



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/168430/>

Version: Accepted Version

Article:

Leadbeater, Daniel Raymond, Oates, Nicola Claire, Bennett, Joe et al. (Accepted: 2020)
Mechanistic strategies of microbial communities regulating lignocellulose deconstruction in
a UK salt marsh. *Microbiome*. ISSN: 2049-2618 (In Press)

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.

Title

Mechanistic strategies of microbial communities regulating lignocellulose deconstruction in a UK salt marsh

Authors

Daniel R. Leadbeater^{1§}, Nicola C. Oates¹, Joseph P. Bennett¹, Yi Li¹, Adam A. Dowle², Joe D. Taylor⁴, Juliana Sanchez Alponi¹, Alexander T. Setchfield¹, Anna M. Alessi¹, Thorunn Helgason³, Simon J. McQueen-Mason^{1§}, Neil C. Bruce^{1§}

¹ Centre for Novel Agricultural Products, Department of Biology, University of York, York, YO10 5DD, UK.

² Bioscience Technology Facility, Department of Biology, University of York, York, YO10 5DD, UK.

³ Department of Biology, University of York, York, YO10 5DD, UK.

⁴ School of Chemistry and Biosciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP, UK.

[§] Corresponding authors: N.C.B. (email: neil.bruce@york.ac.uk), S.J.M.-M. (email: simon.mcqueenmason@york.ac.uk) or D.R.L. (email: daniel.leadbeater@york.ac.uk).

1 **Abstract**

2 **Background**

3 Salt marshes are major natural repositories of sequestered organic carbon with high burial rates of
4 organic matter, produced by highly productive native flora. Accumulated carbon predominantly
5 exists as lignocellulose which is metabolized by communities of functionally diverse microbes.
6 However, the organisms that orchestrate this process and the enzymatic mechanisms employed that
7 regulate the accumulation, composition and permanence of this carbon stock is not yet known. We
8 applied meta-exo-proteome proteomics and 16S rRNA gene profiling to study lignocellulose
9 decomposition *in situ* within the surface level sediments of a natural established UK salt marsh.

10 **Results**

11 Our studies revealed a community dominated by *Gammaproteobacteria*, *Bacteroidetes* and
12 *Deltaproteobacteria* that drive lignocellulose degradation in the salt marsh. We identify 42 families
13 of lignocellulolytic bacteria of which the most active secretors of carbohydrate active enzymes were
14 observed to be *Prolixibacteraceae*, *Flavobacteriaceae*, *Cellvibrionaceae*, *Saccharospirillaceae*,
15 *Alteromonadaceae*, *Vibrionaceae* and *Cytophagaceae*. These families secreted lignocellulose-active
16 glycoside hydrolase (GH) family enzymes GH3, GH5, GH6, GH9, GH10, GH11, GH13 and GH43 that
17 were associated with degrading *Spartina* biomass. While fungi were present, we did not detect a
18 lignocellulolytic contribution from fungi which are major contributors to terrestrial lignocellulose
19 deconstruction. Oxidative enzymes such as laccases, peroxidases and lytic polysaccharide
20 monooxygenases that are important for lignocellulose degradation in the terrestrial environment
21 were present but not abundant, while a notable abundance of putative esterases (such as
22 carbohydrate esterase family 1) associated with decoupling lignin from polysaccharides in
23 lignocellulose was observed.

24 **Conclusions**

25 Here, we identify a diverse cohort of previously undefined bacteria that drive lignocellulose
26 degradation in the surface sediments of the salt marsh environment and describe the enzymatic
27 mechanisms they employ to facilitate this process. Our results increase the understanding of the
28 microbial and molecular mechanisms that underpin carbon sequestration from lignocellulose within
29 salt marsh surface sediments *in situ* and provide insights into the potential enzymatic mechanisms
30 regulating the enrichment of polyphenolics in salt marsh sediments.

31

32 **Keywords**

33 Salt marsh, Lignocellulose, CAZyme, Carbon cycling, Carbohydrate esterase, CE1, Proteomics,
34 Transcriptomics, Community profiling

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54 **Introduction**

55 Salt marshes are highly productive intertidal ecosystems that generate an abundance of organic
56 carbon in the form of lignocellulose, where net aerial primary productivity often exceeds $1\text{-}2\text{ kg C m}^{-2}$
57 y^{-1} [1-3]. This productivity is intrinsically linked to organic carbon burial rates, estimated to be $57\text{-}245$
58 $\text{g C m}^{-2} \text{y}^{-1}$ [4-6]. This indicates that salt marshes are among the most effective carbon sequestering
59 ecosystems per unit area on the planet with a total estimated sequestration capacity of 4.8 to 87.2
60 Tg C y^{-1} [7] despite occupying only $22\ 000\text{-}400\ 000\ \text{km}^2$ [6-8]. These processes contribute to an
61 increasing pool of inaccessible carbon as the salt marsh accretes. Organic carbon is introduced into
62 the ecosystem as grass lignocellulose which represents the major component of surface to shallow
63 sub-surface level carbon [9]. Furthermore, the composition of organic carbon changes with depth,
64 with an enrichment in persistent lignin derivatives while polysaccharides are lost [10, 11]. Deposited
65 lignin is subject to passive leaching of soluble and often biologically available phenols which diffuse
66 throughout the sedimentary column, adding further recalcitrance to the remaining phenolics,
67 degradation of which is suppressed in anoxic conditions [12-14]. Traditionally, primary productivity,
68 surface area, sediment deposition and transport rates, leaching and sorption govern carbon capture
69 in coastal sediments [15]. Mineral protection and preferential retention of recalcitrant organic
70 carbon are major themes governing carbon sequestration [16]; however, throughout this natural
71 biogeochemical carbon processing, microbial mechanisms of carbon transformation are present and
72 likely operate as a system-level decomposition process that influences the permanence of
73 lignocellulose and stored carbon in marsh sediments. Currently, this process is orchestrated by an
74 undefined consortium of organisms prior to entry into stable deeper sediments where this material
75 persists for millennia [6, 17].

76 Lignocellulose is a strong fibre composite material which provides mechanical support and the
77 vessels for long distance water transport in plants, and is highly resistant to degradation. It is a
78 macromolecular complex formed from cellulose microfibrils embedded in a matrix of branched
79 polysaccharides known as hemicellulose. This polysaccharide complex is interpenetrated and sealed

80 by lignin, a phenolic heteropolymer, making lignocellulose more hydrophobic and difficult to
81 degrade enzymatically [18]. The sheer abundance of lignocellulose in the terrestrial biosphere, along
82 with its complexity and recalcitrance to digestion has led to the evolution of a diverse range of
83 lignocellulolytic enzymes across the tree of life [19]. Yet very little is known about the factors that
84 regulate lignocellulose decomposition in salt marshes despite large annual inputs into these systems
85 as microdetritus that is predominantly retained and degraded on site [9, 20, 21].

86 The dominant flora in salt marsh ecosystems are perennial such as *Spartina* species. Dieback of
87 these plants introduces vast quantities of lignocellulosic biomass into the marine environment. The
88 first phase of decay occurs during dieback where terrestrial fungal plant pathogens, usually mycelial
89 *Ascomycetes*, attack the senescent plant biomass [22, 23]. These fungi target standing senescent
90 tissue that resides above-ground in a terrestrial setting and act to break open the plant cell walls as a
91 means to access the nutritionally rich cell contents leading to the resultant infected tissues
92 becoming nitrogen depleted and lignocellulose enriched [24-26]. The senescent standing tissue then
93 weakens, and the lignocellulose enriched biomass detaches from the root-rhizome becoming
94 deposited onto the sediment surface where it transitions into a predominantly marine environment
95 with significantly greater and more variable physico-chemical pressures than terrestrial zones [27].
96 Here, it redistributes around the salt marsh surface or aggregates on the strandline where it is
97 subject to a different phase of decay.

98 Studies have established degradation rates of deposited lignocellulose *in situ* at surface levels
99 using litterbag methodologies [9, 20, 28-31], however; very little is known about the microbial
100 framework that regulates this decomposition or the enzymatic mechanisms employed to
101 deconstruct the complex lignocellulosic substrate. *In vitro* studies have suggested that bacteria, such
102 as *Cyclobacteriaceae*, *Desulfobacteraceae*, *Flavobacteriaceae*, *Halomonadaceae*, *Oceanospirillales*
103 *Pseudomonadaceae* and *Spirochaetaceae* are involved in lignocellulose degradation in this
104 environment with fungi becoming competitively displaced [32-34]. Beyond this, our understanding

105 of the functional groups involved in the decomposition process and the biocatalytic strategies they
106 employ to achieve this are poorly understood. *In vitro* studies are divorced from environmental
107 factors and the findings require cautious interpretation, as results cannot be directly extrapolated
108 into the context of ecosystem processes. Additionally, the function of an organism cannot be
109 determined by its presence or the presence of a gene as this only deduces a potential propensity to
110 function.

111 Direct monitoring of ecological processes *in situ* has the potential to capture functional, molecular
112 and phylogenetic information at their environmental interface. To identify the microbial community
113 that regulate the initial decomposition of introduced lignocellulose at the surface level, we applied
114 meta-exo-proteome proteomics, ribosomal 16S rRNA gene phylogenetic profiling and lignocellulose
115 composition analysis to *Spartina anglica* biomass in litterbags *in situ* along a 300 m transect within
116 an established salt marsh (Welwick, UK) for 16 weeks (Figure 1, Additional file 1: Figure S1). We
117 identify lignocellulolytic enzymes from the meta-exo-proteome, ascertain the taxonomic origin to
118 identify functional groups and determine the mechanistic strategies they employ to depolymerize
119 lignocellulose.

120

121 **Materials and Methods**

122 **Experimental design**

123 The field experiment was conducted in Welwick salt marsh, Hull, Humber estuary, UK,
124 53°38'55"N, 0°01'19"E from 16/7/15 to 6/11/15 (Additional file 1: Figure S1). To mimic natural
125 lignocellulose cycling, senescent above ground *Spartina anglica* biomass was collected prior to
126 deposition during winter dieback in February-March 2015 on an adjacent intertidal mud flat (Cherry
127 Cobb sands, Humber estuary, Hull, UK). The biomass was washed free of sediment, dried at 65°C for
128 48 hours and size fractionated with a Retsch Cutting Mill SM 300 at 2300rpm. The final biomass
129 fraction consisted of 80% of >1.12mm fraction and 20% <1.12mm - >500µm fraction. Nylon 66

130 monofilament woven bags ('litterbag') (18.5cmx18.5cm) of aperture size 200µm were filled with 50g
131 of biomass and sealed with 100% polyester thread.

132 Bags were placed in a 3 x 3 conformation in five stainless steel cages (711.2mm x 711.2mm x
133 63.5mm) with 25cm legs that were interspersed by 75m along a 300m parallel elevation transect,
134 defined by plant zonation patterns (dominance of *Spartina anglica*, *Puccinellia maritima* and
135 *Salicornia europaea*). Prior to deployment, the under canopy was removed and the cages placed
136 with the bags interfacing with the sediment to facilitate crosstalk to mimic surface-interfacing
137 detritus and mapped to position with GPS coordinates (Additional file 1: Table S1).

138 Sampling was performed by removing a single litterbag from each cage. During deployment, the
139 uppermost 1-5mm of sediment surrounding the cages were sampled to act as a non-lignocellulose
140 enriched sediment day 0 outgroup control. Sampling was randomised *a priori* and occurred weekly
141 for the first six weeks and thereafter at week eight, ten and 16 for a total of 46 samples (including
142 the day 0 outgroup). Sampling began at the point of low tide and was completed within two hours.
143 Sampled bags were kept at 4°C during transport and processing began within four hours of harvest.

144

145 **DNA and RNA extraction**

146 Each biological replicate at each time point (week one-six, eight, ten and 16) were treated
147 independently. Harvested biomass was equilibrated twice with 40mL ice cold 0.5x PBS pH 8.15 and
148 centrifuged for 10 min at 4500 xg. Five 0.5g biomass aliquots per litterbag (per cage; 25 total per
149 week) were taken forward for DNA and RNA extraction. The five 0.5g biomass aliquots were added
150 to screw cap tubes (2mL) containing 0.5g 0.5mM glass beads (Sigma G9268) and 0.5g 0.1mM glass
151 beads (Sigma G8893). Cetyl trimethylammonium bromide buffer (0.5mL) containing 10% CTAB (m/v)
152 in 0.7M NaCl, 240mM potassium phosphate pH 8 and 0.1% β-mercaptoethanol and 0.4mL
153 phenol/chloroform/isoamyl alcohol (25:24:1) pH 8 were added. The samples were homogenized in a

154 TissueLyser II (Qiagen) for 2 x 2.5 min at 30/s. The tubes were then centrifuged at 4°C at 16500g for
155 15 min. The aqueous phase was transferred to a new tube and an equal volume of
156 chloroform:isoamyl alcohol (24:1) was added and centrifuged as previously and three 7.5g aliquots
157 of decaying *Spartina* biomass per litterbag (one litterbag per each of the five cages for a total of 15
158 aliquots per week; only for weeks one, three, five and ten) were taken forward for protein extraction
159 following Alessi *et al.* [35] The aqueous phase was precipitated for 16 hours at 4°C with two volumes
160 of PEG precipitation solution containing 20% (w/v) PEG8000 (Sigma) in 1.6M NaCl. The nucleic acid
161 pellet was collected by centrifugation as above for 30 min at 4°C. The pellet was washed twice in
162 75% ethanol. Total RNA from weeks one, three, five and ten (the same samples used to extract the
163 proteins within the meta-exo-proteome) were taken forward for metatranscriptomic processing.

164

165 **Meta-exo-proteome extraction**

166 Each biological replicate in the protein extraction was treated independently. Per week, for each
167 of the five cages, three aliquots (only for weeks one, three, five and ten, the same weeks utilised for
168 metatranscriptome extraction to generate complimentary paired-in-time databases) were taken
169 forward for protein extraction to generate meta-exo-proteome libraries. Each 7.5g aliquot of
170 harvested biomass was washed twice with 40mL ice cold 0.5x PBS pH 8.15 and centrifuged for 10
171 min at 4500g. The extracellular and transmembrane proteins were labelled in triplicate and 2.5g
172 biomass aliquots for each of the biological replicates were resuspended in 10mM EZ-link-Sulfo-NHS-
173 SS-biotin (Thermo Scientific #21331) in 0.5x PBS and incubated at 4°C for 1 hour. The biomass was
174 centrifuged for 10 min at 4500g as above, the supernatant was discarded and the biotinylation
175 reaction was quenched with 25mL 50mM Tris-HCl pH 8 and incubated for 30 min at 4°C. Excess Tris-
176 HCl and residual biotin was removed with two washes with 20mL ice cold 0.5X PBS pH8 with
177 centrifugation steps for 5 min at 4500g.

178 Proteins were extracted from the biomass with 10mL 2% (w/v) SDS pre-heated to 60°C and
179 incubated for 1 hour. The supernatant was extracted, and the proteins were precipitated with five
180 volumes of pre-chilled (-20°C) 100% acetone and incubated at -20°C for 16 hours. Precipitated
181 proteins were pelleted by centrifuging at 4500rpm for 20 min and the residual acetone was
182 discarded. The pellets were air dried and resuspended in 1mL 0.1% SDS in 1x PBS, filtered through
183 0.22µm syringe driven filter units and loaded onto 1mL HiTrap Streptavidin HP columns (GE
184 Healthcare #17-5112-01) and incubated for 1 hour at 4°C. The proteins were eluted with 1mL 50mM
185 dithiothreitol (DTT) in 1x PBS, the column was incubated for a further 1 hour and eluted again, this
186 was performed three times and the first two 1mL fractions were pooled.

187 The protein fractions were desalted, and buffer exchanged into H₂O using 5mL Zeba™ Spin 7k
188 MWCO columns (Thermo 89882) according to the manufacturer's protocol. To concentrate, the buffer
189 exchanged protein was frozen in liquid nitrogen, lyophilized using a Heto PowerDry LL3000 Freeze
190 Dryer (Thermo) and resuspended in 210µL H₂O. All five biological replicates for each time point were
191 pooled in equal concentrations. The proteins were stored for LC-MS/MS analysis by solubilizing in
192 NuPAGE LDS sample buffer (Life Technologies) and incubating at 70°C for 10 mins prior to a short (6
193 min) run into a 7 cm NuPAGE Novex 10% Bis-Tris Gel (Life Technologies) at 200 V. The gels were stained
194 with SafeBLUE protein stain (NBS biologicals) for 1 hour before de-staining with H₂O for 1 hour. The
195 stained gels were sliced into 1mm² fragments and stored at -20°C prior to LC-MS/MS analysis.

196

197

198 **Meta-exo-proteomics, protein identification, functional annotation and taxonomic origin**

199 To generate paired-in-time reference metatranscriptome databases, total extracted nucleic acids
200 from each biological replicate were pooled in equal ratios for each time point (week one, three, five
201 and ten) and DNA depleted. Messenger RNA (mRNA) was enriched by depleting ribosomal RNA

202 (rRNA) using Ribo-Zero™ Magnetic Epidemiology rRNA removal kit (RZE1224/MRZ11124C; Illumina).
203 RNA-seq libraries were prepared using a NEBnext RNA Ultra Library preparation kit with NEBnext
204 single 6bp indexing primers (New England BioLabs, Herts, UK) and pooled in equimolar ratios. The
205 pooled RNA-seq library was spiked with 1% PhiX and sequenced on a single lane of an Illumina HiSeq
206 3000 2 x 150 base pair chip. Sequencing resulted in 82 966 97, 99 319 32, 95 318 91 and 105 517 252
207 raw reads for the metatranscriptomics databases for week one, three, five and ten respectively
208 (383,122,461 reads in total); statistics for the four individual metatranscriptomic databases and
209 totals are available in Additional file 1: Table S2.

210 To leverage the depth of sequencing and capitalise on the diversity within the temporally
211 interspersed metatranscriptomes maximise protein identification, the metatranscriptomic databases
212 for week one, three, five and ten were concatenated into a single master metatranscriptome. Raw
213 reads were searched against Silva_115 database to identify ribosomal RNA (rRNA) genes using the
214 Bowtie2 software package [35, 36]. Orphan reads in the paired reads, rRNA reads and poor quality
215 sequences were removed with the ngsShoRT software [37]. Duplicated libraries were assembled
216 *de novo* with the Trinity software package [38]. Read counts and gene abundance were obtained
217 with the Trinity utility programs. The *de novo* assembled metatranscriptomic databases contained 29
218 938 868 contiguous sequences (contigs). Contigs ≤ 500 bp were filtered, split into open reading
219 frames (ORFs) using Emboss GETORF (<http://www.bioinformatics.nl/cgi-bin/emboss/getorf>) that
220 were ≥ 300 bp and includes alternative initiation codons and dereplicated resulting in 2 400 360
221 unique ORFs within the master metatranscriptome.

222 To generate paired-in-time exo-meta-proteome databases, biological replicates at week one,
223 three, five and ten were pooled and protein identification was performed in triplicate for each pool
224 at each time point (N=3 for each of week one, three, five and ten). Tryptic digestion was performed
225 for in-gel proteins post reduction with DTE and S-carbamidomethylation with iodoacetamide. Gel
226 pieces were washed twice with 50% (v:v) aqueous acetonitrile containing 25 mM ammonium

227 bicarbonate and finally washed with acetonitrile and then dried. Modified porcine trypsin (Promega,
228 Southampton, UK) was dissolved in 50 mM acetic acid and diluted with 25 mM ammonium
229 bicarbonate to 0.02 µg/µL. 25 µL of trypsin solution was added and incubated for 10 min before
230 adding 25 mM ammonium bicarbonate to submerge to gel pieces and incubated further for 16 hours
231 at 37°C. Three washes were performed with 50% (v:v) aqueous acetonitrile containing 0.1% TFA
232 (v:v), dried and reconstituted in aqueous 0.1% trifluoroacetic acid (v:v).

233 The acquisition of peptide spectra was performed in triplicate for each time point and was
234 achieved using a nanoLC system interfaced with a maXis HD LC-MS/MS system and a CaptiveSpray
235 ionisation source (Bruker Daltonics, Coventry, UK). Positive ESI-MS and MS/MS spectra were
236 acquired using AutoMSMS mode. Instrument control, data acquisition and processing were
237 performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker
238 Daltonics). Instrument settings were: dry gas; 3 L/min, ion acquisition range; m/z 150-2 000, MS/MS
239 spectra rate; 5 Hz at 2 500 cts to 20 Hz at 250 000 cts, quadrupole low mass; 300 m/z , cycle time; 1 s,
240 ion spray voltage; 1,450 V, collision RF; 1,400 Vpp, transfer time; 120 ms, MS spectra rate; 5 Hz, dry
241 gas temperature; 150°C, absolute threshold 200 counts, preferred charge states: 2–4, singly charged
242 ions excluded. A single MS/MS spectrum was acquired for each precursor and former target ions
243 were excluded for 0.8 min unless the precursor intensity increased fourfold.

244 Our approach of shotgun LC-MS/MS-based proteomics allows in-depth proteomic analysis, but is
245 only effective if the peptide spectra can be matched to a corresponding sequence database. Because
246 the salt marsh environment has been little explored at the molecular level, we used the
247 metatranscriptome libraries, that were generated at the same time points (paired-databases in time)
248 as the proteomic studies, as reference libraries to map peptide spectra to their originating sequence.
249 We concatenated the four metatranscriptomic databases (week one, three, five and ten) into a
250 master databases to capitalise on the diversity within the temporally interspersed
251 metatranscriptomes and used this gene expression data to identify meta-exo-proteome proteins

252 from peptide spectra, shedding new light on the communities of microbes in this environment and
253 their activities.

254 To identify proteins from LC-MS/MS spectra, peptide spectra generated from the digested meta-
255 exo-proteome proteins were mapped back to originating sequences in the ORF library generated
256 from the concatenated metatranscriptomic assemblies. Firstly, redundant sequences in the ORF
257 database were removed by leveraging non-redundant sequences in an initial round of high
258 stringency searching ($p=0.05$) against 21 subsets of $\sim 115\,000$ sequences (252 searches total),
259 followed by the concatenation of sequence hits into a secondary 'true hit' database (containing 42
260 894 sequences) with minimal redundancy, the final search against the true hit database ($p=0.1$)
261 yielded 11,268 unique proteins; individual peptide spectral matches were filtered to require expect
262 scores of 0.1 or better. Peptide spectra generated with LC-MS/MS were cross referenced with ORF
263 sequences using Mascot version 2.5.1 (Matrix Science Ltd.), through the ProteinScape interface
264 version 2.1 [39]. The search criteria for both searches were +2/+3/+4 peptide charge, peptide
265 tolerance ± 10 ppm, modifications; carbamidomethyl and oxidation. Analysis was performed using
266 the quantitative exponentially modified protein abundance index (emPAI) [40]. emPAI values for
267 each protein were then normalized to generate the molar percentage.

268 dbCAN was used to identify carbohydrate active enzymes (CAZyme) within the meta-exo-
269 proteome and the metatranscriptomic databases using HHMER3 [41]. The meta-exo-proteome was
270 also searched against the NCBI non-redundant protein database (NR_prot; 1:62) using BLAST+
271 (BlastP) version 2.2.31 with an expect value threshold of $1e^{-5}$ [42]. The resulting best-hit was
272 obtained for each protein in the meta-exo-proteome and NCBI Accession and TaxID database was
273 compiled and the most likely taxonomic origin of these proteins were established using tools within
274 the Environment for Tree Exploration (ETE) version 3 toolkit [43]. To delineate functional members
275 of the microbial community associated with the *Spartina* biomass, we cross-referenced the 16S rRNA
276 gene phylogenetic profile with the taxonomic origin of the meta-exo-proteome proteins.

277

278 **16S rRNA gene and ITS2 amplicon sequencing and analyses**

279 Biological replicates were treated independently (N=5) for each of the 9 time points (week one-
280 six, eight, ten, 16 and the day 0 outgroup for a total of 46 data points). Total extracted nucleic acids
281 were RNase A treated in triplicate. The ribosomal 16S rRNA gene V4 region was targeted with
282 primers; 515f-Y GTGYCAGCMGCCGCGGTAA (5'-3') [44] and 806R GGACTACNVGGGTWTCTAAT (5'-3')
283 [45]. The internal transcribed region 2 (ITS2) region was targeted with primers; fITS7
284 GTGARTCATCGAATCTTTG (5'-3') [46] and ITS4ngs TCCTSCGCTTATTGATATGC (5'-3') [47]. Cluster
285 identification was enhanced with a random dodecamer sequence NNNHNNNWN (5'-3')
286 prepended to the forward primer [48].

287 16S rRNA gene polymerase chain reactions (PCR) were performed in 25µL volumes containing
288 200µM dNTPs, 0.5µM 515fY-MN, 0.5µM 806rMN, 50ng gDNA, 0.5U Phusion HF polymerase
289 (#M0530) and 1x Phusion HF Buffer. Thermocycling conditions included an initial denaturation at
290 98°C for 30s, followed by 28 cycles of 98°C for 10s, 53°C for 30s and 72°C for 15s and 72°C for 10
291 min. ITS2 PCR were performed as above with thermocycling conditions including an initial
292 denaturation at 98°C for 30s, followed by 34 cycles of 98°C for 10s, 57°C for 30s and 72°C for 20s,
293 with final extension at 72°C for 5 min. Indexing was performed using the Nextera XT™ library
294 preparation kit (Illumina FC-131-1001). The libraries were pooled in equimolar concentrations to
295 4nM, spiked to 1% PhiX and run on a MiSeq 250bp x 2 cartridge (MiSeq Reagent Kit v2 (500 cycles)
296 MS-102-2003, Illumina).

297 The generated 16S rRNA genes libraries averaged 54,929 sequences. Fastq merging was
298 performed with Vsearch version 1.11.1 [49]. The generated ITS2 libraries averaged 50,843
299 sequences and were processed with ITSx [50] to filter non-fungi sequences. The resulting fungi only
300 ITS2 libraries averaged 28,972 sequences. The primer sequences were trimmed using Cutadapt
301 (version 1.11.). Sequences were trimmed to global lengths of 250bp using Usearch (version 9; -

302 fastx_truncate) [51]. Amplicon profiles were dereplicated, purged of singletons, assigned abundance
303 and sorted by size using Usearch (version 7: -derep_fulllength) [51]. Clustering was performed using
304 the UPARSE algorithm [52], with concurrent *de novo* chimera detection using Usearch (version 9 -
305 cluster_otus) with a 97% identity threshold resulting in 5122 non-chimeric OTUs that were taken
306 forward for analysis. Representative sequences for each OTU were then mapped to original
307 sequences using Usearch (version 7: -usearch_global). Taxonomy was assigned with QIIME [53]
308 (version 1.9; assign_taxonomy.py) using SILVA 132 [35] for the 16S rRNA libraries and UNITE v7.1
309 [54]. Rarefaction analysis [53] displayed curves that begin to reach asymptotic levels, indicating
310 sufficient depth for analysis but not complete diversity coverage (Additional file 1: Figure S2). The
311 taxonomy of any unassigned OTUs (N=610), using UNITE in the ITS2 libraries were further classified
312 using BLASTn against the GenBank non-redundant nucleotide database. Non-fungal OTUs were
313 discarded (N=) and missing taxonomies of on target OTU sequences were manually curated (N=393)
314 resulting in a total of 920 fungal OTUs which were subsequently analysed. Fungal OTUs were
315 classified into functional guilds using FUNGuild [55] which assigned a functional guild to 419 OTUs
316 from 724 matches of the original 920, this represented $51.4 \pm 2.12\%$ mean OTU abundance across
317 all time points and was taken forward for analysis. All commands for the analysis pipeline are
318 available in Additional file 1: Table S3.

319

320 **Determining highly productive groups**

321 A productivity index was used to elucidate taxonomic groups with disproportionately greater
322 CAZyme production per unit abundance, given by $\log_{10}(\sum \bar{x} \text{ mol\%/abundance})$. Disproportionately
323 productive groups were determined as those with an index >0.3 in at least 1 observation.

324

325 **Network associations**

326 Network associations between meta-exo-proteome CAZyme classes and taxonomic classes were
327 constructed by grouping annotated domains ($\leq 1e^{-10}$) into CAZyme classes by average $\sum \bar{x}$ mol%
328 across the entire time course and connecting these nodes to the taxonomic classes the domains
329 originated from. Classes are presented by their mean $\sum \bar{x}$ mol% output. Taxa $< 0.025 \sum \bar{x}$ mol% with \leq
330 5 edges (connections) and CAZyme classes $< 1.25 \times 10^{-3} \sum \bar{x}$ mol% were filtered for clarity. Plots were
331 generated with NetworkX [56].

332

333 **Biomass composition analysis**

334 Biomass was washed free of sediment through 100 μ m mesh with free flowing dH₂O. Total
335 biomass was measured as the mass balance of lyophilised material. Ash was determined with 1g of
336 biomass incubated 600°C for 24 hours. Matrix polysaccharides were measured using triflouracetic
337 acid methodology [57]. Cellulose was subsequently determined using the Updegraff and Saeman
338 hydrolysis [58, 59]. Lignin was measured as acetyl bromide soluble lignin [60] using a previously cited
339 extinction coefficient of 17.75 for grasses [61].

340

341 **Statistics**

342 One-way ANOVAs and Tukey's HSD tests were performed using SciPy [62] and Scikit [63]
343 respectively. All data were assessed for normality using the Shapiro-Wilk test. Statistical analyses
344 were performed on non-normalised data.

345

346 **Results**

347 **Functional assignment of the meta-exo-proteome**

348 To identify the lignocellulolytic enzymes involved in biomass breakdown, we employed a
349 metaproteomic analysis of extracellular proteins (meta-exoproteome), accomplished by an affinity
350 tagging process using a membrane-impermeable biotinylation tag [64]. Because lignocellulose is an
351 insoluble macromolecule, it generally has to be broken down by extracellular enzymes. Many of the
352 enzymes involved adhere to the lignocellulose or the microbe, and the use of surfactants to extract
353 them leads to cell lysis and contamination with intracellular proteins. The tagging approach avoids
354 the problem of intracellular contaminants, allowing a focus on extracellular and cell surface proteins.

355 Annotation of the transcriptome revealed 103 CAZyme families ($\leq 1e^{-5}$ and transcripts per million
356 (TPM) ≥ 1) across 44,334 ORFs (excluding glycosyl transferases), the total proportion of CAZymes
357 across all transcriptomic databases was 429.27 ± 62.16 TPM (Additional file 1: Figure S3). Proteomic
358 analysis identified 11,268 proteins within the meta-exo-proteome, of which 320 ($\leq 1e^{-10}$) were
359 annotated as putative carbohydrate active domains (CAZyme) within 252 peptide matching ORFs
360 across 81 CAZyme families. Families present within the metatranscriptomic databases that were
361 absent from the meta-exo-proteomes were largely families not specific to lignocellulose degradation
362 or families usually associated with core intracellular activities (AA6, CE14, GH32, GH57, GH73, GH92
363 and GH108) or CAZyme families containing enzymes with both intracellular and extracellular
364 localisations (GH1, GH2, CE7) with the exception of a small subset of predominantly pectin-targeting
365 CAZymes; CE2, CE7, CE8, CE12, CE15, GH28 and GH105 and AA4 (Additional file 1: Figure S3).
366 Instead, the exo-meta-proteome predominantly consisted of pectin-targeting CAZyme families CE3,
367 CE4, CE6, PL1, PL4, GH35 and GH43.

368 CAZyme homologs ($\leq 1e^{-10}$) represented only 0.72-0.99 mol% of the total meta-exo-proteome
369 concordant with previous *in vitro* reports [65]. The meta-exo-proteome CAZyme profile revealed
370 three dominant Euclidean clusters of temporally abundant classes which contain a diverse collection
371 of activities (Figure 2c). Glycoside hydrolases (GH) were the most abundant class with 37 families
372 identified. GH3, GH5 and GH6 family enzymes were abundant, these classes are typically related to

373 cellulose degradation, many of which were associated to carbohydrate binding domains (CBMs). The
374 CBM profile of our data highlighted two abundant Euclidean clusters (Figure 2d); the dominant of
375 which contained CBM2 and CBM44 motifs associated with cellulose and matrix polysaccharide
376 binding and a secondary cluster containing CBM10, CBM5 and CBM60 which have been associated
377 with cellulose, hemicellulose and chitin binding, respectively. Families associated with hemicellulose
378 degradation were abundant, notably GH10, GH11 and GH16 typically associated with xylan
379 degradation.

380 A rapid loss of dry mass was observed with a reduction of 69% during the 16 week period. The
381 distribution of CAZyme family proteins coupled with the biomass composition revealed successional
382 targeting of the major lignocellulose biopolymers (Figure 2a-b), that temporally synchronized with
383 the abundance of CAZyme proteins within the meta-exo-proteome. The largest loss in cellulose
384 occurs during the first week, most likely conducted by the highly abundant GH6 and GH5 family
385 enzymes coordinated with CBM2 and CBM44 domains targeting exposed cellulose microfibrils
386 generated as a result of the mechanical fractionation of the *Spartina anglica* biomass, while lignin
387 degradation appears rate limiting (weeks one and two). During weeks three to five, there was an
388 increased rate of matrix polysaccharide loss which corresponds to an increased abundance of GH11,
389 GH10, GH13 and GH43 family enzymes coupled with a concomitant decline in the rate of cellulose
390 hydrolysis, suggesting matrix polysaccharides limited cellulose access. During weeks 6 to 16, the rate
391 of cellulose deconstruction increases and a degradative equilibrium was established.

392 An interesting finding was that carbohydrate esterases (CE) were more abundant than many GH
393 family enzymes (Figure 2b), particularly those from family 1 (CE1) that predominantly presented as
394 feruloyl esterases and acetyl xylan esterases. Auxiliary activities (AA) established largely as
395 encompassing oxidative enzymes, while present within the enzymatic profile, were not abundant.
396 We only identified two AA families; AA7 (glucooligosaccharide oxidases and chitoooligosaccharide

397 oxidases) which were transiently present during week three and AA2 (containing class II lignin-
398 modifying peroxidases) that were present at low abundances throughout the study.

399

400 **Taxonomic affiliation of meta-exo-proteome proteins**

401 Fungi and archaea were poorly represented in our metaproteome annotations and were only
402 responsible for 0.28-1.46 and 0.04-0.2 mol% of the total meta-exo-proteome, respectively. Bacteria
403 produced 99-100 mol% CAZymes. Indeed, within the CAZyme profile, the only notable proteins not
404 of bacteria/archaea origin showed homology to Annelida (AA2) and Chlorophyta (AA3) enzymes.
405 This was concordant with the total meta-exo-proteome, of which 66.5-79.5 mol% originated from
406 bacteria/archaea.

407 Proteins that originated from families also identified in the 16S rRNA gene derived community
408 profile accounted for $75 \pm 6.9\%$ CAZyme mol%. The results indicate *Proteobacteria* and *Bacteroidetes*
409 are the dominant producers of lignocellulolytic enzymes (Figure 3a). *Gammaproteobacteria* and
410 *Deltaproteobacteria* were responsible for $39.03 \pm 13.65\%$ and $7.48 \pm 3.95\%$ of total CAZyme mol%,
411 respectively, while *Bacteroidia*, *Flavobacteriia* and *Cytophagia* were responsible for $12.45 \pm 6.30\%$,
412 $9.25 \pm 2.55\%$ and $7.45 \pm 3.03\%$ of the total CAZyme mol%, respectively. This is concordant with the
413 16S rRNA gene abundance of these two phyla, which is maintained at $78.43 \pm 4.10\%$. Investigations
414 revealed *Alteromonadaceae* (*Alteromonas*, *Rheinheimera* and *Catenovulum*), *Vibrionaceae* (*Vibrio*),
415 *Flavobacteriaceae* (predominantly *Lutibacter*, *Wenyngzhuangia* and *Flavobacterium*),
416 *Cellvibrionaceae*, *Saccharospirillaceae* and *Reinekea*, *Prolixibacteraceae* (predominantly
417 *Draconibacterium*, *Prolixibacter* and *Sunxiuqinia*), *Marinilabiliaceae* (*Saccharicrinis*),
418 *Saccharospirillaceae* (*Reinekea*) and *Bacteroidaceae* (*Bacteroides*) as dominant CAZyme producers
419 (Figure 3). Groups with disproportionately high CAZyme productivity relative to their abundance
420 were revealed as *Bacteroidaceae* (*Bacteroides*), *Paludibacteraceae* (*Paludibacter*),
421 *Flammeovirgaceae* (*Flexithrix*), *Sphingobacteriaceae*, *Melioribacteraceae* (*Melioribacter*),

422 *Chromatiaceae*, *Peptococcaceae* and *Salinivirgaceae* (*Salinivirga*) (Additional file 1: Figure S5).
423 CAZyme productive but poorly resolved genera included *Teredinibacter*, *Sporocytophaga*,
424 *Aquimarina*, *Hyunsoonleella*, *Planococcus*, *Pseudosphingobacterium*, *Desulfosporosinus*, *Formosa*,
425 *Simiduia*, *Sorangium*, *Lentimicrobium*, *Arcticbacter*, *Desulfobulbus*, *Saccharophagus* and
426 *Chitinophaga* (Additional file 1: Figure S4).

427 Fungi were identified within the sediment and lignocellulosic material but no CAZymes originating
428 from fungi were detected. Significant changes in fungal OTU richness was observed (ANOVA, $F_{8,36} =$
429 14.95 , $p < 2.29 \times 10^{-9}$) with a significant increase (ANOVA, $F_{1,8} = 29.17$, $p < 0.0006$) between week one to
430 the observed peak during week two from 253 ± 35 to 360 ± 18.54 respectively, before entering a
431 gradual but continuous decline to week 16 ($N = 180 \pm 24.2$) (Additional file 1: Figure S11). Fungal
432 taxonomy was poorly resolved with 558 of 920 OTUs identified to class level. Identified fungi were
433 predominately *Ascomycetes* ($56.38 \pm 4.07\%$) with a small contribution from *Basidiomycota* ($5.22 \pm$
434 1.57%); however, in the day 0 sediment *Rozellomycota*, *Chytridiomycota* and *Zygomycota* were
435 observed as very minor components.

436 *Saccharomycetales* and *Pleosporales* were consistently abundant within the lignocellulose
437 associated fungal community throughout the 16 week time course ($25.6 \pm 3.23\%$ and $10.85 \pm 1.74\%$
438 respectively) (Additional file 1: Figure S11). Notable components of the early fungal profile included
439 *Hypocreales*, *Capnodiales*, *Tremellales* (weeks one through to three) before rapidly declining and
440 seemingly displaced by *Microascales* which enrich and dominate the profile between week five (1.82
441 $\pm 1.2\%$) and six onwards ($10.2 \pm 4.45\%$). Functional classification of these OTUs in terms of nutrient
442 acquisition strategy revealed the dominant guild to be saprotroph, followed by pathotroph-
443 saprotroph (Additional file 1: Figure S12), of which the most prevalent trophic modes were
444 undefined saprotroph which enriched gradually from 32.3% in the day 0 sediment outgroup to 86.1
445 $\pm 3.32\%$ at week 16, endophyte-lichen parasite-plant pathogen-undefined saprotroph which were
446 consistent between week two to 16 ($20.1 \pm 1.91\%$) and animal pathogen-endophyte-lichen parasite-

447 plant-pathogen-soil saprotroph-wood saprotroph which was a large component only during weeks
448 one to three ($13.8 \pm 1.72\%$; Additional file 1: Figure S11), both of which are poorly resolved
449 definitions. Interestingly, modes associated with the turnover of lignocellulosic substrates such as
450 wood saprotroph and leaf saprotroph were more abundant in the day 0 sediment outgroup than in
451 the lignocellulose associated community.

452 Filtering out non-CAZyme productive lineages revealed a rapid enrichment for CAZyme producing
453 families relative to the day 0 sediment outgroup. This suggests that within the sediment a maximum
454 of 13.9% of the bacteria/archaea microbiome at family level functioned as lignocellulose degraders
455 whilst OTU richness was highest (2277). During the first week within the biomass, we observed an
456 enrichment in CAZyme productive lineages of 3.77 ± 0.11 fold to $52.40 \pm 1.51\%$ of the total
457 community whilst OTU richness declined (1020 ± 65), increasing to $59.56 \pm 1.66\%$ in week two
458 (Figure 3b-c). We observed significant variation in OTU richness over time (ANOVA, $F_{8,36} = 17.59$,
459 $p < 0.000005$), increasing significantly from week one to three and all time points thereafter (Tukey
460 HSD, $p < 0.015$). OTU richness continued to increase toward day 0 outgroup levels while no significant
461 decline in the abundance of CAZyme producing members was observed during the time course
462 (ANOVA, $F_{8,36} = 1.78$, $p > 0.114$) suggesting the colonisation of diverse heterotrophs and secondary
463 metabolizers. Concordantly, the total CAZyme mol% was not significantly different throughout the
464 time series (ANOVA, $F_{3,8} = 1.06$, $p > 0.42$).

465 We noted a degree of congruence between the enzymatic profiles of the most productive groups
466 within *Proteobacteria* and *Bacteroidetes* (Figure 4, Additional file 1: Figures S4-7). The most
467 abundant CAZyme classes (Figures 2c-b) with the exception of GH6, CBM2, CBM10 and CBM44 that
468 were produced exclusively by *Gamma*proteobacteria, represented a core suite of activities (Figure 4)
469 and were produced by multiple divergent lineages, suggesting a common mechanistic strategy was
470 employed by the major CAZyme-producing consortia (Figure 4).

471 *Gammaproteobacteria* maintain unparalleled levels of CAZyme production across the time course
472 despite a reduction in their overall abundance. This is due to an enrichment in clades exhibiting high
473 CAZyme production, e.g. *Alteromonadaceae*, *Saccharospirillaceae* and *Vibrionaceae*.
474 *Cellvibrionaceae*, *Alteromonadaceae* and *Saccharospirillaceae* are not abundant in sediments but
475 progressively became major components of the *Gammaproteobacteria* profile in both the
476 community profile and their CAZyme output. *Vibrionaceae* appear transient with peak abundance
477 during week one ($22.98 \pm 3.27\%$) which precedes a steady decline ($5.25 \pm 1.04\%$ and $2.70\% \pm 0.70\%$
478 in weeks two and three, respectively), indicating that this clade represent rapid colonizers and
479 opportunistic oligotrophs. *Vibrionaceae* was predominantly comprised of two genera, *Vibrio* and
480 *Photobacterium*. Subsequently, *Vibrionaceae* appears to be outcompeted by *Alteromonadaceae*,
481 *Saccharospirillaceae* (*Reinekea*), and *Cellvibrionaceae* (predominantly *Marinagarivorans*); accounting
482 for much of the decline in the *Gammaproteobacteria* profile in weeks one to five. Identifiable
483 *Alteromonadaceae* genera included *Alteromonas*, *Glaciecola* and *Paraglaciecola*.
484 *Gammaproteobacteria* abundance is supplanted by *Deltaproteobacteria* groups; *Desulfobulbaceae*,
485 *Desulfuromonadaceae* and *Desulfovibrionaceae* and *Bacteroidetes* groups; *Prolixibacteraceae*
486 (*Draconibacterium* and *Roseimarinus*), *Flavobacteriaceae* (*Lutibacter*) and *Marinilabiliaceae*
487 (*Labilibacter*). Families within *Firmicutes* and *Verrucomicrobia* were active CAZyme secretors despite
488 low apparent abundances; particularly *Peptococcaceae*, *Planococcaceae* and *Paenibacillaceae* and
489 *Rubritaleaceae*.

490

491 **Discussion**

492 We examined the process of surface level lignocellulose decomposition within a natural salt
493 marsh environment demonstrating a framework wherein lignocellulose decomposition can be
494 monitored *in situ*. Our data suggest a large proportion of the total native microbiome is
495 lignocellulose responsive and capable of rapid colonization and restructuring to take advantage of
496 this annual influx of carbon. Our metaproteomic studies revealed an enrichment for activities that

497 target linkages between lignin and polysaccharides as well as glycanohydrolases and a marked
498 sparseness of oxidative enzymes that attack lignin.

499 It is notable that although the total biomass in our mesh bags was reduced by about 70% over a
500 16 week period, our results conform to a first order decay model alluded to in previous experiments
501 [28, 66-68]. These previous studies suggest the majority of particulate decomposition occurs within
502 the first year of entry into the system and proceeds through three phases; the leaching of soluble
503 compounds, decomposition and a final refractory phase characterized by diminished rates of
504 decomposition [28, 66-68]. Valiela et al., [66] suggest the refractory period is confined to
505 decomposition rates below 0.4% day⁻¹. In our study, the decomposition rates for the refractory
506 period during weeks eight and 16 were observed to be 0.22 ± 0.087% day⁻¹ and 0.176 ± 0.03% day⁻¹
507 respectively, suggesting our experiment ran into the refractory period.

508 We assessed surface level, aerobic lignocellulose decomposition as it has been shown to be
509 significantly more efficient than sub-surface decay [30]. Valiela et al. [30] explored the relative
510 composition of *Spartina alterniflora* for 24 months beginning in winter. While the study findings are
511 not directly comparable to our own due to differing location, start date and species of biomass,
512 which significantly affects decomposition [68], there is an undeniable synchrony between the
513 profiles of lignocellulose degradation in both studies. Both studies demonstrated an initial increase
514 in cellulose and hemicellulose, sequentially followed by lignin degradation, then hemicellulose
515 degradation. The hemicellulose degradation then coincides with cellulose degradation whilst lignin
516 increases.

517 The relative enrichment in lignin, accepted as the most recalcitrant component of lignocellulose
518 [18], suggests this biopolymer is not actively targeted for metabolism by the microbial community.
519 Salt marsh sediments are known to be significantly enriched in lignin-derived high molecular weight
520 polyphenols, with these increasing in concentration with depth [10, 11, 30, 69]. Conversely, the
521 more biologically available polysaccharides reduce with depth as they are known to be preferentially

522 targeted [70-72]. As lignin interpenetrates the core polysaccharides in the lignocellulosic matrix it
523 must be removed before the internal polysaccharides are accessible for digestion. Oxidative
524 enzymes are the predominant mechanism exhibited in terrestrial systems to modify and degrade
525 lignin, yet in our study only AA2 family members were present at low abundances. These enzymes
526 attack lignin moieties to modify the structure and it is unlikely they are responsible for cleaving high
527 molecular weight phenolics that are observed in salt marsh sediments. These findings suggest that
528 native salt marsh organisms have enzymes responsible for lignin modification that are not yet known
529 or that they adopt other mechanisms able to facilitate access to the valuable sugars present in
530 lignocellulose.

531 Instead, we note that the salt marsh meta-exo-proteome has a high representation of
532 carbohydrate esterases (CE), particularly from family 1 (CE1). CE1 family enzymes function non-
533 oxidatively to remove cinnamoyl and acetyl esters from xylans, disrupting the lignin-carbohydrate
534 complex interface between hemicellulose and lignin, and hemicellulose and cellulose respectively
535 [18, 73]. Lignin-carbohydrate complex linkages are thought to consist mainly of aryl ester (from
536 ferulic acid to arabinose in grasses like *Spartina anglica*) and aryl ether bonds, hydrolysis of which
537 decouples the lignin, exposing the surface of the remaining polysaccharides [74]. The CE1 family
538 includes a range of esterases, especially those which hydrolyse ester links between arabinoxylans
539 and ferulic and coumaric acid residues. Ferulic acid residues in arabinoxylans are particularly
540 important in providing linkages between arabinoxylan chains and between arabinoxylans and lignin,
541 thereby contributing significantly to lignocellulose recalcitrance [18, 75, 76]. CE1 also contains xylan
542 acetylesterases that remove acetyl groups from arabinoxylan, having major impacts on their three
543 dimensional conformation and ability to bind cellulose [77]. Previous compositional analysis of
544 decomposed lignocellulose in salt marshes have revealed trans-ferulic acid was responsible for 57-
545 82% of the total lignin loss which agrees with the mechanism identified in our study [78]. This
546 indicates that the linkages holding lignin to the polysaccharides of lignocellulose may be major
547 targets to allow GHs access to their substrates. We contend that this mechanism is favourable within

548 salt marshes in contrast to terrestrial systems due to the liquid medium facilitating desorption of
549 dissociated lignin macromolecules into the surrounding waters, circumventing the requirement for
550 total deconstruction. This mechanism could explain the enrichment of persistent lignin-rich particles
551 known to accumulate in salt marsh sediments through the cleavage of high molecular weight
552 phenolics. These phenolics are then likely subject to oxidative modification by the low abundance
553 AA2 family enzymes causing them to slowly become biologically available.

554 Previous studies suggest lignocellulose degradation within sediments is driven by bacteria, which
555 is supported by our data [32, 34], yet fungi are known to populate salt marsh sediments but their
556 function, community ecology and interactions remain elusive [79]. We did identify a handful of
557 fungal families with potential historical connections to lignocellulose, predominantly *Pleosporaceae*,
558 *Hypocreaceae*, *Nectriaceae*, *Sordariaceae* and *Saccharomycetales* [80]. Nutrient acquisition
559 strategies of the identified fungi revealed the dominant trophic mode to be saprotroph (acquire
560 nutrients from dead organic matter) and to a lesser extent pathotroph (acquire nutrients by
561 attacking cells) and combinations thereof. This suggests most fungi were acquiring nutrients from
562 alternative dead organic matter or were utilising a pathotrophic acquisition strategy where
563 lignocellulose is not a primary target. A notable observation was that wood saprotrophs and leaf
564 saprotrophs, which would be expected to thrive on the dead *Spartina* biomass which included
565 stems, leaves and sheaths, were present in the sediment but were not abundant on the
566 lignocellulosic material. As fungi are orders of magnitude less abundant than bacteria in this system
567 [81, 82] and we did not detect lignocellulolytic enzymes from these groups within the meta-exo-
568 proteome despite fungal enzyme sequences being well represented in archive databases, our data
569 would suggest their influence on lignocellulose decomposition for material within salt marsh
570 sediments is negligible and they likely target alternative sources of organic matter that are present
571 or cohabit within the lignocellulosic aggregate.

572 Bacterial families that have been implicated with salt marsh lignocellulose degradation based on
573 isotope probe experiments include *Desulfobacteraceae*, *Spirochaetaceae*, *Kangiellaceae* [32] and
574 selective enrichments include *Flavobacteriaceae*, *Cyclobacteriaceae*, *Pseudomonadaceae* and
575 *Halomonadaceae* [34]. We did not observe the groups reported by Darjany *et al.* [32] to be active
576 lignocellulose degraders, since the majority of lignocellulose deconstruction occurs within the
577 extracellular matrix the breakdown products are available to all microbes within proximity, therefore
578 the ¹³C approach employed by Darjany *et al.* [32] possibly identified benefactors of breakdown
579 products rather than organisms actively degrading lignocellulose. We did identify all major groups
580 reported by Cortes-Totalpa *et al.* [34] in our *in situ* study which confirm these groups to be active
581 secretors of lignocellulolytic enzymes. We also identified an additional 38 families that were not
582 previously known to actively secrete lignocellulose active enzymes. The 42 CAZyme producing
583 families reported here underpin long term carbon sequestration using a mechanism that appears to
584 favour the degradation of complex polysaccharides by selectively avoiding lignin degradation. This
585 process not only expands the pool of stored carbon but also reduces complex carbohydrates to
586 biologically available molecules within the extracellular space for the wider microbial community.

587 The CAZyme-producing *Gammaproteobacteria* described here appeared to be early colonizers of
588 lignocellulose that undergo taxonomic restructuring to favor heterotrophic lineages.
589 *Gammaproteobacteria* are displaced by CAZyme-producing groups belonging to *Bacteroidetes* and
590 *Deltaproteobacteria* clades. The results suggest the *Gammaproteobacteria* families are the
591 ecologically dominant surface level lignocellulose degraders. The divergent families identified within
592 *Bacteroidetes* and *Deltaproteobacteria* suggest they are highly active at surface levels, but likely
593 dominate carbon cycling in the oxygen depleted cores of biomass aggregates and in shallow to deep
594 subsurface sediments as they have been identified in abundance within deeper sediments. However;
595 their ecological functions were previously unknown [83].

596 Well studied examples of marine lignocellulolytic *Gammaproteobacteria* include *Saccharophagus*
597 *degradans* and the closely related *Teredinibacter turnerae*, belonging to families *Alteromonadaceae*
598 and *Cellvibrionaceae*, respectively. Both families were abundant within the lignocellulose responsive
599 microbiome and identified to be highly productive of CAZymes and interestingly, neither family were
600 well represented within the day 0 sediment outgroup suggesting they function as saprotrophs within
601 the salt marsh. *S. degradans* is a well characterised free-living heterotroph that appears fully capable
602 of deconstructing complex plant cell wall polysaccharides and many other biopolymers [84, 85]. The
603 use of these bacteria as a source of enzyme cocktails for lignocellulose saccharification has been
604 explored due to the broad complement of CAZymes [86] and full cellulolytic system [87]. Dominant
605 CAZymes within *S. degradans* culture supernatant include GH3, GH5, GH6, GH9, GH10 and GH16
606 many of which are multi-domain with a prevalence of CBM2 and CBM10 containing proteins as well
607 as CBM6, CBM13 and CBM32 [88], all of which collectively correspond to highly abundant
608 *Gammaproteobacteria* associated CAZyme families identified within our exo-meta-proteomics.

609 *T. turnerae* is a facultative intracellular endosymbiont found in wood-boring bivalves, it is
610 cellulolytic with demonstrated cellulose degrading capability and more recently it has been found to
611 harbour a complex array of xylan degrading enzymes and lytic polysaccharide monooxygenases [89,
612 90], yet it possesses a relatively small repertoire of CAZymes that only target woody plant biomass
613 within its genome compared to *S. degradans* [91]. These are predominantly GH5, GH6 and GH11
614 that also include multi-domain proteins often associated with CBM5 and CBM10 domains that were
615 well represented in our corresponding *Gammaproteobacteria* exo-meta-proteome [91-93]. The
616 prevalence of multi-domain proteins containing CBMs, particularly within the cellulases, observed
617 both within the well-characterised marine isolates and observed to be highly abundant within this
618 study have been suggested to function as a tether as opposed to a catalytic enhancement [88]. Less
619 than 40% of terrestrially derived cellulases contain CBMs that are conventionally thought to increase
620 the effective concentration of enzyme at the substrate surface and therefore rates of activity [94]. In
621 salt marshes, adsorption of CBM containing enzymes would act to tether and localise the enzyme to

622 the substrate, improving substrate beneficiation to the secretor by preventing them from being
623 washed away in these intertidal regions. It is possible that the predominance of CAZymes possessing
624 CBMs has facilitated the *Gammaproteobacteria* to flourish within the early stage of surface level
625 decomposition within the salt marsh as observed in both our exo-meta-proteome and independent
626 16S rRNA amplicon profile.

627 Carbohydrate active enzymes are an incredibly broad designation of enzymes that includes both
628 the intracellular and extracellular biosynthesis and breakdown of complex and simple
629 polysaccharides. Accurately determining extracellular localisations of proteins from transcripts,
630 particularly in an underexplored environment such as a salt marsh is challenging. Therefore, we
631 applied exo-meta-proteomics to accurately determine proteins existing within the extracellular
632 matrix. We successfully identified 81 CAZyme families within the exo-meta-proteome of the 103
633 identified within the metatranscriptome libraries suggesting sufficient depth of coverage of the
634 extracellular encompassing families. However, due to the salt marsh ecosystem being intertidal, it is
635 possible our analysis has not detected transiently localised enzymes. This may explain the small
636 subset of predominantly pectin-targeting CAZyme families that were present within the
637 transcriptome but not detected within the exo-meta-proteome as pectin is widely considered the
638 most soluble component of the plant secondary cell wall.

639 Our approach targeted the rapid surface level deconstruction phase of lignocellulose. While the
640 salt marsh microbiome varies marginally with elevation at the sediment surface [95], it is
641 significantly variable with depth [83]. These changes are a function of oxygen depletion, leaching
642 rates, sorption characteristics and alternative respiratory terminal electron acceptor availability.
643 Considered together with the relative enrichment in lignin derived polyphenolics [10, 11, 96, 97] this
644 suggests the lignocellulose-active community could be stratified with depth. These communities may
645 employ alternate mechanisms than identified here that target the most recalcitrant, lignin enriched
646 material that has once passed through the initial surface level decomposition phase we describe.

647 Our results only capture the initial rapid surface level decomposition phase, these findings cannot be
648 extrapolated throughout the salt marsh sedimentary column where the majority of the carbon stock
649 persists. Further exploration of the microbial communities at depth is required to elucidate the
650 functional taxa and the mechanisms they employ to degrade the lignin enriched carbohydrate
651 complexes that progressively accumulate and contribute to the extensive pool of sequestered
652 carbon as substrate composition is known to modulate the mechanisms employed [98].

653

654 **Conclusions**

655 Our study captured lignocellulolytic organisms as they functioned *in situ* at their environmental
656 interface within surface sediments in a salt marsh. We identified 42 families that actively secrete
657 enzymes that act to deconstruct lignocellulosic polymers, 38 of these families had no previously
658 proven ecological function. Our data suggest that bacteria primarily orchestrate this process within
659 sediments with no detectable contribution from fungi despite being present. Our proteomic analysis
660 of the meta-exo-proteome highlighted *Gammaproteobacteria* as early lignocellulolytic colonisers
661 that are temporally displaced by *Bacteroidetes* and *Deltaproteobacteria* groups and these taxa
662 concurrently produce a core suite of diverse enzymes that act upon lignocellulose. This also revealed
663 a potential mechanism of deconstruction, driven by carbohydrate esterase family 1 enzymes, which
664 are capable of dissociating lignin macromolecules from the core polysaccharides within the
665 lignocellulosic complex. This degradative strategy potentially explains the accretion of lignin derived
666 polyphenolics within salt marsh sediments. As our study assessed early stage surface level
667 degradation, further research is required to elucidate mechanisms that drive organic carbon storage
668 and turnover in deeper sediments.

669

670

671 **Abbreviations**

672 **AA:** Auxiliary activity enzyme

673 **CAZyme:** Carbohydrate-active enzymes

674 **CBM:** Carbohydrate-binding module

675 **CE:** Carbohydrate esterase

676 **GH:** Glycoside hydrolase

677 **ITS2:** Internal transcribed spacer region 2

678 **KO:** KEGG Ontology categories

679 **LC-MS/MS:** Liquid chromatography with tandem mass spectrometry

680 **ORF:** Open reading frame

681 **OTUs:** Operational taxonomic units

682 **PBS:** Phosphate buffered saline

683 **PCR:** Polymerase chain reaction

684 **PL:** Polysaccharide lyase

685 **TPM:** Transcripts per million

686

687 **Ethics approval and consent to participate**

688 Not applicable

689

690 **Consent for publication**

691 Not applicable

692

693 **Availability of data and material**

694 Metaproteomic and metatranscriptomic databases generated during this research are available at

695 MassIVE (<https://massive.ucsd.edu/>) MSV000083872 and ProteomeXchange

696 (<http://www.proteomexchange.org/>) PXD014068. The raw 16S rRNA gene sequencing data
697 generated and analysed in this study is available from the European Nucleotide Archive
698 (<https://www.ebi.ac.uk/ena>) under accessions PRJEB32810 and ERP115532. A curated dataset is
699 available in the supplemental dataset found in Additional file 2. In-house scripts, datasets and
700 dependencies for reproducing this analysis is available at [https://github.com/leadbot/Salt-marsh-](https://github.com/leadbot/Salt-marsh-metasecretome-analysis)
701 [metasecretome-analysis](https://github.com/leadbot/Salt-marsh-metasecretome-analysis).

702

703 **Competing interests**

704 The authors declare no conflict of interest.

705

706 **Funding**

707 This work was funded by Biotechnology and Biological Sciences Research Council (BBSRC)
708 Grants BB/K020358/1, BB/I018492/1 and BB/L001926/1. D.R.L. and N.C.O. were supported by a
709 studentship from the BBSRC Doctoral Training Programme (BB/J014443/1). J.S.A. was supported by
710 the Brazilian National Council for Scientific and Technology Development (CNPq; process number:
711 232506/2014-0).

712

713 **Authors' contributions**

714 N.C.B. and S.J.M.-M. conceived the idea, provided expertise and edited the manuscript. D.R.L.,
715 N.C.B., S.J.M.-M. and T.H. designed experiments. D.R.L. conducted experiments and analyzed the
716 data, wrote and edited the manuscript. Y.L. performed RNA-seq assembly. A.A.D. performed mass
717 spectroscopy and assisted with the MS/MS analysis. N.C.O. and J.P.B., assisted in data handling,

718 protein expression and provided significant expertise and intellectual input. J.D.T. provided expertise
719 and assisted in developing the phylogenetic analysis pipeline. J.S.A. provided intellectual input. A.S.
720 assisted in cloning. A.M.A. assisted in method development.

721

722 **Acknowledgements**

723 The authors thank Dr. Deborah Rathbone and Susan Heywood at the Biorenewables Development
724 Centre (York, UK) for access to research facilities and providing invaluable expertise and assistance.
725 The York Centre of Excellence in Mass Spectrometry was created thanks to investment through
726 Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and
727 subsequent support from EPSRC (EP/K039660/1; EP/M028127/1).

728

729 **Authors' information**

730 Not applicable

731 **References**

732

- 733 1. Morris JT, Sundberg K, Hopkinson CS: Salt Marsh Primary Production and Its Responses to Relative
734 Sea Level and Nutrients in Estuaries at Plum Island, Massachusetts, and North Inlet, South
735 Carolina, USA. *Oceanography*. 2013;26:78-84.
- 736 2. Trilla GG, De Marco S, Marcovecchio J, Vicari R, Kandus P: Net Primary Productivity of *Spartina*
737 *densiflora* Brong in an SW Atlantic Coastal Salt Marsh. *Estuaries and Coasts*. 2010;33:953-962.
- 738 3. Vera F, Gutierrez JL, Ribeiro PD: Aerial and detritus production of the cordgrass *Spartina densiflora*
739 in a southwestern Atlantic salt marsh. *Botany-Botanique*. 2009;87:482-491.
- 740 4. Chmura GL: What do we need to assess the sustainability of the tidal salt marsh carbon sink?
741 *Ocean & Coastal Management*. 2013;83:25-31.
- 742 5. Hopkinson CS, Cai WJ, Hu XP: Carbon sequestration in wetland dominated coastal systems - a
743 global sink of rapidly diminishing magnitude. *Current Opinion in Environmental Sustainability*.
744 2012;4:186-194.
- 745 6. Ouyang X, Lee SY: Updated estimates of carbon accumulation rates in coastal marsh sediments.
746 *Biogeosciences*. 2014;11:5057-5071.
- 747 7. McLeod E, Chmura GL, Bouillon S, Salm R, Bjork M, Duarte CM, Lovelock CE, Schlesinger WH,
748 Silliman BR: A blueprint for blue carbon: toward an improved understanding of the role of
749 vegetated coastal habitats in sequestering CO₂. *Frontiers in Ecology and the Environment*. 2011;
750 9:552-560.

- 751 8. Chmura GL, Anisfeld SC, Cahoon DR, Lynch JC: Global carbon sequestration in tidal, saline wetland
752 soils. *Global Biogeochemical Cycles*. 2003;17:12.
- 753 9. Bouchard V, Lefeuvre JC: Primary production and macro-detritus dynamics in a European salt
754 marsh: carbon and nitrogen budgets. *Aquatic Botany*. 2000;67:23-42.
- 755 10. Fogel ML, Sprague EK, Gize AP, Frey RW: Diagenesis of organic-matter in georgia salt marshes.
756 *Estuarine Coastal and Shelf Science*. 1989;28:211-230.
- 757 11. Benner R, Fogel ML, Sprague EK, Hodson RE: Depletion of C-13 in lignin and its implications for
758 stable carbon isotope studies. *Nature*. 1987;329:708-710.
- 759 12. Barry SC, Bianchi TS, Shields MR, Hutchings JA, Jacoby CA, Frazer TK: Characterizing blue carbon
760 stocks in *Thalassia testudinum* meadows subjected to different phosphorus supplies: A lignin
761 biomarker approach. *Limnology and Oceanography*. 2018;63:2630-2646.
- 762 13. Young LY, Frazer AC: The fate of lignin and lignin-derived compounds in anaerobic environments.
763 *Geomicrobiology Journal*. 1987;5:261-293.
- 764 14. Hernes PJ, Robinson AC, Aufdenkampe AK: Fractionation of lignin during leaching and sorption
765 and implications for organic matter "freshness". *Geophysical Research Letters*. 2007;34:6.
- 766 15. Bianchi TS, Cui XQ, Blair NE, Burdige DJ, Eglinton TI, Galy V: Centers of organic carbon burial and
767 oxidation at the land-ocean interface. *Organic Geochemistry*. 2018;115:138-155.
- 768 16. Hemingway JD, Rothman DH, Grant KE, Rosengard SZ, Eglinton TI, Derry LA, Galy VV: Mineral
769 protection regulates long-term global preservation of natural organic carbon. *Nature*.
770 2019;570:228.
- 771 17. Macreadie PI, Allen K, Kelaher BP, Ralph PJ, Skilbeck CG: Paleoreconstruction of estuarine
772 sediments reveal human-induced weakening of coastal carbon sinks. *Global Change Biology*.
773 2012;18:891-901.
- 774 18. Marriott PE, Gomez LD, McQueen-Mason SJ: Unlocking the potential of lignocellulosic biomass
775 through plant science. *New Phytologist*. 2016;209:1366-1381.
- 776 19. Cragg SM, Beckham GT, Bruce NC, Bugg TDH, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison
777 J, McGeehan JE, et al: Lignocellulose degradation mechanisms across the Tree of Life. *Current*
778 *Opinion in Chemical Biology*. 2015;29:108-119.
- 779 20. Jackson D, Long SP, Mason CF: Net primary production, decomposition and export of spartina-
780 anglica on a suffolk salt-marsh. *Journal of Ecology*. 1986;74:647-662.
- 781 21. Dame RF, Stilwell D: Environmental-factors influencing macrodetritus flux in north inlet estuary.
782 *Estuarine Coastal and Shelf Science*. 1984;18:721-726.
- 783 22. Buchan A, Newell SY, Butler M, Biers EJ, Hollibaugh JT, Moran MA: Dynamics of bacterial and
784 fungal communities on decaying salt marsh grass. *Applied and Environmental Microbiology*.
785 2003;69:6676-6687.
- 786 23. Torzilli AP, Sikaroodi M, Chalkley D, Gillevet PM: A comparison of fungal communities from four
787 salt marsh plants using automated ribosomal intergenic spacer analysis (ARISA). *Mycologia*.
788 2006;98:690-698.
- 789 24. Calado MD, Carvalho L, Barata M, Pang KL: Potential roles of marine fungi in the decomposition
790 process of standing stems and leaves of *Spartina maritima*. *Mycologia* 2019, 111:371-383.
- 791 25. Torzilli AP, Andrykovitch G: Degradation of spartina lignocellulose by individual and mixed
792 cultures of salt-marsh fungi. *Canadian Journal of Botany-Revue Canadienne De Botanique*.
793 1986;64:2211-2215.
- 794 26. Wilson JO, Buchsbaum R, Valiela I, Swain T: Decomposition in salt-marsh ecosystems - phenolic
795 dynamics during decay of litter of *Spartina alterniflora*. *Marine Ecology Progress Series*.
796 1986;29:177-187.
- 797 27. Baumann H, Wallace RB, Tagliaferri T, Gobler CJ: Large Natural pH, CO₂ and O₂ Fluctuations in a
798 Temperate Tidal Salt Marsh on Diel, Seasonal, and Interannual Time Scales. *Estuaries and Coasts*.
799 2015;38:220-231.

- 800 28. Negrin VL, Trilla GG, Kandus P, Marcovecchio JE: Decomposition and nutrient dynamics in a
801 *Spartina alterniflora* marsh of the bahia blanca estuary, Argentina. Brazilian Journal of
802 Oceanography. 2012;60:259-263.
- 803 29. White DA, Trapani JM, Thien LB, Weiss TE: Productivity and decomposition of dominant salt-
804 marsh plants in Louisiana. Ecology. 1978;59:751-759.
- 805 30. Valiela I, Wilson J, Buchsbaum R, Rietsma C, Bryant D, Foreman K, Teal J: Importance of chemical-
806 composition of salt-marsh litter on decay-rates and feeding by detritivores. Bulletin of Marine
807 Science. 1984;35:261-269.
- 808 31. Bouchard V, Creach V, Lefeuvre JC, Bertru G, Mariotti A: Fate of plant detritus in a European salt
809 marsh dominated by *Atriplex portulacaoides* (L.) Aellen. Hydrobiologia. 1998;374:75-87.
- 810 32. Darjany LE, Whitcraft CR, Dillon JG: Lignocellulose-responsive bacteria in a southern California
811 salt marsh identified by stable isotope probing. Frontiers in Microbiology. 2014;5:9.
- 812 33. Benner R, Newell SY, Maccubbin AE, Hodson RE: Relative contributions of bacteria and fungi to
813 rates of degradation of lignocellulosic detritus in salt-marsh sediments. Applied and
814 Environmental Microbiology. 1984;48:36-40.
- 815 34. Cortes-Tolalpa L, Norder J, van Elsas JD, Salles JF: Halotolerant microbial consortia able to
816 degrade highly recalcitrant plant biomass substrate. Applied Microbiology and Biotechnology.
817 2018;102:2913-2927.
- 818 35. Alessi AM, Bird SM, Bennett JP, Oates NC, Li Y, Dowle AA, Polikarpov I, Young JPW, McQueen-
819 Mason SJ, Bruce NC: Revealing the insoluble metasecretome of lignocellulose-degrading microbial
820 communities. Scientific Reports. 2017;7.
- 821 36. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: The SILVA
822 ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic
823 Acids Research. 2013;41:D590-D596.
- 824 37. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nature Methods.
825 2012;9:357-U354.
- 826 38. Chen C, Khaleel SS, Huang H, Wu CH: Software for pre-processing Illumina next-generation
827 sequencing short read sequences. Source code for biology and medicine. 2014;9:8.
- 828 39. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
829 Raychowdhury R, Zeng QD, et al: Full-length transcriptome assembly from RNA-Seq data without a
830 reference genome. Nature Biotechnology. 2011;29:644-U130.
- 831 40. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: Probability-based protein identification by
832 searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551-
833 3567.
- 834 41. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: Exponentially modified
835 protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the
836 number of sequenced peptides per protein. Molecular & Cellular Proteomics. 2005;4:1265-1272.
- 837 42. Yin YB, Mao XZ, Yang JC, Chen X, Mao FL, Xu Y: dbCAN: a web resource for automated
838 carbohydrate-active enzyme annotation. Nucleic Acids Research. 2012;40:W445-W451.
- 839 43. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST plus :
840 architecture and applications. BMC Bioinformatics. 2009;10:9.
- 841 44. Huerta-Cepas J, Serra F, Bork P: ETE 3: Reconstruction, Analysis, and Visualization of
842 Phylogenomic Data. Molecular Biology and Evolution. 2016;33:1635-1638.
- 843 45. Parada AE, Needham DM, Fuhrman JA: Every base matters: assessing small subunit rRNA primers
844 for marine microbiomes with mock communities, time series and global field samples.
845 Environmental Microbiology. 2016;18:1403-1414.
- 846 46. Apprill A, McNally S, Parsons R, Weber L: Minor revision to V4 region SSU rRNA 806R gene primer
847 greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology. 2015;75:129-
848 137.
- 849 47. Ihrmark K, Bodeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J,
850 Brandstrom-Durling M, Clemmensen KE, Lindahl BD: New primers to amplify the fungal ITS2

851 region - evaluation by 454-sequencing of artificial and natural communities. *Fems Microbiology.*
852 *Ecology* 2012;82:666-677.

853 48. Tedersoo L, Bahram M, Polme S, Koljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM,
854 Thu PQ, Suija A, et al: Global diversity and geography of soil fungi. *Science.* 2014;346:1078.

855 49. Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL: Practical innovations for high-
856 throughput amplicon sequencing. *Nature Methods.* 2013;10:999.

857 50. Rognes T, Flouri T, Nichols B, Quince C, Mahe F: VSEARCH: a versatile open source tool for
858 metagenomics. *Peerj.* 2016;4:22.

859 51. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sanchez-
860 Garcia M, Ebersberger I, de Sousa F, et al: Improved software detection and extraction of ITS1 and
861 ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental
862 sequencing data. *Methods in Ecology and Evolution.* 2013;4:914-919.

863 52. Edgar RC: Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.*
864 2010;26:2460-2461.

865 53. Edgar RC: UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature*
866 *Methods.* 2013;10:996.

867 54. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG,
868 Goodrich JK, Gordon JI, et al: QIIME allows analysis of high-throughput community sequencing
869 data. *Nature Methods.* 2010;7:335-336.

870 55. Koljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD,
871 Bengtsson-Palme J, Callaghan TM, et al: Towards a unified paradigm for sequence-based
872 identification of fungi. *Molecular Ecology.* 2013;22:5271-5277.

873 56. Nguyen NH, Song ZW, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG:
874 FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild.
875 *Fungal Ecology.* 2016;20:241-248.

876 57. Hagberg A, Swart P, S Chult D: Exploring network structure, dynamics, and function using
877 NetworkX. *Proceedings of the 7th Python in Science conference.* 2008;11-15.

878 58. Marriott PE, Sibout R, Lapierre C, Fangel JU, Willats WGT, Hofte H, Gomez LD, McQueen-Mason
879 SJ: Range of cell-wall alterations enhance saccharification in *Brachypodium distachyon* mutants.
880 *Proceedings of the National Academy of Sciences of the United States of America.*
881 2014;111:14601-14606.

882 59. Updegraff DM: Semimicro determination of cellulose in biological materials. *Analytical*
883 *Biochemistry.* 1969;32:420.

884 60. Saeman JF: Kinetics of wood saccharification - hydrolysis of cellulose and decomposition of
885 sugars in dilute acid at high temperature. *Industrial and Engineering Chemistry.* 1945;37:43-52.

886 61. Moreira-Vilar FC, Siqueira-Soares RD, Finger-Teixeira A, de Oliveira DM, Ferro AP, da Rocha GJ,
887 Ferrarese MDL, dos Santos WD, Ferrarese O: The Acetyl Bromide Method Is Faster, Simpler and
888 Presents Best Recovery of Lignin in Different Herbaceous Tissues than Klason and Thioglycolic Acid
889 Methods. *Plos One.* 2014;9:7.

890 62. Foster CE, Martin TM, Pauly M: Comprehensive compositional analysis of plant cell walls
891 (lignocellulosic biomass) part I: lignin. *Journal of visualized experiments: JoVE.* 2010.

892 63. Jones E, Oliphant T, Peterson P: SciPy: Open source scientific tools for Python. 2001.

893 64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P,
894 Weiss R, Dubourg VJJomlr: Scikit-learn: Machine learning in Python. 2011;12:2825-2830.

895 65. Alessi AM, Bird SM, Oates NC, Li Y, Dowle AA, Novotny EH, deAzevedo ER, Bennett JP, Polikarpov
896 I, Young JPW, et al: Defining functional diversity for lignocellulose degradation in a microbial
897 community using multi-omics studies. *Biotechnology for Biofuels.* 2018;11:16.

898 66. Valiela I, Teal JM, Allen SD, Vanetten R, Goehring D, Volkman S: Decomposition in salt-marsh
899 ecosystems - the phases and major factors affecting disappearance of above-ground organic-
900 matter. *Journal of Experimental Marine Biology and Ecology.* 1985; 89:29-54.

- 901 67. Curco A, Ibanez C, Day JW, Prat N: Net primary production and decomposition of salt marshes of
902 the Ebre delta (Catalonia, Spain). *Estuaries*. 2002;25:309-324.
- 903 68. Simoes MP, Calado ML, Madeira M, Gazarini LC: Decomposition and nutrient release in
904 halophytes of a Mediterranean salt marsh. *Aquatic Botany*. 2011;94:119-126.
- 905 69. Cragg SM, Friess DA, Gillis LG, Trevathan-Tackett SM, Terrett OM, Watts JEM, Distel DL, Dupree
906 P: Vascular Plants Are Globally Significant Contributors to Marine Carbon Fluxes and Sinks. In
907 *Annual Review of Marine Science*. 2020;12:469-497: *Annual Review of Marine Science*.
- 908 70. Benner R, Fogel ML, Sprague EK: Diagenesis of belowground biomass of *Spartina alterniflora* in
909 salt-marsh sediments. *Limnology and Oceanography*. 1991;36:1358-1374.
- 910 71. Wilson JO, Valiela I, Swain T: Carbohydrate dynamics during decay of litter of *Spartina*
911 *alterniflora*. *Marine Biology*. 1986;92:277-284.
- 912 72. Maccubbin AE, Hodson RE: Mineralization of detrital lignocelluloses by salt-marsh sediment
913 microflora. *Applied and Environmental Microbiology*. 1980;40:735-740.
- 914 73. Nakamura AM, Nascimento AS, Polikarpov I: Structural diversity of carbohydrate esterases.
915 *Biotechnology Research and Innovation*. 2017;1:35-51.
- 916 74. Buanafina MMD: Feruloylation in Grasses: Current and Future Perspectives. *Molecular Plant*.
917 2009;2:861-872.
- 918 75. Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P,
919 Harholt J, Keasling JD, et al: XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to
920 rice xylan. *Proceedings of the National Academy of Sciences of the United States of America*.
921 2012;109:17117-17122.
- 922 76. de Souza WR, Martins PK, Freeman J, Pellny TK, Michaelson LV, Sampaio BL, Vinecky F, Ribeiro
923 AP, da Cunha B, Kobayashi AK, et al: Suppression of a single BAHG gene in *Setaria viridis* causes
924 large, stable decreases in cell wall feruloylation and increases biomass digestibility. *New*
925 *Phytologist*. 2018;218:81-93.
- 926 77. Grantham NJ, Wurman-Rodrich J, Terrett OM, Lyczakowski JJ, Stott K, Iuga D, Simmons TJ,
927 Durand-Tardif M, Brown SP, Dupree R, et al: An even pattern of xylan substitution is critical for
928 interaction with cellulose in plant cell walls. *Nature Plants*. 2017;3:859-865.
- 929 78. Haddad RI, Newell SY, Martens CS, Fallon RD: Early diagenesis of lignin-associated phenolics in
930 the salt-marsh grass *Spartina alterniflora*. *Geochimica Et Cosmochimica Acta*. 1992;56:3751-3764.
- 931 79. Alzarhahi AK, Clark DR, Underwood GJC, Ford H, Cotton TEA, Dumbrell AJ: Are drivers of root-
932 associated fungal community structure context specific? *Isme Journal*. 2019;13:1330-1344.
- 933 80. Wilhelm RC, Singh R, Eltis LD, Mohn WW: Bacterial contributions to delignification and
934 lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope
935 probing. *Isme Journal*. 2019;13:413-429.
- 936 81. Lee SH, Megonigal PJ, Langley AJ, Kang H: Elevated CO₂ and nitrogen addition affect the
937 microbial abundance but not the community structure in salt marsh ecosystem. *Applied Soil*
938 *Ecology*. 2017;117:129-136.
- 939 82. Chaudhary DR, Kim J, Kang H: Influences of Different Halophyte Vegetation on Soil Microbial
940 Community at Temperate Salt Marsh. *Microbial Ecology*. 2018;75:729-738.
- 941 83. Cleary DFR, Coelho F, Oliveira V, Gomes NCM, Polonia ARM: Sediment depth and habitat as
942 predictors of the diversity and composition of sediment bacterial communities in an inter-tidal
943 estuarine environment. *Marine Ecology-an Evolutionary Perspective*. 2017;38:15.
- 944 84. Taylor LE, Henrissat B, Coutinho PM, Ekborg NA, Hutcheson SW, Weiner RA: Complete cellulase
945 system in the marine bacterium *Saccharophagus degradans* strain 2-40(T). *Journal of*
946 *Bacteriology*. 2006;188:3849-3861.
- 947 85. Ekborg NA, Gonzalez JM, Howard MB, Taylor LE, Hutcheson SW, Weiner RM: *Saccharophagus*
948 *degradans* gen. nov., sp nov., a versatile marine degrader of complex polysaccharides.
949 *International Journal of Systematic and Evolutionary Microbiology*. 2005;55:1545-1549.

950 86. Jung YH, Kim HK, Song DS, Choi IG, Yang TH, Lee HJ, Seung D, Kim KH: Feasibility test of utilizing
951 *Saccharophagus degradans* 2-40(T) as the source of crude enzyme for the saccharification of
952 lignocellulose. *Bioprocess and Biosystems Engineering*. 2014;37:707-710.

953 87. Zhang HT, Hutcheson SW: Complex Expression of the Cellulolytic Transcriptome of
954 *Saccharophagus degradans*. *Applied and Environmental Microbiology*. 2011;77:5591-5596.

955 88. Weiner RM, Taylor LE, Henrissat B, Hauser L, Land M, Coutinho PM, Rancurel C, Saunders EH,
956 Longmire AG, Zhang HT, et al: Complete genome sequence of the complex carbohydrate-
957 degrading marine bacterium, *Saccharophagus degradans* strain 2-40(T). *Plos Genetics*. 2008;4:13.

958 89. Fowler CA, Hemsworth GR, Cuskin F, Hart S, Turkenburg J, Gilbert HJ, Walton PH, Davies GJ:
959 Structure and function of a glycoside hydrolase family 8 endoxylanase from *Teredinibacter*
960 *turnerae*. *Acta Crystallographica Section D-Structural Biology*. 2018;74:946-955.

961 90. Fowler CA, Sabbadin F, Ciano L, Hemsworth GR, Elias L, Bruce N, McQueen-Mason S, Davies GJ,
962 Walton PH: Discovery, activity and characterisation of an AA10 lytic polysaccharide oxygenase
963 from the shipworm symbiont *Teredinibacter turnerae*. *Biotechnology for Biofuels*. 2019;12:11.

964 91. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB, Radune D, Toms BS,
965 Henrissat B, Coutinho PM, et al: The Complete Genome of *Teredinibacter turnerae* T7901: An
966 Intracellular Endosymbiont of Marine Wood-Boring Bivalves (Shipworms). *Plos One*. 2009;4:17.

967 92. Ekborg NA, Morrill W, Burgoyne AM, Li L, Distell DL: CelAB, a multifunctional cellulase encoded
968 by *Teredinibacter turnerae* T7902(T), a culturable symbiont isolated from the wood-boring marine
969 bivalve *Lyrodus pedicellatus*. *Applied and Environmental Microbiology*. 2007;73:7785-7788.

970 93. Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, Waterbury J: *Teredinibacter turnerae* gen.
971 nov., sp nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated
972 from the gills of wood-boring molluscs (*Bivalvia* : *Teredinidae*). *International Journal of Systematic
973 and Evolutionary Microbiology*. 2002;52:2261-2269.

974 94. Varnai A, Siika-aho M, Viikari L: Carbohydrate-binding modules (CBMs) revisited: reduced
975 amount of water counterbalances the need for CBMs. *Biotechnology for Biofuels*. 2013;6:11.

976 95. Bowen JL, Morrison HG, Hobbie JE, Sogin ML: Salt marsh sediment diversity: a test of the
977 variability of the rare biosphere among environmental replicates. *Isme Journal*. 2012;6:2014-2023.

978 96. Andrews JE, Samways G, Dennis PF, Maher BA: Origin, abundance and storage of organic carbon
979 and sulphur in the Holocene Humber Estuary: emphasizing human impact on storage changes.
980 *Geological Society, London, Special Publications*. 2000;166:145.

981 97. Lamb AL, Vane CH, Wilson GP, Rees JG, Moss-Hayes VL: Assessing delta C-13 and C/N ratios from
982 organic material in archived cores as Holocene sea level and palaeoenvironmental indicators in
983 the Humber Estuary, UK. *Marine Geology*. 2007;244:109-128.

984 98. Sechovcova H, Kulhava L, Fliegerova K, Trundova M, Morais D, Mrazek J, Kopečný J: Comparison
985 of enzymatic activities and proteomic profiles of *Butyrivibrio fibrisolvens* grown on different
986 carbon sources. *Proteome Science*. 2019;17:12.

987 99. Hunter JD: Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*.
988 2007;9:90-95.

989
990
991
992
993
994
995
996

997

998

999 **Figure 1.** Schematic representation of the integrated 'omics approach undertaken in this study.

1000

1001 **Figure 2.** Temporal changes in lignocellulose composition and the distribution of carbohydrate active
1002 enzyme domains within the meta-exo-proteome. **a;** Lignocellulose composition of remaining *in situ*
1003 *Spartina anglica* biomass. **b;** Rate of compositional change within the lignocellulose displayed as
1004 $\mu\text{g.mg biomass}^{-1}$, the dashed line represents 0 change, *L*: Lignin, *H*: Hemicellulose, *C*: Cellulose. **c;**
1005 Euclidean clustering of the enzyme class profile ($\leq 1e^{-10}$), the 30 most abundant classes are displayed,
1006 *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *PL*: Polysaccharide lyase.
1007 **d;** Euclidean clustering of the carbohydrate binding domain (CBM) profile ($\leq 1e^{-10}$). Error bars (**a**)
1008 represent SE (n=25). Figure plotted with [99].

1009

1010 **Figure 3.** CAZyme producing taxa at family level resolution and their respective CAZyme
1011 contributions. Microbiome and proteomic data is displayed as the mean of n=5 and n=3 respectively.
1012 **a;** Distribution of CAZyme producing lineages with respective CAZyme productivity ($\leq 1e^{-10}$), taxa
1013 below the dashed line were not identified in the community profile. **b;** bacteria profiles elucidated
1014 from 16S rRNA gene sequence homology, each time point is the mean of five biological replicates. **c;**
1015 CAZyme productive bacteria profile, the non-CAZyme productive taxa have been filtered, boxes
1016 display OTU richness, no further filtering was undertaken for these data. *NA*: Not assigned, *t*:
1017 CAZyme producer. Figure plotted with [99].

1018

1019 **Figure 4.** Network associations among meta-exo-proteome CAZyme classes and taxonomic lineages.
1020 Data displayed is the mean of the four time points. Node area is proportional to productivity and

- 1021 abundance for taxa and CAZyme class respectively. Edge color and width is relative to output size.
- 1022 Taxa $< 0.025 \sum \bar{x}$ mol% with ≤ 5 edges and CAZyme classes $< 1.25 \times 10^{-3} \sum \bar{x}$ mol% have been filtered.
- 1023 Glycoside hydrolases (GH) families; blue nodes, carbohydrate esterases (CE) families; red nodes,
- 1024 auxiliary activities (AA) families; orange nodes, polysaccharide lyases (PL) families; yellow nodes,
- 1025 carbohydrate binding domains (CBM) families; purple nodes. *NA*: Not assigned.

1 Supplementary Information for:

2

3 Mechanistic strategies of microbial communities regulating lignocellulose deconstruction in a UK salt
4 marsh

5

6 **Authors**

7 Daniel R. Leadbeater^{1§}, Nicola C. Oates¹, Joseph P. Bennett¹, Yi Li¹, Adam A. Dowle², Joe D. Taylor⁴,
8 Juliana Sanchez Alponi¹, Alexander T. Setchfield¹, Anna M. Alessi¹, Thorunn Helgason³, Simon J.
9 McQueen-Mason^{1§}, Neil C. Bruce^{1§}

10 ¹ Centre for Novel Agricultural Products, Department of Biology, University of York, York, YO10 5DD,
11 UK.

12 ² Bioscience Technology Facility, Department of Biology, University of York, York, YO10 5DD, UK.

13 ³ Department of Biology, University of York, York, YO10 5DD, UK.

14 ⁴ School of Chemistry and Biosciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP,
15 UK.

16 [§] Corresponding authors: N.C.B. (email: neil.bruce@york.ac.uk), S.J.M.-M. (email:
17 simon.mcqueenmason@york.ac.uk) or D.R.L. (email: daniel.leadbeater@york.ac.uk).

18

19

20

21

22

23

24

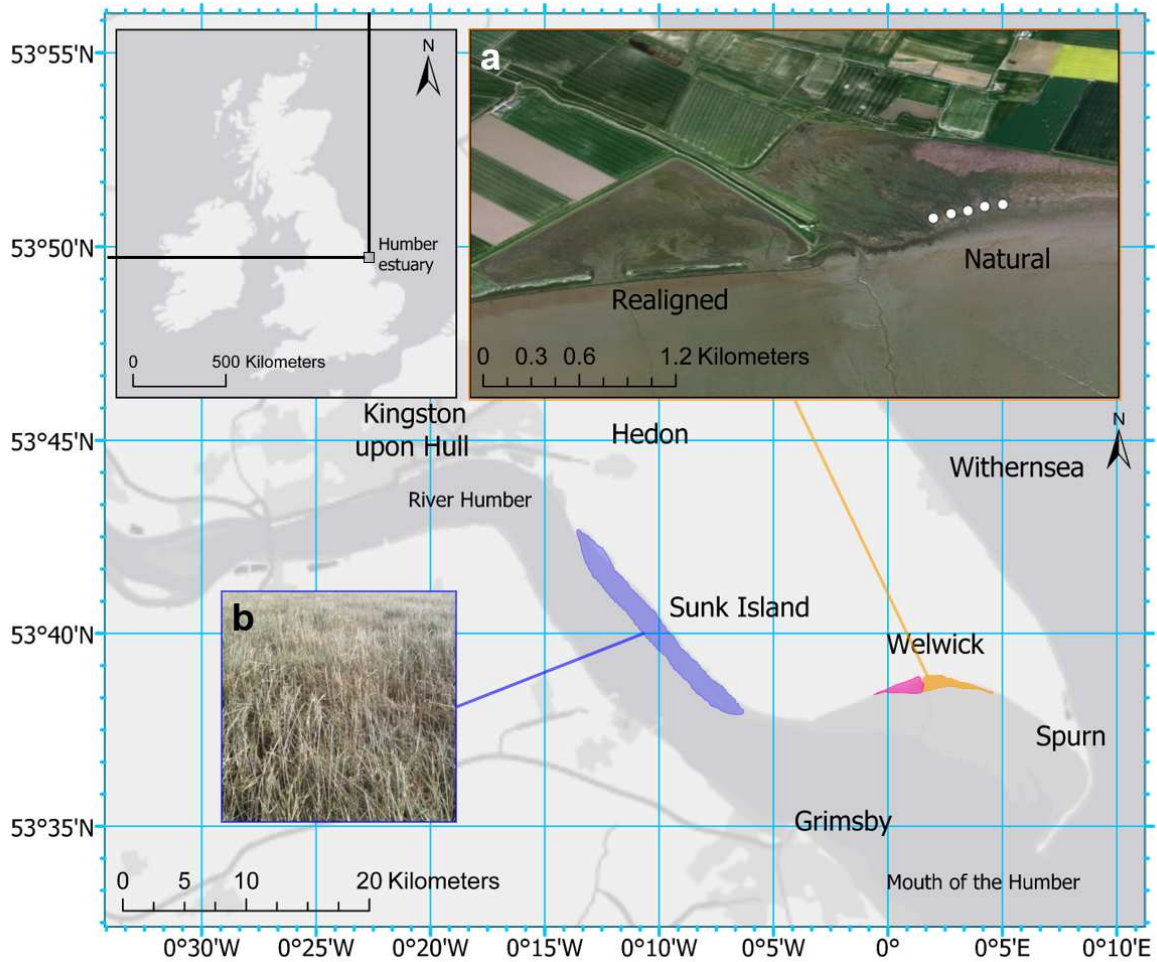
25

26

27

28

29



30

31 **Figure S1.** Experiment location. The Humber estuary (Hull, UK). The North bank of the Humber
 32 estuary, Welwick salt marsh (inlet a) is highlighted in orange and the cage locations along the
 33 transect in white within the inlet. Cherry Cobb Sands is highlighted in blue with the location of the
 34 collected *Spartina anglica* (inlet b).

35

36

37

38

39

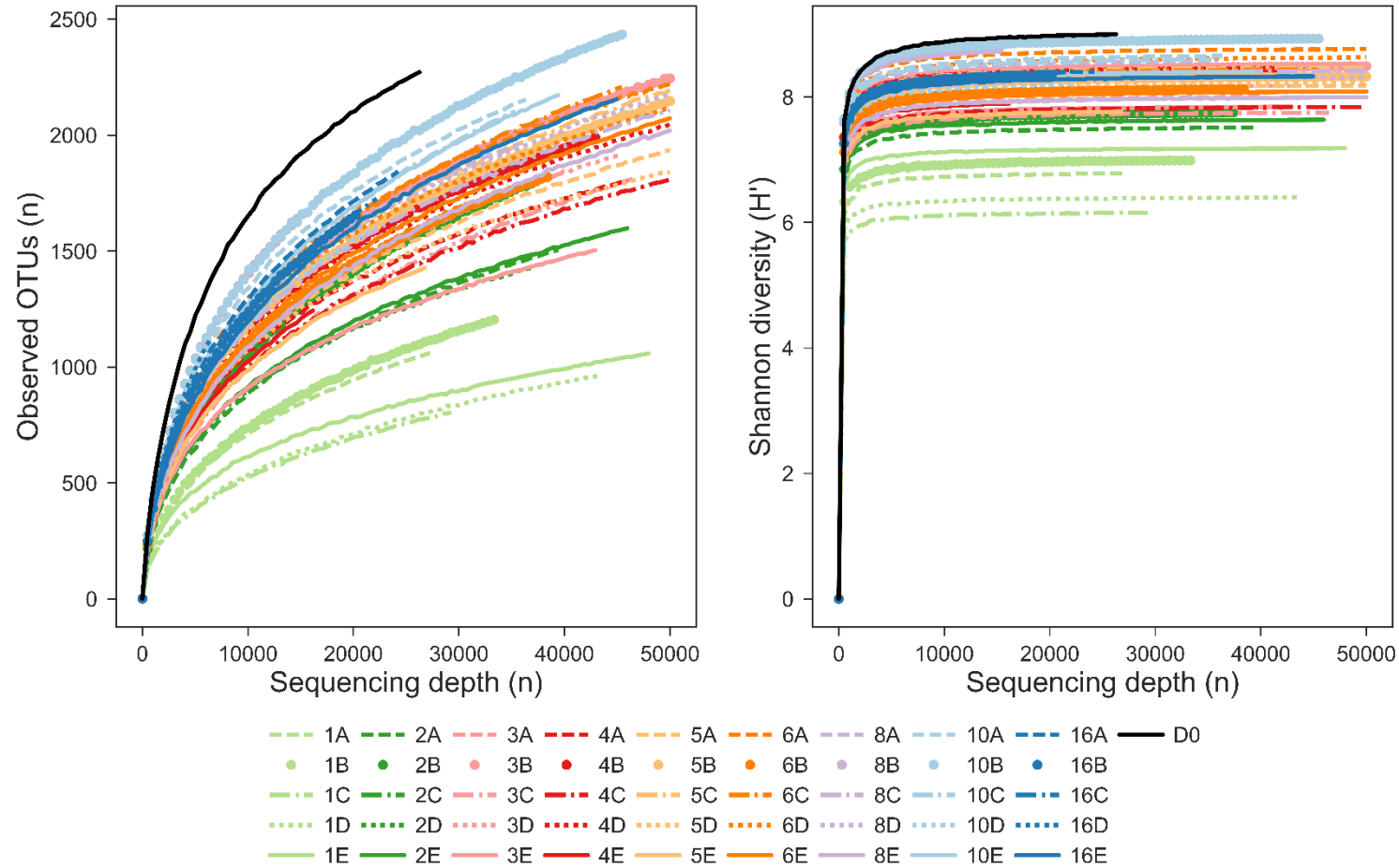
40

41

42

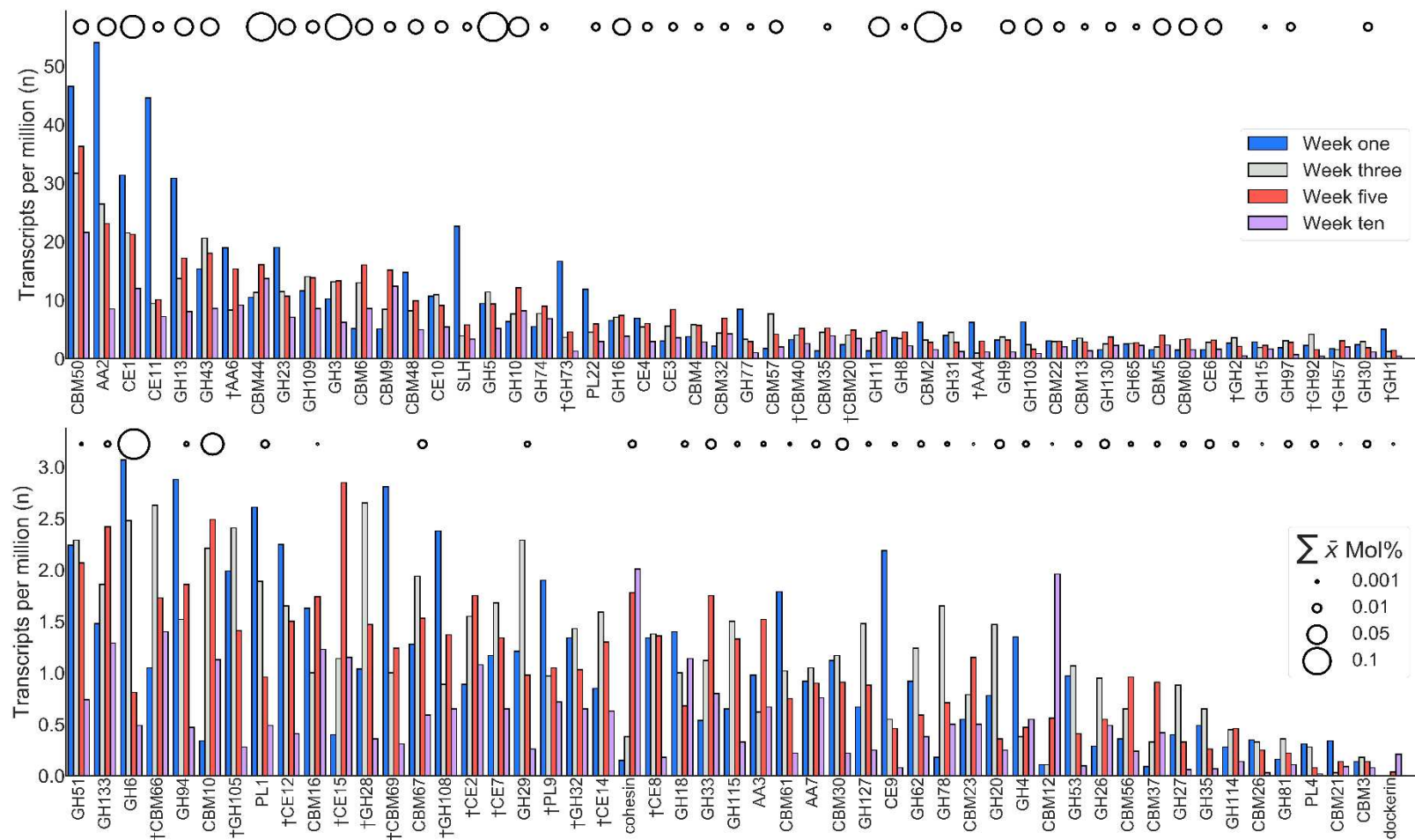
43

44



45

46 **Figure S2.** Coverage estimates for each 16S rRNA amplicon library for all biological replicates across each time point. **a;** Rarefaction curves identified by
 47 number of observed operational taxonomic units (OTUs) and sequencing depth. **b;** Shannon diversity index collector curves as a function of sequencing
 48 depth.

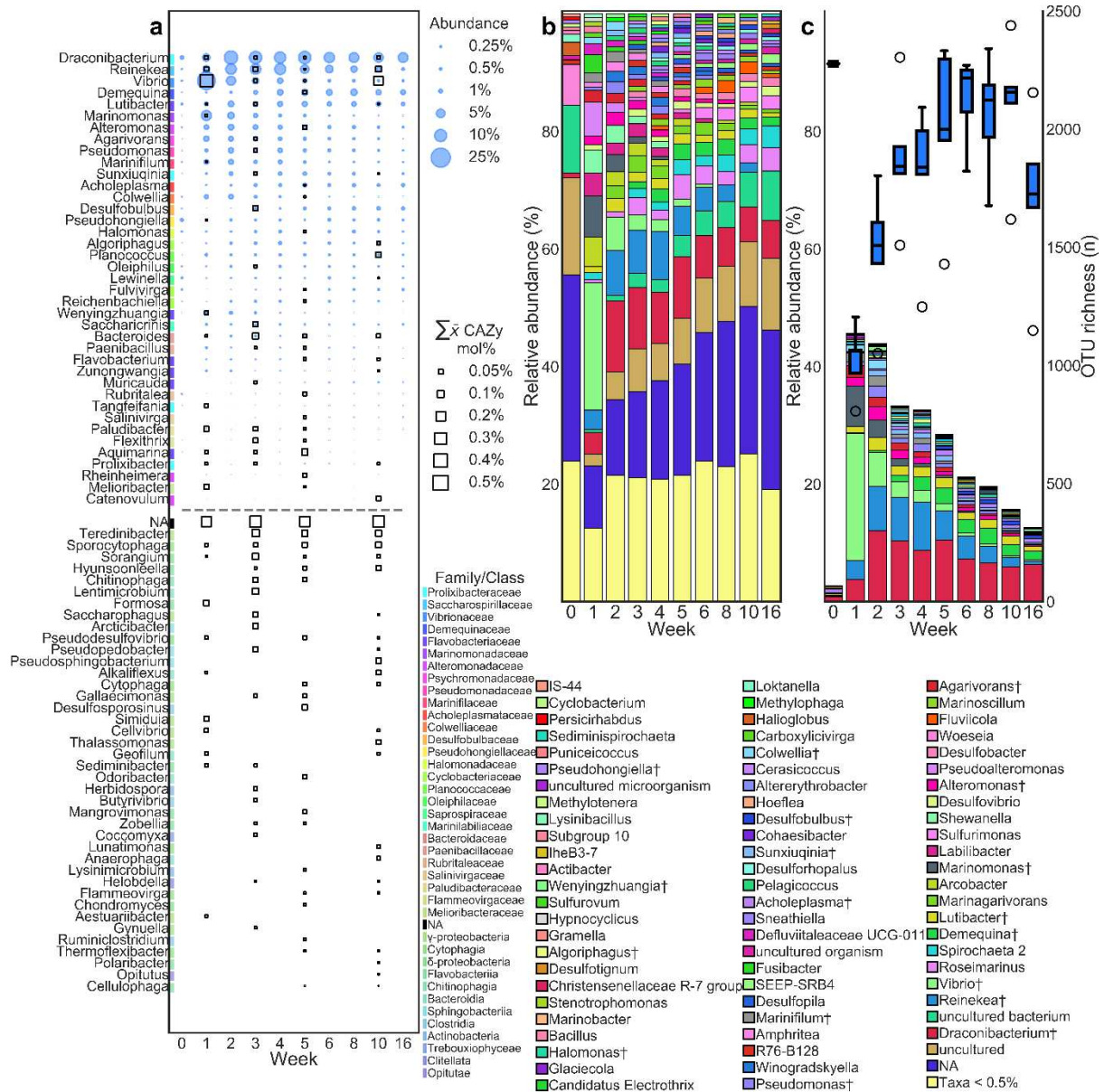


49

50 **Figure S3.** CAZyme families identified within the metatranscriptomic databases presented as transcripts per million (TPM). Glycosyl transferase families
 51 have been filtered and only annotations with expect values $\leq 1e^{-5}$ and TPM values ≥ 1 have been included for analysis. † delineates the CAZyme family was
 52 not identified within the meta-exo-proteome. Marker size is proportional to the mean sum of the molar percentage across the exo-meta-proteome for each
 53 week. *GH*: Glycoside hydrolase, *CE*: Carbohydrate esterase, *AA*: Auxiliary activity, *PL*: Polysaccharide lyase, *CBM*: Carbohydrate binding domain.
 54 Transcriptomic data includes both intracellular and extracellular representatives of the respective CAZyme families.

55

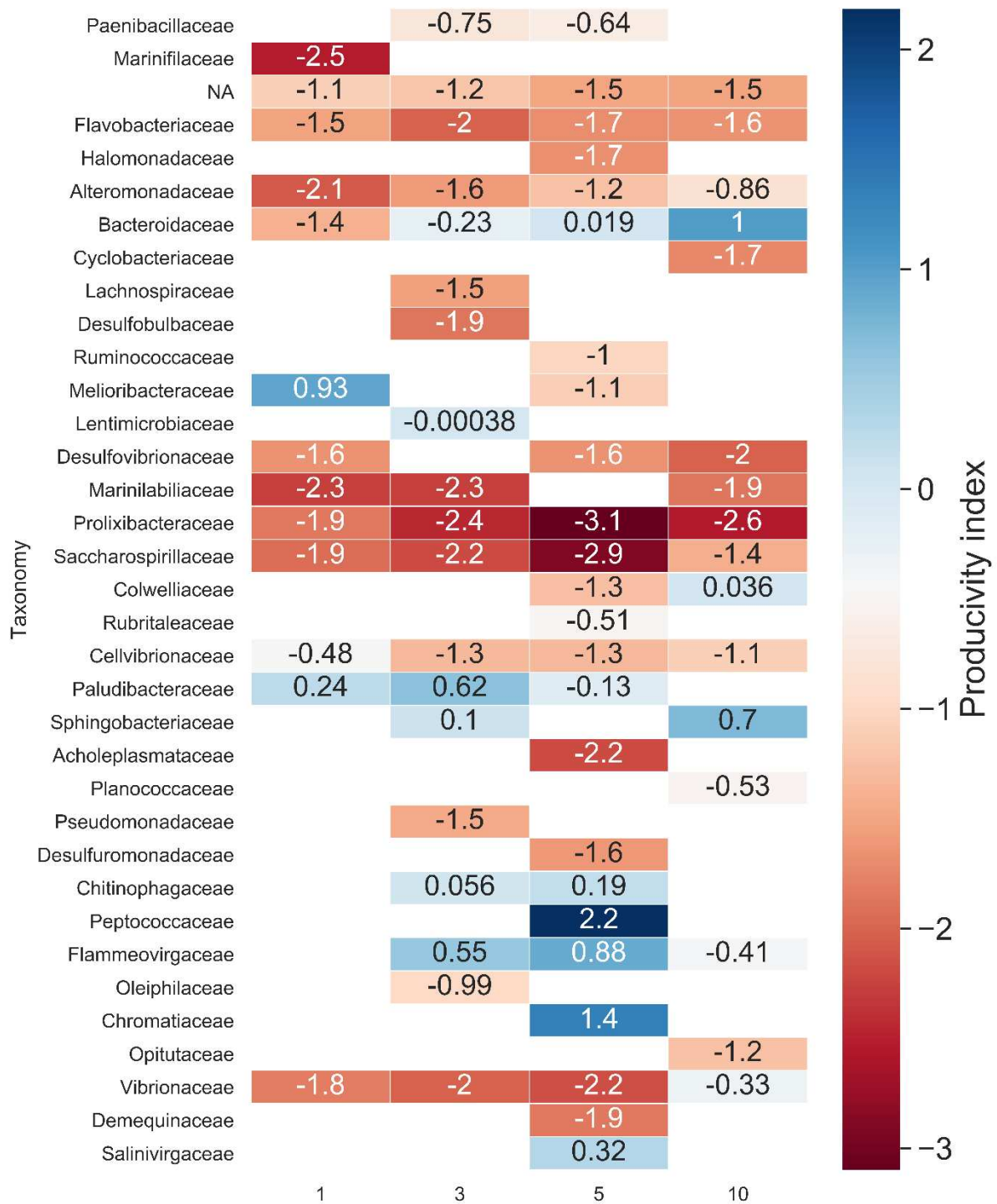
56



57

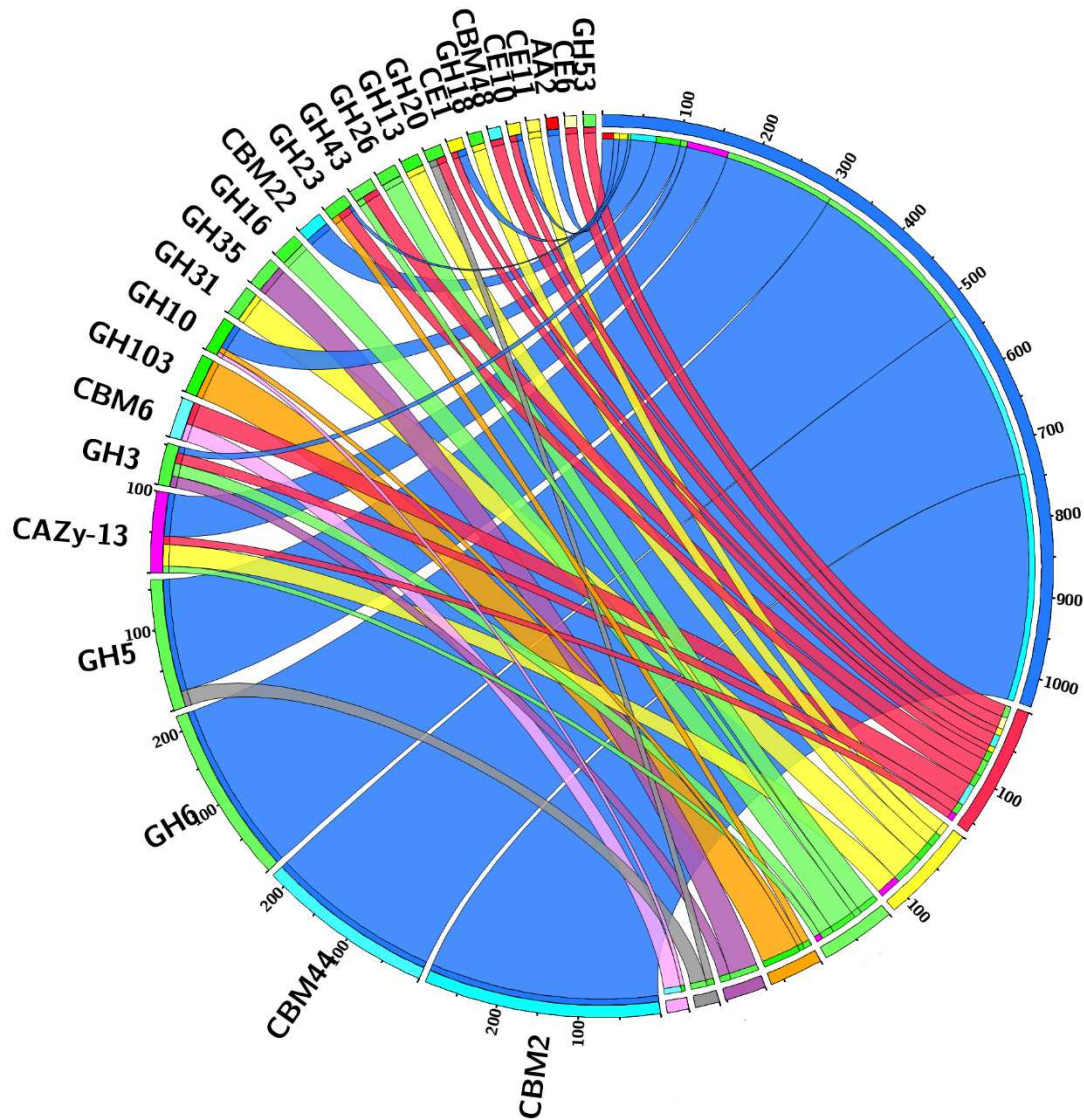
58 **Figure S4.** CAZyme producing genera and their respective CAZyme contributions. Microbiome and
 59 proteomic data is displayed as the mean of n=5 and n=3 respectively. **a;** Distribution of CAZyme
 60 producing lineages with respective CAZyme productivity ($\leq 1e^{-10}$), taxa below the dashed line were
 61 not identified in the community profile, ordered by descending abundance by total temporal sum. **b;**
 62 Bacteria profiles elucidated from 16S rRNA sequence homology, each time point is the mean of five
 63 biological replicates. **c;** CAZyme productive bacteria profile, the non-CAZyme productive taxa have
 64 been filtered, boxes display OTU richness, no further filtering was undertaken for these data. NA:
 65 Not assigned, †: CAZyme producer.

66

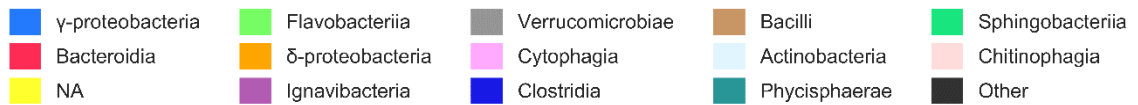


67

68 **Figure S5.** Productivity index for CAZyme producing taxa at family level resolution. A higher index
 69 indicates disproportionately more CAZyme produced per unit abundance. Values displayed as $\log_{10}(\Sigma$
 70 \bar{x} mol%/Relative abundance).



71



72

73

74

75

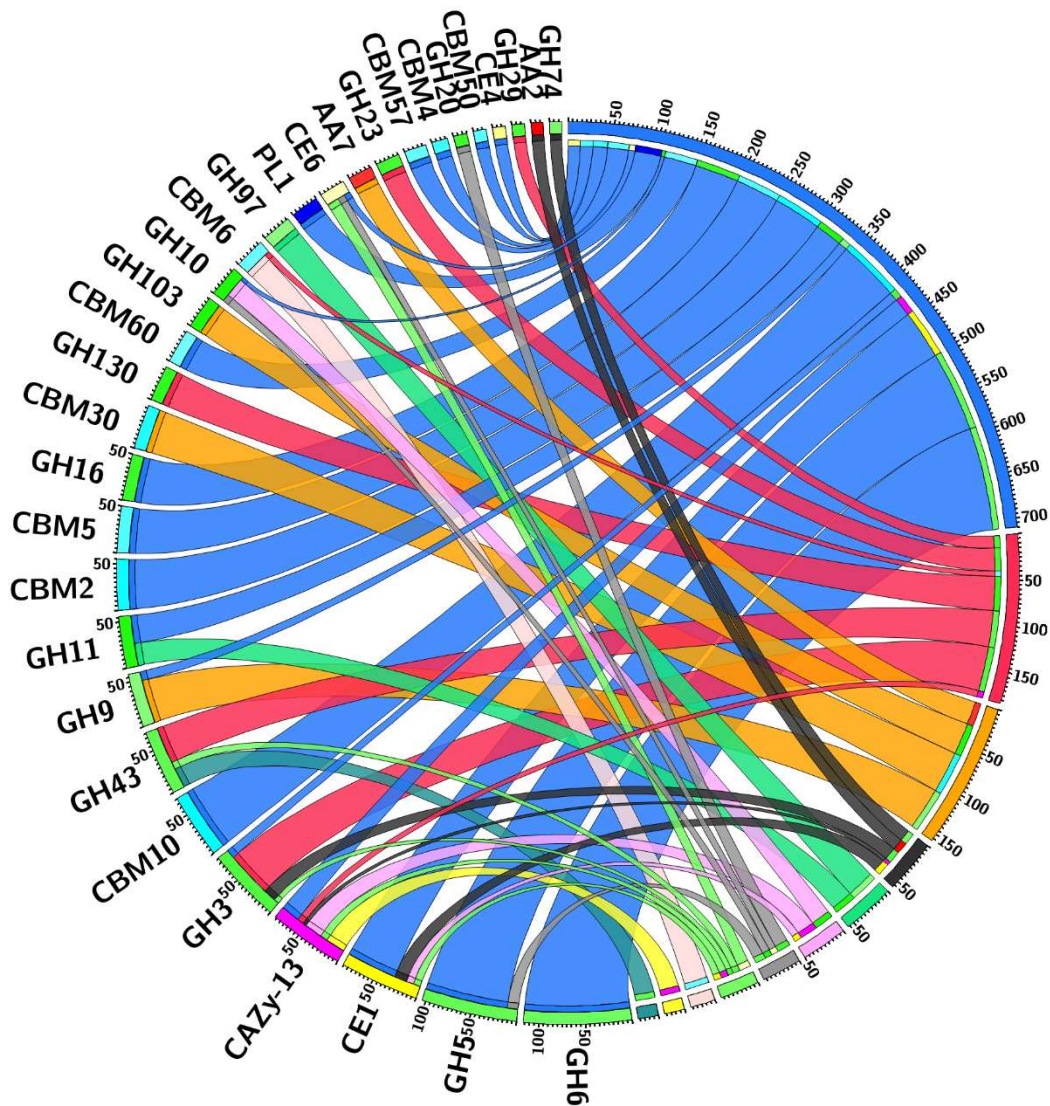
76

77

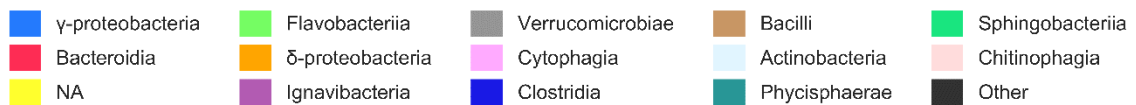
78

79

Figure S6. Phylogenetic distribution of CAZyme classes at class resolution for week one. Absolute values of segments are the sum molar percentage $1e^3$ (1000 = 1%). Ribbon size is relative to the contribution to or of the taxa/ CAZyme and are only interspecifically comparable. Taxa producing < 1% of the total CAZyme and CAZyme classes < 0.85% of the total have been filtered, no further proteomic filtering criteria was applied to this data. Chord diagrams were generated using the Circos software package [1]. NA: Not assigned.



80



81

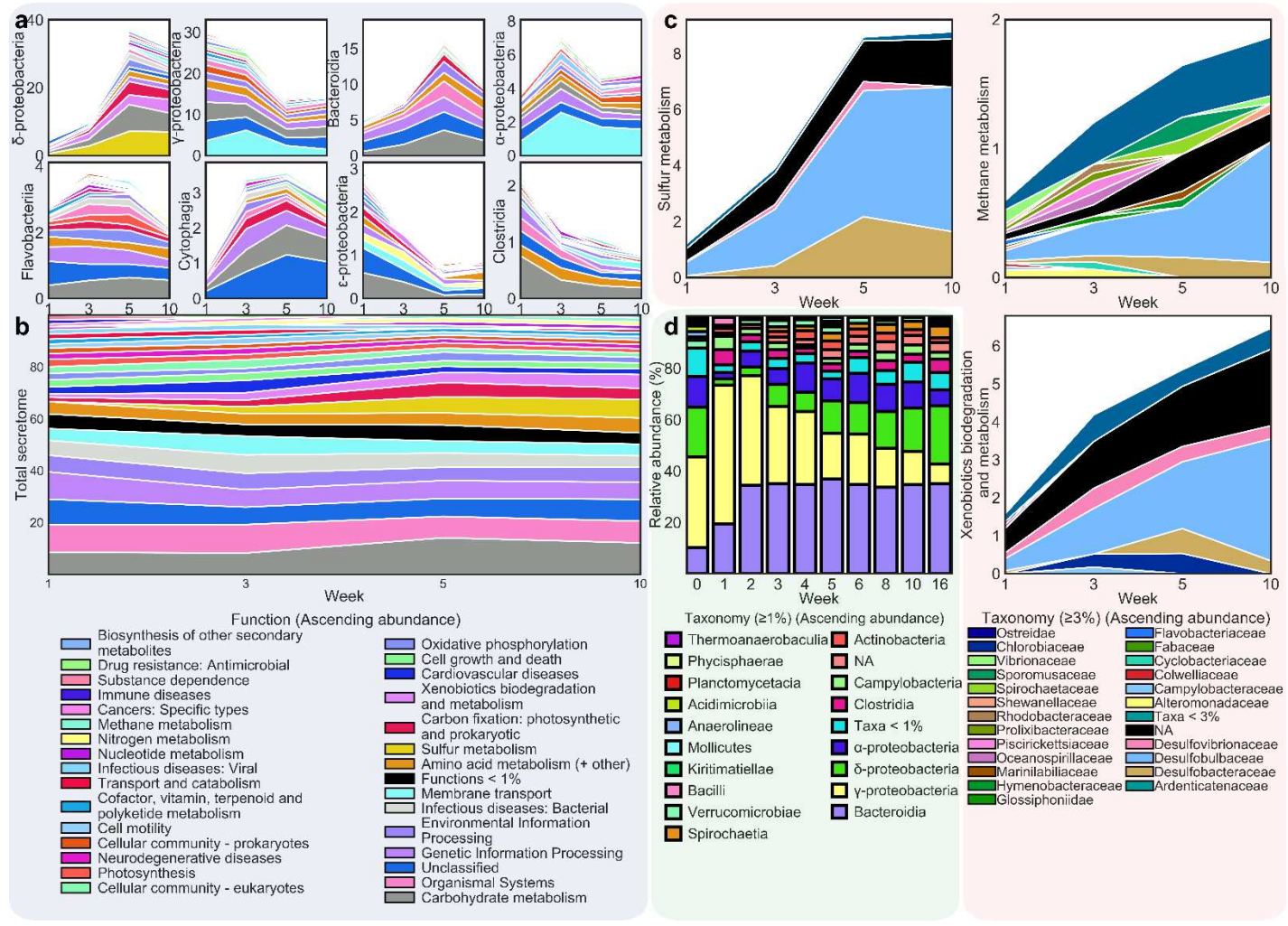
82 **Figure S7.** Phylogenetic distribution of CAZyme classes at class resolution for week three. Absolute
 83 values of segments are the sum molar percentage $1e^3$ (1000 = 1%). Ribbon size is relative to the
 84 contribution to or of the taxa/ CAZyme and are only interspecifically comparable. Taxa producing <
 85 1% of the total CAZyme and CAZyme classes < 0.85% of the total have been filtered, no further
 86 proteomic filtering criteria was applied to this data. Chord diagrams were generated using the Circos
 87 software package [1]. NA: Not assigned.

88

89

90

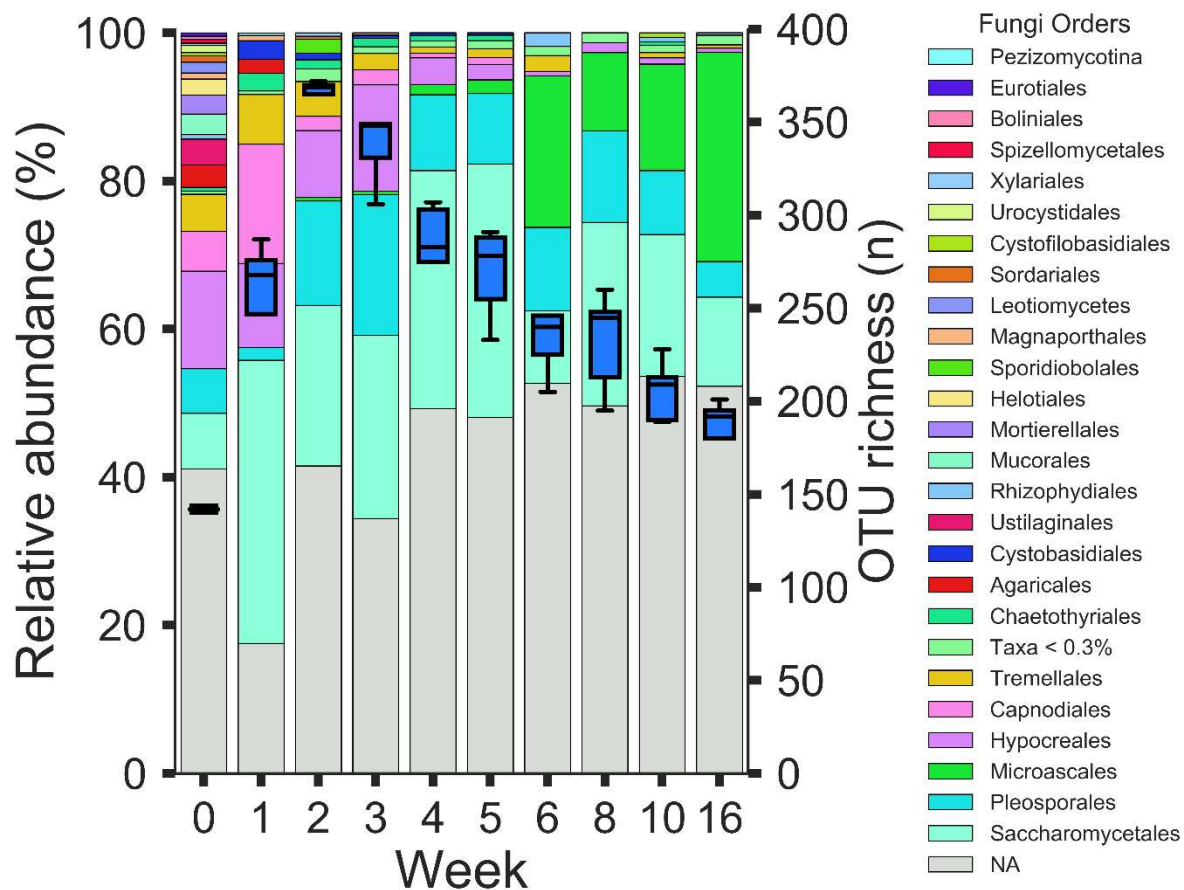
91



113

114 **Figure S10.** Functional classification of proteins within the metasecretome. Unclassified orthologs are not displayed. One to many mapping applied. **a**;
 115 Temporal distribution of functional KEGG orthologs [2] attributable to taxonomic classes ($\sum \bar{x}$ mol%). **b**; Functional assignment of the total metasecretome
 116 ($\sum \bar{x}$ mol%). **c**; Taxonomic distribution of major functional pathways identified from the metasecretome ($\sum \bar{x}$ mol%). **d**; Bacteria profile. NA: Not assigned.

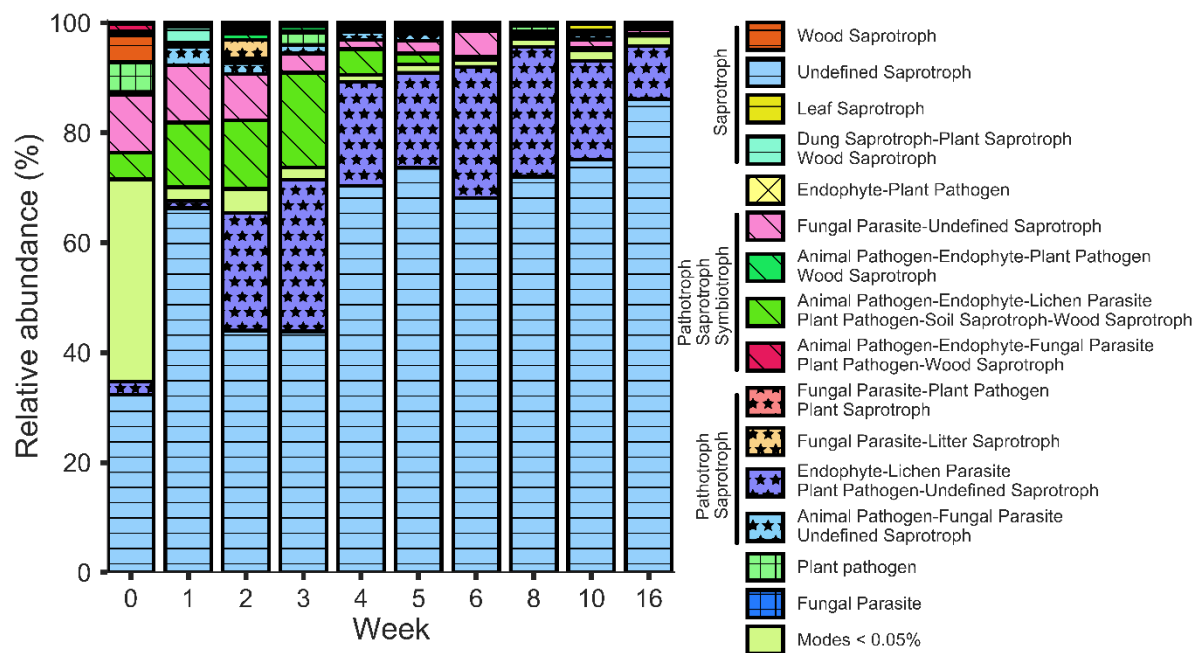
117
118
119
120
121
122



123
124
125
126
127
128
129
130
131
132

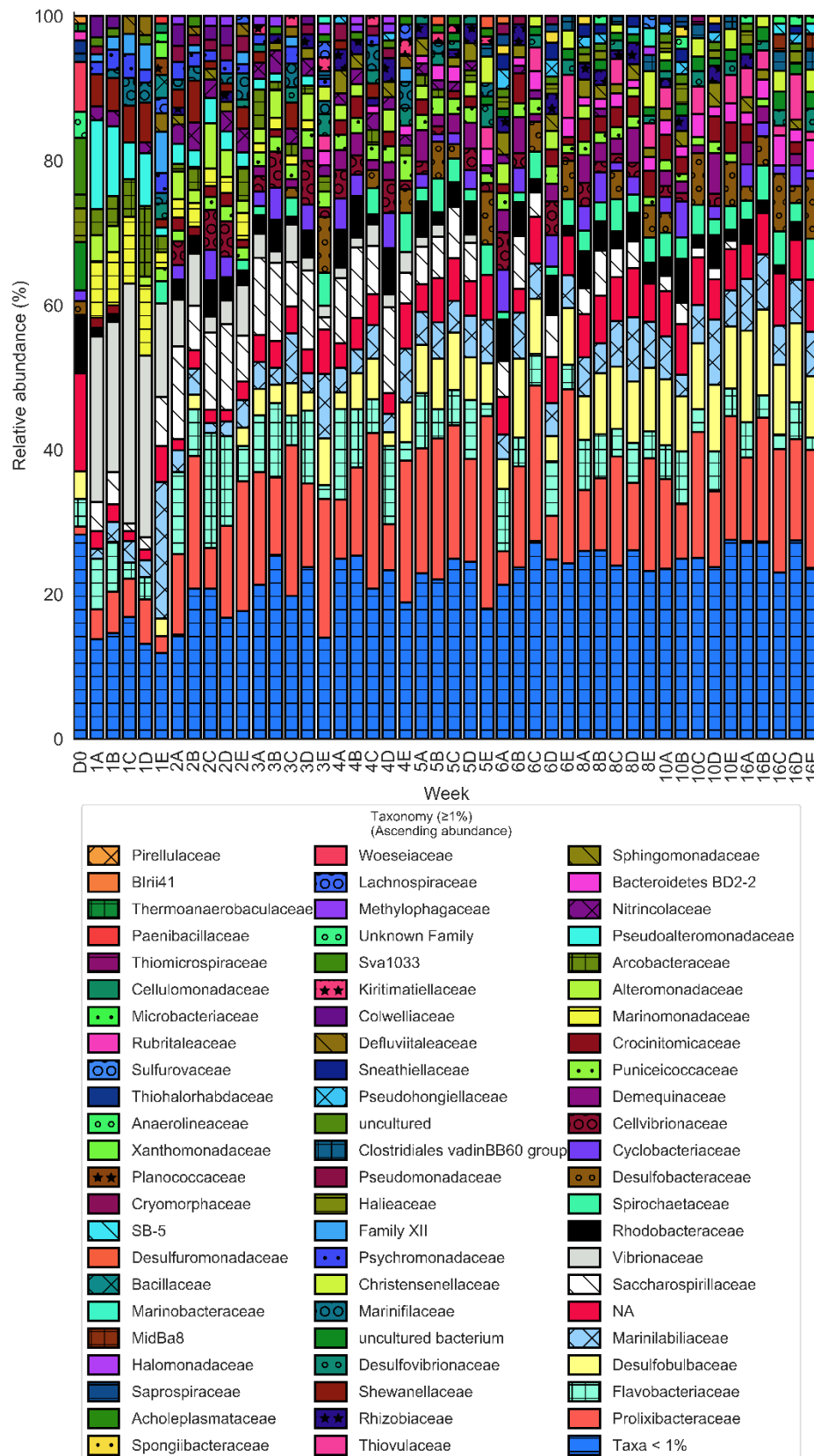
Figure S11. Fungal profiles and OTU richness elucidated from internal transcribed spacer region 2 amplicon sequencing across the 16-week time course. Data is displayed as the mean of n=5.

133
 134
 135
 136
 137
 138
 139
 140



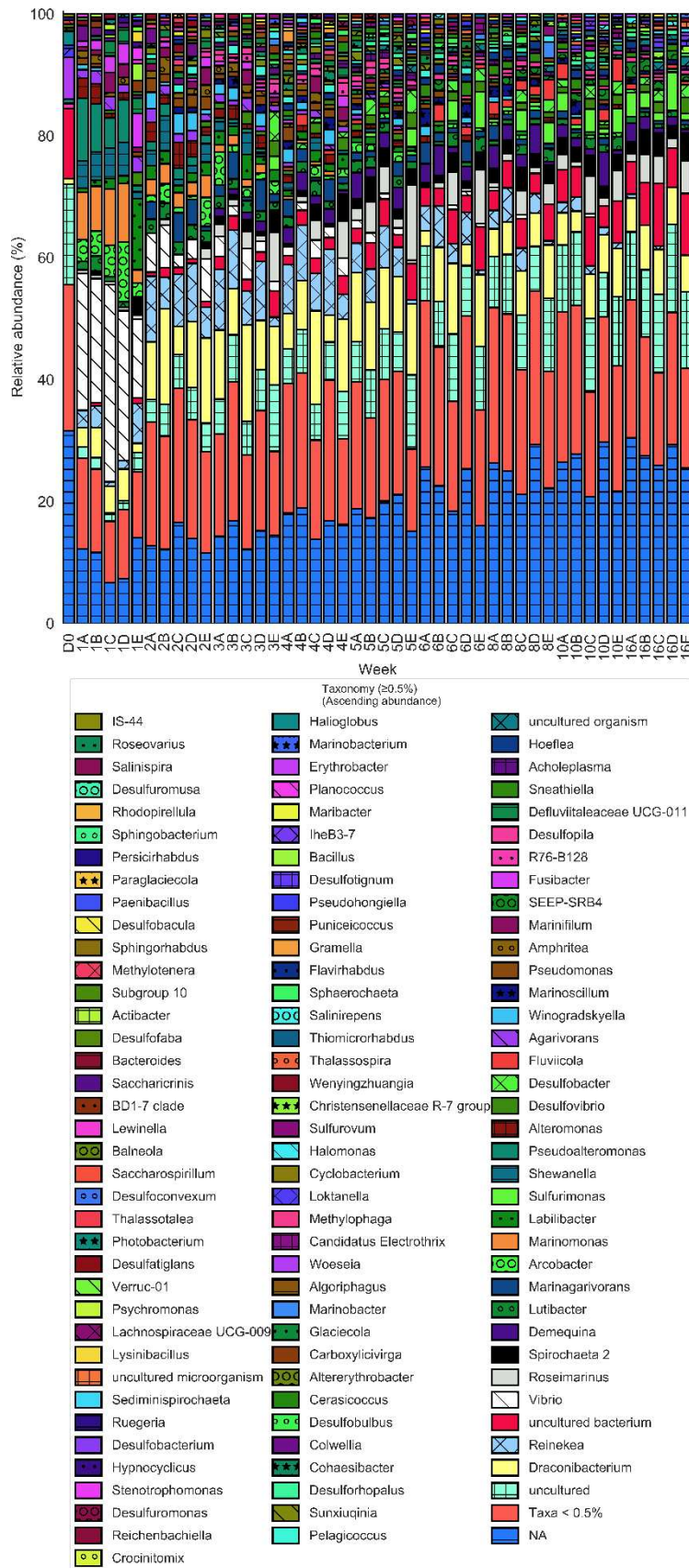
141
 142
 143
 144
 145

Figure S12. Nutrient acquisition strategy of fungi profile across the 16-week time course. Data is displayed as the mean of n=5. Guild and trophic mode are displayed.



147

148 **Figure S13.** Bacteria profiles elucidated from 16S rRNA sequence homology, all biological replicates
 149 are displayed at family level.



150

151 **Figure S14.** Bacteria profiles elucidated from 16S rRNA sequence homology, all biological replicates
 152 are displayed at genus level.

153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186

Table S1. Transect position of the five biological replicates in Welwick salt marsh. *OS GR*: ordnance survey grid reference.

Cage	OS GR	Coordinates	Latitude/Longitude	Elevation (m)
A	TA 33850 18626	53.647022, 0.023465794	53°38'49.28"N/ 0° 1'24.48"E	2
B	TA 33915 18656	53.647271,0.024455525	53°38'50.18"N/ 0° 1'28.04"E	2.2
C	TA 33977 18676	53.647439, 0.025407313	53°38'50.78"N/ 0° 1'31.47"E	2.7
D	TA 34038 18704	53.647678,0.026343800	53°38'51.64"N/ 0° 1'34.84"	3
E	TA 34105 18719	53.647792, 0.027360971	53°38'52.05"N/ 0° 1'38.50"E	3

† Coordinate accuracy ±4m

187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211

Table S2. Sequence reads throughout RNA sequence processing and assembly.

	Week 1	Week 3	Week 5	Week 10	Total
Raw reads	82 966 972	99 319 322	95 318 915	105 517 252	383 122 461
Percentage reads rRNA	10.71%	1.58%	2.74%	2.04%	-
Paired end reads (- rRNA, -duplicates)	68 307 441	91 948 286	85 685 871	93 709 355	339 650 953
Paired end reads (-post QC)	68 302 150	91 926 491	85 676 818	93 700 742	339 606 201
Contigs (Trinity)	4 100 865	9 720 697	7 501 904	8 615 202	29 938 668
Genes (Trinity)	3 869 004	9 132 509	7 071 997	8 153 028	28 226 538
Median contig length (bp)	187	195	185	201	-
Average contig length (bp)	268.28	247.46	255.67	249.49	-

212
 213
 214
 215
 216
 217
 218

Table S3. Commands for the replication of the 16S rRNA amplicon database processing pipeline. Emboldened characters are specific input or output files. Only steps utilising third party software is shown, custom steps are not shown. A1-A12 represent input or output files.

Step	Function	Command
1	Merge	vsearch --fastq_mergepairs X1_001.fastq.gz --reverse X2_001.fastq.gz --fastqout X_16Smerged
2A	Trim adapter	cutadapt -o nexterearemoved/ X -a CTGTCTCTTATACACATCTGACGCTGCCGACGA X
2B	Trim 13N	cutadapt --cut 13 -o X X
3	Fastq split	convert_fastaqual_fastq.py -c fastq_to_fastaqual -f X split_libraries.py -f X --max_ambig 0 -r -k -B -H 10 -M 2 -e 2 -b 7 -o X -m MAP
4	Demultiplex	
5	Trim primer	cutadapt --cut 19 --cut -20 -o X X
6	Concatenate	cat *.fna > catfile.fna
7	Format header	Format headers to >barcode label=sample_id;sequence_number_integer - custom
8	Global trim	usearch_v9 -fastx_truncate A1 -truncLen 250 -fastaout A2 usearch_v7 --derep_fulllength A2 --output A3 --log=log --sizeout --minuniquesize 2
9	Dereplicate	
10	Sort by size	Usearch_v7 -sortbysize A3 -output A4 -minsize2
11	Cluster	usearch_v9 -cluster_otus A4 -cat16S.fna -otus A5 -minsize 2
12	Relabel	fasta_number.py A5 OTU_> A6
13	Map OTUs	Usearch_v7 -usearch_global A1 -db A6 -strand plus -id 0.95 -uc A7
14	Assign taxonomy	Assign_taxonomy.py -i A6 -o A8 --similarity 0.9 -r ref_set.fna -t ref_set.txt
15	OTU table (.txt)	uc2otutab.py A7 > A9.txt
16*	.biom	Biom convert -i A9 -o A10 --table-type="OTU table" --to-hdf5
17*	.tsv	Biom convert -i A10 -o A11 --to-tsv

*Optional

219
 220
 221
 222
 223
 224
 225
 226

227

228 **References**

229

230 1. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: Circos: An
231 information aesthetic for comparative genomics. *Genome Research*. 2009; 19:1639-1645.

232 2. Kanehisa M, Sato Y, Morishima K: BlastKOALA and GhostKOALA: KEGG Tools for Functional
233 Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*.
234 2016;428:726-731.

235