 1 hour as a result of fast degradation of xylooligomers with degree of polymerization (DP) ≥4 by compost21_GH11. Higher DP xylooligomers have a higher inhibitory effect, thus its rapid degradation seems beneficial. The commercial cocktail used was able to degrade the added xylooligomers within 24h. In consequence, after 96h, reactions with and without compost21_GH11 achieved about the same glucose yields.

However, at initial stages, addition of compost21_GH11 strongly improved glucose production rate.

Tables

Table 1

	Week 1	Week 2	Week 3	Week 4	Week 5	Total
Total reads generated	20,119,184	14,715,430	7,767,801	5,879,612	17,825,076	66,307,103
Reads after quality filtering	15,916,667	12,072,389	6,757,570	5,111,343	14,535,407	54,393,376
mRNA reads	12,645,050	10,652,024	3,608,818	3,971,178	11,004,487	41,881,555
Assembled transcripts	104,408	64,923	27,821	34,023	71,786	302,961
N50	947	888	789	817	858	912
average size	963	921	820	847	884	887
Predicted open reading frames	104,425	59,885	23,156	31,198	64,692	283,356
Lignocellulose active	3,012	1,523	534	654	1,473	7,196

Metatranscriptome sequencing and processing metrics.

Table 2

Data collection	
Wavelength/beamline	1.45866/MX2, LNLS
Space group	$P4_32_12$
Unit cell dimensions (Å)	64.33; 64.33; 105.87
Molecules/asymmetry unit	1
Matthews coefficient ($Å^3$/Da)	2.24
Solvent content (%)	45.0
Resolution (Å)	1.76
Number of unique reflections	22,638 (1,226)
Mosaicity (°)	0.205
Multiplicity	24.0
Completeness	99.6
Refinement	
Number of amino acid residues	216
Number of waters	282
R_{work}/R_{free} (%)	19.4/21,8
RMS bond lengths (Å)	0.069
RMSD bond angles (°)	1.144
Mean overall B-factor ($Å^2$)	19.5
Ramachandran in most favored regions (%)	96.73
Ramachandran outliers (%)	0
PDB ID	5VQJ

1090

1091 Data collection and refinement statistics of compost21_GH11 structure. Values in
1092 parenthesis refer to the outer shell. R_{free} was calculated with 5% of the reflections that
1093 were randomly chosen and excluded from the refinement.

1094