

This is a repository copy of *Targeted metatranscriptomics of compost derived consortia reveals a GH11 exerting an unusual exo-1,4- β -xylanase activity.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/126688/>

Version: Accepted Version

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>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

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

3

4 Bruno L. Mello¹, Anna M. Alessi², Diego M. Riaño-Pachón^{3,4}, Eduardo R.
5 deAzevedo¹, Francisco E. G. Guimarães¹, Melissa C. Espirito Santo¹, Simon
6 McQueen-Mason², Neil C. Bruce², Igor Polikarpov¹

7

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

12 ²Department of Biology, University of York, Wentworth Way, York, UK, YO10
13 5DD, emails: anna.szczepanska@york.ac.uk; simon.mcqueenmason@york.ac.uk;
14 neil.bruce@york.ac.uk

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

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

21

22 **Corresponding author** Igor Polikarpov, Instituto de Física de São Carlos,
23 Universidade de São Paulo, Av. Trabalhador São-carlense 400, São Carlos, SP,
24 13560-970, Brazil

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

26 **Abstract**

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

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

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

50

51

52 **Keywords**

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

55

56 **Background**

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

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

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

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

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

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

93

94 **Results**

95 **Compositional and morphological changes in sugarcane bagasse**

96 The compositional analysis of sugarcane bagasse collected weekly from in
97 vitro composting cultures was investigated using ssNMR. In order to obtain ^{13}C
98 quantitative spectra in an achievable measuring time, the spectra were acquired using
99 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

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

128

129 **Functional and phylogenetic characteristics of the sugarcane bagasse degrading** 130 **microbial community**

131 To examine the transcriptional responses of the sugarcane bagasse degrading
132 microbial community, we performed RNA-seq metatranscriptomic analysis on weekly
133 cultures grown for up to five weeks. Although it yielded 66 million paired-end reads
134 (Table 1), the rarefaction analysis showed that the sequencing did not reach saturation
135 (Fig. S1). The rarefaction analysis also revealed that week 1 cultures were more
136 diverse than week 5 cultures since a larger proportion of new reads was obtained at
137 the same sequencing depth. Resulting high quality, non-ribosomal RNA reads (63%)
138 were de novo assembled into 302,961 transcripts and used to predict biochemical
139 capabilities of the microbial community by mapping putative transcripts to the KEGG
140 orthology. The distribution of genes classified to KEGG functions presented a similar
141 profile for all time points (Fig. 3). Transcripts assigned to translation processes (mean
142 = 8.5%) showed the highest relative abundance, followed by genes involved in
143 energy, carbohydrate and amino acids metabolism and signal transduction (mean from
144 6 to 8%). We observed that the microbial community was more actively growing and
145 breaking down the lignocellulose in initial stages of culture since a number of
146 transcripts assigned to energy and carbohydrate metabolism were higher in week 1
147 compared to week 5 cultures.

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

150 involved in lignocellulose degradation. From the 283,356 predicted open reading
151 frames (ORFs), 2.5% (number of sequences = 7,196) showed homology to CAZymes.
152 The CAZymes were distributed between carbohydrate binding modules (34.4%),
153 glycoside hydrolases (34.3%), glycosyl transferases (19.3%), carbohydrate esterases
154 (15.9%), auxiliary activities (3.3%), polysaccharide lyases (2.2%) and cohesin and
155 dockerin modules (0.8%). The expression level for the majority of CAZyme classes
156 was higher in week 1 cultures (Fig. 4). KEGG assigned transcripts involved in
157 carbohydrate metabolism presented a similar profile. Transcripts predicted as glycosyl
158 transferases showed contrasting behavior with the highest expression level at week 5.

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

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

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

198

199 **Heterologous expression and characterization of putative CAZymes**

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

207 BLASTP results against the NCBI-nr database showed that the protein
208 compost7_GH6 has 49% identity to a GH6 β -1,4-glucan cellobiohydrolase from
209 *Sorangium cellulosum*. Substrate screening showed that compost7_GH6 had activity
210 against β -glucan and lichenan. No enzymatic activity was detected towards filter
211 paper and carboxymethyl cellulose (CMC). compost7_GH6 displayed highest activity
212 towards β -glucan at pH 10.0 and was able to retain 70% or more activity until the pH
213 dropped to 4.0 (Fig. S3). Thermal shift assays (see method, Fig. S4) confirmed the
214 alkaliphilic behavior of the enzyme with the highest thermostability at pH 6 to 9. The
215 optimal temperature was assayed at both pH 6.0, where the enzyme was found to
216 maintain >70% maximum activity, and 10.0. The highest activity at pH 6.0 and 10.0
217 was found at 50 and 45 °C, respectively. Next, the optimal pH and temperature were
218 used to test enzyme specificity. The results showed that compost7_GH6 had highest
219 specific activity towards β -glucan (2.0 U/mg) and lichenan (1.5 U/mg) amongst the
220 substrates tested (Fig. S3c). We also determined that compost7_GH6 maintained 50%
221 of its initial activity after 24 h incubation at 45 °C at pH 6.0 and 100% activity after
222 96 h incubation at 40 °C at pH 10.0.

223 The protein C13 was predicted as a member of GH10 family and had 91% and
224 89% identity to an endoglucanase and endo- β -1,4-xylanase from *S. cellulosum*,

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

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

246

247 **Characterization and structure of compost21_GH11**

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

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

258 In order to understand the molecular basis of *exo*-1,4- β -xylanase activity, the
259 crystal structure of compost21_GH11 was solved at 1.76 Å resolution. The data
260 collection and refinement statistics are summarized in Table 2. The final model was
261 constructed from the first to last residue of the crystallized protein, which was cloned
262 without the first 13 residues that were predicted as a disordered region and reported as
263 a signal peptide [37]. Despite its low identity to the closest homologue in PDB (ID
264 1XNK, 36% identity), compost21_GH11 presents a typical GH11 fold. It consists of
265 one α -helix and 15 β -sheets, labeled A1 to A6 and B1 to B9 (Fig. 7a). The curvature
266 of the β -sheets B1 to B9 forms a cleft where the catalytic residues (nucleophile Glu98
267 in β -sheet B6 and proton donor Glu200 in β -sheet B4) are located. Two extra loops
268 (EL1 and EL2) are found in the compost21_GH11 structure. They are created by
269 additional residues, which stretch between β -sheets B5 and B6 and in β -sheet B4,
270 respectively. To gain insight into substrate binding, 1XNK, which was co-crystallized
271 with a modified xylotrioside, was aligned to compost21_GH11 (root mean square
272 deviation of 0.898 Å when EL1 and 2 are ignored). EL2 blocks substrate interaction
273 at subsite -3, which accommodates the non-reducing end of xylan (Fig. 7d). This loop
274 is stabilized by a number of hydrogen bonds within the loop main/side chain residues

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

279

280 **compost21_GH11 activity improves performance of a commercial enzyme** 281 **mixture**

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

300

301 **Discussion**

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

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

314 Next, we explored the metatranscriptome-assembled library, by focusing on
315 screening the resulting database for CAZymes. Although the predicted GHs
316 accounted for a small fraction (0.87%) in our composting community
317 metatranscriptome, this was similarly observed by others who investigated various
318 lignocellulytic communities such as rice-straw enriched compost (0.97%) [1], soil-
319 contacting sugarcane bagasse (0.97%) [3], termite lumen (0.78%) [11], bovine rumen
320 (0.78%) [10] and macropod foregut (0.71%) [14]. Amongst GHs, oligosaccharide-
321 degrading enzymes from GH3 family were highly expressed in our study. These
322 enzymes are fundamental in lignocellulolytic processes [5] and were abundant in
323 other lignocellulolytic environments [40]. Mhuantong et al. explored the metagenome
324 of a microbial community extracted from soil-contacting sugarcane bagasse [3]. Six

325 out of the 10 most abundant GH families in the reported metagenome are amongst the
326 most expressed GHs in each week of our metatranscriptome. Therefore, despite the
327 different environments and techniques used, these communities have a reasonable
328 level of similarity. Enzymes from auxiliary activity families, attributed to lignin
329 oxidative modification and lytic polysaccharides degradation, accounted for a very
330 small fraction (3.3%) of all predicted CAZymes in our metatranscriptome, supporting
331 the lack of sugarcane bagasse lignin removal or structural changes in this polymer.
332 This could be associated with a low relative abundance of fungi in our composting
333 cultures, especially in early stages of the time course. Experimental design that uses
334 liquid culturing of compost inoculum could have an effect on fungal growth and
335 hence ligninolytic enzymes expression [41]. Recent studies showed that the
336 composting conditions without liquid phase were preferable for CAZymes enrichment
337 [42]. Other factors such as medium composition [43], temperature, agitation and
338 inoculum source could also play critical role for suppressing fungal growth.

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

345 Interestingly, in the later stages of composting process, CAZymes expression
346 shifted towards Eukaryotes and Animal kingdom. Representatives of nematodes,
347 protists and other groups will be present in a composting spot, but it is unlikely that
348 they would survive weeks in the submerged cultures. One explanation can be that the
349 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.

400

401 **Methods**

402 **Sample collection and culture**

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

413 **Biomass analysis**

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

418 NMR experiments were performed using a Bruker Avance 400 spectrometer,
419 equipped with a Bruker 4-mm magical angle spinning double-resonance probe, at ¹³C
420 and ¹H frequencies of 100.5 and 400.0 MHz, respectively. The spinning frequency at
421 14 kHz was controlled by a pneumatic system that ensures a rotation stability higher
422 than ~1 Hz. Typical $\pi/2$ pulse lengths of 4 and 3.5 μ s were applied for ¹³C and ¹H,
423 respectively. Proton decoupling field strength of $\gamma B_1/2\pi = 100$ kHz was used. ¹³C
424 quantitative spectra were measured by using the Multi-CP excitation method

425 described by Johnson and Schmidt-Rohr [21]. A total of nine cross-polarization
426 blocks were implemented with 1 ms and RF amplitude increment (90-100%), while
427 the cross-polarization before acquisition was executed with 0.8 ms and the same
428 amplitude increment. The recycle delay was 2 s and the duration of the repolarization
429 period t_z was 0.9 s [22]. To obtain the fraction of cellulose, hemicellulose and lignin,
430 the spectra were normalized with respect to their area and integrated over the specific
431 regions for cellulose (1 and 1'), lignin (2 and 2') and hemicellulose (3 and 3'). The
432 relative percentage of each component was obtained by dividing the calculated value
433 for the biomass after microbial growth with the value obtained in the control sample.
434 Finally, this fraction was multiplied by the initial percentage of the component in the
435 raw sugarcane bagasse, as reported by Lima *et al.* [32]. Chemical shifts were assigned
436 based on published studies [23-31]. Approximately 4,000 scans were measured to
437 acquire each spectrum. Chemical shift was assigned based on published studies [23-
438 31].

439 Confocal microscopy was performed using a Zeiss LSM 780 confocal inverted
440 microscope with a Coherent Chameleon laser (Ti:sapphire) as source for two-photons
441 (2P) excitation at 800 nm. Ground sugarcane bagasse was hydrated for 24 h and
442 observed with a C-Apochromat objective lens (20X, numerical aperture 0.8); the
443 images were taken in the opposite side of the cover slip. The images were obtained by
444 the average of 2 scans and no appreciate variation was observed. The spatial
445 resolution was approximately 350 nm (considering the numerical aperture and the
446 wavelength of excitation).

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

450 a drop was directly applied to the sample pedestal and dried at room temperature for
451 12 h. After drying, samples were gold coated using a metalizer model MED 020 (Bal-
452 tec, Liechtenstein). Images were obtained under vacuum. At least 10 images per
453 sample were acquired from different areas to certify the reproducibility of the results.

454

455 **Nucleic acid extraction from sugarcane composting cultures**

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

472 All solutions and glassware were treated with 0.1% DEPC overnight at 37 °C
473 under homogenization and autoclaved to create and RNase-free environment. Only
474 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

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

506

507 **Metatranscriptomics assembly and annotation**

508 Sequenced reads were preprocessed with Trimmomatic [68] to remove
509 adaptors, low quality and short sequences. SortMeRNA [69] was used to merge and
510 remove contaminant ribosomal RNA sequences, which were identified using Silva
511 [70] and Rfam [71] reference databases with an e-value cutoff of 1×10^{-5} . Non-
512 ribosomal RNA reads were de novo assembled with Trinity [72] and genes were
513 predicted using TransGeneScan [73]. Expression levels were calculated with eXpress
514 [74] and Bowtie2 [75]. The phylogenetic origin of predicted ORFs was analysed
515 using MEGAN v6 [76] and the Lowest Common Ancestor algorithm at default
516 values. HTSeq [77] software was used to subsample the sequenced reads. Rarefaction
517 curves were plotted using the software BBMap (sourceforge.net/projects/bbmap/) for
518 sequencing depth analysis. Functional annotation was performed with HMMER
519 alignment tool [34] against the dbCAN database [35]. Kyoto Encyclopedia of Genes
520 and the Genomes (KEGG) [78] orthology classification was performed using the
521 online tool GhostKOALA [79] and default values.

522

523 **Target genes cloning, expression and purification**

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

535 Cells were grown overnight in the LB medium in presence of kanamycin (50
536 µg/mL) and chloramphenicol (34 µg/mL) at 150 rpm shaking at 37 °C. 1 L of LB
537 medium was inoculated with overnight culture (1% v/v) and incubated under 150 rpm
538 shaking at 37 °C until the optical density at 600 nm reached 0.8. Incubation
539 temperature was decreased to 17 °C and expression was carried out for 16 hours after
540 induction with 0.5 mM IPTG. The cells were harvested at 9,000 x g for 20 minutes
541 and resulting pellet was resuspended in 20 mL of lysis buffer (20 mM Tris-HCl, 300
542 mM NaCl, 5 mM imidazole, 5% (v/v) glycerol, 10 mM β-mercaptoethanol (β-ME), 1
543 mM phenylmethylsulfonylfluoride (PMSF), 0.2 mg/mL lysozyme, pH 8.0) with or
544 without 0.6% (w/v) sarkosyl. Cells suspension was incubated on ice for 1 hour,
545 sonicated for 6 minutes and centrifuged at 23,000 x g for 30 minutes. Supernatant was
546 loaded on a column with 2 mL of nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen,
547 Hilden, Germany) previously equilibrated with 10 volumes of lysis buffer. The
548 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 μ L 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 $^{\circ}$ C,
591 with stepwise increments of 1 $^{\circ}$ 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 CH₃COONa, [segment 2] 20-24' up to
608 100 mM CH₃COONa, [segment 3] 24-30' isocratic 100 mM CH₃COONa.

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

624 concentration. The program used was: flow rate 0.6 mL/min, 65 °C, isocratic 5 mM
625 H₂SO₄.

626

627 **Xylooligosaccharide hydrolysis inhibition**

628 2% (w/v) Avicel hydrolysis by Accellerase 1500 (DuPont, Wilmington, DE,
629 USA) was performed in 50 mM sodium citrate pH 5.0 to access xylooligosaccharides
630 inhibition (at 8 g/L) and activity improvement by compost21_GH11 addition (at 0.1
631 g/L). 0.02% (w/v) sodium azide was used to prevent microbial growth. Reactions
632 were started by addition of Accellerase 1500 (DuPont, Wilmington, DE, USA) and
633 Novozyme 188 (Sigma-Aldrich, St. Louis, MO, USA) diluted 25 and 250 times to
634 give 5 FPU/g and 10 CBU/g respectively. The flasks were incubated at 35 °C with
635 150 rpm agitation. Substrate blanks without enzyme and enzyme blanks without
636 substrate were also set. 0.5 mL aliquots taken after 1, 4, 7, 24, 48, 72 and 96 h of
637 hydrolysis were immediately boiled for 10 min to inactivate enzymes, centrifuged at
638 10,000 x g for 1 min, filtered through a 0.45 µm PVDF filter and frozen at -20 °C.
639 Two independent experiments were performed. The cellulose conversion yields were
640 analyzed with HPLC, as discussed before, using glucose standards. This experimental
641 set up was based on NREL standard protocol [86].

642

643 **Crystallization and data collection**

644 Crystallization conditions were screened for crystal growth using a HoneyBee
645 crystallization robot 931 (Genomic Solutions, Ann Arbor, MI, USA) and commercial
646 available screens. Crystals were obtained for protein compost21_GH11 in multiple
647 conditions. Crystals grown at 18 °C in 0.1 M BIS-TRIS propane pH 7.5, 20% (w/v)
648 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

674 **Consent for publication**

675 Not applicable.

676

677 **Availability of data and materials**

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

682

683 **Competing interests**

684 The authors declare that they have no competing interests.

685

686 **Funding**

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

693

694 **Authors' contributions**

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

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

702

703 **Acknowledgments**

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

706 **References**

- 707 1. Wang C, Dong D, Wang H, Müller K, Qin Y, Wu W. Metagenomic analysis of
708 microbial consortia enriched from compost: new insights into the role of
709 Actinobacteria in lignocellulose decomposition. *Biotechnol Biofuels*. 2016;9:22.
- 710 2. Simmons CW, Reddy AP, D'haeseleer P, Khudyakov J, Billis K, Pati A, et al.
711 Metatranscriptomic analysis of lignocellulolytic microbial communities involved
712 in high-solids decomposition of rice straw. *Biotechnol Biofuels*. 2014;7:495.
- 713 3. Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsornruang S,
714 Champreda V. Comparative analysis of sugarcane bagasse metagenome reveals
715 unique and conserved biomass-degrading enzymes among lignocellulolytic
716 microbial communities. *Biotechnol Biofuels*. 2015;8:16.
- 717 4. Jiang Y, Xiong X, Danska J, Parkinson J. Metatranscriptomic analysis of
718 diverse microbial communities reveals core metabolic pathways and
719 microbiome-specific functionality. *Microbiome*. 2016;4:2.
- 720 5. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases
721 in bacteria. *Appl Environ Microbiol*. 2013;79:1545-54.
- 722 6. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al.
723 Metagenomic discovery of biomass-degrading genes and genomes from cow
724 rumen. *Science*. 2011;331:463-7.
- 725 7. Martins LF, Antunes LP, Pascon RC, de Oliveira JC, Digiampietri LA,
726 Barbosa D, et al. Metagenomic analysis of a tropical composting operation at the
727 são paulo zoo park reveals diversity of biomass degradation functions and
728 organisms. *PLoS One*. 2013;8:e61928.
- 729 8. Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S, et al.
730 Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost
731 community. *PLoS One*. 2010;5:e8812.
- 732 9. Evans R, Alessi AM, Bird S, McQueen-Mason SJ, Bruce NC, Brockhurst MA.
733 Defining the functional traits that drive bacterial decomposer community
734 productivity. *ISME J*. 2017, doi:10.1038/ismej.2017.22.
- 735 10. Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale
736 EA, et al. Gene-centric metagenomics of the fiber-adherent bovine rumen
737 microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci U S*
738 *A*. 2009;106:1948-53.
- 739 11. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege
740 JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-
741 feeding higher termite. *Nature*. 2007;450:560-5.
- 742 12. Zhu L, Wu Q, Dai J, Zhang S, Wei F. Evidence of cellulose metabolism by
743 the giant panda gut microbiome. *Proc Natl Acad Sci U S A*. 2011;108:17714-9.
- 744 13. Engel P, Martinson VG, Moran NA. Functional diversity within the simple
745 gut microbiota of the honey bee. *Proc Natl Acad Sci U S A*. 2012;109:11002-7.
- 746 14. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA, et al.
747 Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside
748 hydrolase profiles different from other herbivores. *Proc Natl Acad Sci U S A*.
749 2010;107:14793-8.
- 750 15. Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB.
751 Comparative fecal metagenomics unveils unique functional capacity of the swine
752 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, McMahan 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.

- 850 47. Danchin EG, Rosso MN, Vieira P, Almeida-Engler J, Coutinho PM, Henrissat
851 B, et al. Multiple lateral gene transfers and duplications have promoted plant
852 parasitism ability in nematodes. *Proc Natl Acad Sci USA*. 2010;107:17651-6.
- 853 48. Kern M, McGeehan JE, Streeter SD, Martin RN, Besser K, Elias L, et al.
854 Structural characterization of a unique marine animal family 7 cellobiohydrolase
855 suggests a mechanism of cellulase salt tolerance. *Proc Natl Acad Sci USA*.
856 2013;110:10189-94.
- 857 49. Camilo CM, Polikarpov I. High-throughput cloning, expression and
858 purification of glycoside hydrolases using Ligation-Independent Cloning (LIC).
859 *Protein Expr Purif*. 2014;99:35-42.
- 860 50. Bacic A, Fincher GB, Stone BA. *Chemistry, biochemistry, and biology of (1-
861 3)-beta-glucans and related polysaccharides*. 1st ed. New York: Elsevier; 2009.
- 862 51. Wood PJ, Weisz J, Blackwell BA. Structural studies of (1-3),(1-4)-beta-d-
863 glucans by c(13)-nuclear magnetic-resonance spectroscopy and by rapid
864 analysis of cellulose-like regions using high-performance anion-exchange
865 chromatography of oligosaccharides released by lichenase. *Cereal Chemistry*.
866 1994;71:301-7.
- 867 52. Sarethy IP, Saxena Y, Kapoor A, Sharma M, Sharma SK, Gupta V, et al.
868 Alkaliphilic bacteria: applications in industrial biotechnology. *J Ind Microbiol
869 Biotechnol*. 2011;38:769-90.
- 870 53. Knob A, Carmona EC. Purification and characterization of two
871 extracellular xylanases from *Penicillium sclerotiorum*: a novel acidophilic
872 xylanase. *Appl Biochem Biotechnol*. 2010;162:429-43.
- 873 54. Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H. Characterization of a
874 bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl
875 Microbiol Biotechnol*. 2011;90:1933-42.
- 876 55. Amel BD, Nawel B, Khelifa B, Mohammed G, Manon J, Salima KG, et al.
877 Characterization of a purified thermostable xylanase from *Caldicoprobacter
878 algeriensis* sp. nov. strain TH7C1(T). *Carbohydr Res*. 2016;419:60-8.
- 879 56. Kataoka M, Akita F, Maeno Y, Inoue B, Inoue H, Ishikawa K. Crystal
880 structure of *Talaromyces cellulolyticus* (formerly known as *Acremonium
881 cellulolyticus*) GH family 11 xylanase. *Appl Biochem Biotechnol*. 2014;174:1599-
882 612.
- 883 57. Jänis J, Hakanpää J, Hakulinen N, Ibatullin FM, Hoxha A, Derrick PJ, et al.
884 Determination of thioxylo-oligosaccharide binding to family 11 xylanases using
885 electrospray ionization Fourier transform ion cyclotron resonance mass
886 spectrometry and X-ray crystallography. *FEBS J*. 2005;272:2317-33.
- 887 58. Vardakou M, Dumon C, Murray JW, Christakopoulos P, Weiner DP, Juge N,
888 et al. Understanding the structural basis for substrate and inhibitor recognition
889 in eukaryotic GH11 xylanases. *J Mol Biol*. 2008;375:1293-305.
- 890 59. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M. Mutational
891 and crystallographic analyses of the active site residues of the *Bacillus circulans*
892 xylanase. *Protein Sci*. 1994;3:467-75.
- 893 60. Paës G, Berrin JG, Beaugrand J. GH11 xylanases:
894 Structure/function/properties relationships and applications. *Biotechnol Adv*.
895 2012;30:564-92.
- 896 61. Valenzuela SV, Lopez S, Biely P, Sanz-Aparicio J, Pastor FI. The Glycoside
897 Hydrolase Family 8 Reducing-End Xylose-Releasing Exo-oligoxyylanase 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.

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

982

983

984

985

986

987

988

989 **Figure Legends**

990 **Fig. 1**

991 Analysis of sugarcane bagasse composition. **a** Solid state nuclear magnetic resonance
992 spectra of sugarcane bagasse prior and after microbial growth. Regions 1 and 1' are
993 assigned to the C1 carbon of cellulose (103-106 ppm) and to C4 carbon of crystalline
994 cellulose (86-92 ppm). Regions 2 and 2' are assigned to C1, C2 and C4 aromatic
995 carbons of lignin (123-142 ppm) and to aryl methoxyl carbons of lignin (50-56 ppm).
996 Regions 3 and 3' are assigned to C1 carbon of hemicellulose and to CH₃ in acetyl
997 groups of hemicelluloses. The relative abundance of cellulose, hemicellulose and
998 lignin was estimated from regions 1, 1', 2, 2', 3 and 3'. **b-e** Confocal images of
999 sugarcane bagasse lignin prior and after microbial growth. **f** Corresponding average
1000 emission spectra.

1001

1002 **Fig. 2**

1003 Scanning electron microscopy imaging of the sugarcane bagasse prior **a-b** and after 5-
1004 weeks **c-d** of microbial growth analyzed in 2x magnifications.

1005

1006 **Fig. 3**

1007 Functional profile of KEGG-assigned genes in sugarcane bagasse composting
1008 community metatranscriptome. Relative abundance of predicted open reading frames
1009 in terms of the KEGG function was assigned. Value for a functional profiles are
1010 normalized by the sum of all functions for each time point.

1011

1012 **Fig. 4**

1013 Expression of putative CAZymes in the microbial community metatranscriptome.
1014 AA: auxiliary activities, CE: carbohydrate esterases, CBM: carbohydrate binding
1015 modules, GH: glycoside hydrolases, GT: glycosyl transferases, PL: polysaccharide
1016 lyases.

1017

1018 **Fig. 5**

1019 Differential expression and phylogenetic distribution of glycoside hydrolase (GH)
1020 families identified in sugarcane bagasse composting community metatranscriptome. **a**
1021 Heat map representation of the GHs expression. Columns represent time when sample
1022 was collected. Rows depict different GHs families identified in the metatranscriptome
1023 dataset. The color key for GH 3 expression at week 1 is out of range with expression
1024 level of 2494. GH families were grouped based on substrate preferences, as reported
1025 elsewhere [93]. **b** Phylogenetic assignment of reads belonging to the most expressed
1026 families using the Lowest Common Ancestor algorithm.

1027

1028 **Fig. 6**

1029 Experimental evidence that compost21_GH11 is an exo-1,4- β -xylanase. **a** Thin layer
1030 chromatography (TLC) of the products formed by: lanes 1 and 3: compost21_GH11
1031 and compost13_GH10 action on xylan, respectively; lane 2: reaction blank; lanes 4 to
1032 8: compost21_GH11 action on X2 to X6; GAc: glucuronic acid; M: X1-X6 standard.
1033 **b** HPAEC-PAD of the products formed by compost21_GH11 action on xylan.

1034

1035 **Fig. 7**

1036 Structural evidence that compost21_GH11 is an exo-1,4- β -xylanase. **a** Amino acid
1037 sequence multiple alignment of compost21_GH11 with closest homologues selected

1038 based on searches in the NCBI-nr and PDB databases. The residues responsible for
1039 formation of extra loops 1 and 2 (EL1 and EL2) are shown in a red box while the
1040 catalytic residues are marked in a blue box. The extra loops are present in other
1041 proteins for which structure has not been solved. **b** Overall secondary structure of
1042 compost21_GH11 colored from blue to red (N- to C-terminal). **c** Superposition of all
1043 available GH11 structures (in gray) with compost21_GH11 (in magenta). The
1044 orientation of the structures are rotated by 90° in relation to the structure presented on
1045 item B. The non-aligned domains from proteins with non-common domains were
1046 hidden from representation. **d** Surface representation of compost21_GH11 (in
1047 magenta) aligned to closest PDB homologue 1XNK (in gray). Note that the non-
1048 reducing end of the ligand methyl 4,4^{II}-dithio- α -xylotrioside present in 1XNK
1049 structure is sterically hindered by EL2 in compost21_GH11 -3 subsite.

1050

1051 **Fig. 8**

1052 Impact of xylooligomers on commercial cocktail performance. **a** Hydrolysis of Avicel
1053 by Accellerase® in presence of xylooligomers and compost21_GH11 during 96h time
1054 course. **b** HPLC analysis showing xylooligomer degradation by Accellerase® and **c**
1055 Accellerase® supplemented with compost21_GH11, as assessed by detector response
1056 in nanocoulombs (nC). Note that in Fig. 3c xylobiose concentration increases after 1
1057 hour as a result of fast degradation of xylooligomers with degree of polymerization
1058 (DP) ≥ 4 by compost21_GH11. Higher DP xylooligomers have a higher inhibitory
1059 effect, thus its rapid degradation seems beneficial. The commercial cocktail used was
1060 able to degrade the added xylooligomers within 24h. In consequence, after 96h,
1061 reactions with and without compost21_GH11 achieved about the same glucose yields.

1062 However, at initial stages, addition of compost21_GH11 strongly improved glucose
1063 production rate.

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

1084 **Tables**1085 **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

1086

1087 Metatranscriptome sequencing and processing metrics.

1088

1089 **Table 2**

Data collection	
Wavelength/beamline	1.45866/MX2, LNLS
Space group	P4 ₃ 2 ₁ 2
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