Low-abundance populations distinguish microbiome performance in plant cell wall deconstruction

Lauren M. Tom[1,2†], Martina Aulitto[1,2†], Yu-Wei Wu[3], Kai Deng[1,2], Yu Gao[1,4], Naijia Xiao[5],
Beatrice Garcia Rodriguez[6], Clifford Louime[6], Trent R. Northen[1,4], Aymerick Eudes[1,4], Jenny C. Mortimer[1,4,7],
Paul D. Adams[1,8,9], Henrik V. Scheller[1,4,10], Blake A. Simmons[1,2], Javier A. Ceja-Navarro[1,2,11,12*] and
Steven W. Singer[1,2*]

Abstract

Background: Plant cell walls are interwoven structures recalcitrant to degradation. Native and adapted microbiomes can be particularly effective at plant cell wall deconstruction. Although most understanding of biological cell wall deconstruction has been obtained from isolates, cultivated microbiomes that break down cell walls have emerged as new sources for biotechnologically relevant microbes and enzymes. These microbiomes provide a unique resource to identify key interacting functional microbial groups and to guide the design of specialized synthetic microbial communities.

Results: To establish a system assessing comparative microbiome performance, parallel microbiomes were cultivated on sorghum (*Sorghum bicolor* L. Moench) from compost inocula. Biomass loss and biochemical assays indicated that these microbiomes diverged in their ability to deconstruct biomass. Network reconstructions from gene expression dynamics identified key groups and potential interactions within the adapted sorghum-degrading communities, including *Actinotalea*, *Filomicrobium*, and *Gemmatimonadetes* populations. Functional analysis demonstrated that the microbiomes proceeded through successive stages that are linked to enzymes that deconstruct plant cell wall polymers. The combination of network and functional analysis highlighted the importance of cellulose-degrading *Actinobacteria* in differentiating the performance of these microbiomes.

Conclusions: The two-tier cultivation of compost-derived microbiomes on sorghum led to the establishment of microbiomes for which community structure and performance could be assessed. The work reinforces the observation that subtle differences in community composition and the genomic content of strains may lead to significant differences in community performance.

Keywords: Biomass deconstruction, Lignocellulose degradation, Microbiome, Transcriptomic network

†Lauren M. Tom and Martina Aulitto contributed equally to this work.

*Correspondence: JCNavarro@lbl.gov; SWSinger@lbl.gov

[2] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[12] Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA
Full list of author information is available at the end of the article

Introduction

Plant cell walls are complex structures that primarily contain the polysaccharide polymers cellulose, hemicellulose, and pectin as well as the aromatic polymer lignin [1]. The primary cell wall of grasses, such as sorghum (*Sorghum bicolor* L. Moench), is a thin layer consisting primarily of cellulose and the hemicellulose xylan, and a

small amount of pectin. The thicker secondary cell wall, deposited after plant cell growth ceases, primarily contains cellulose, lignin, and hemicellulose. Chemical and biological deconstruction of plant cell walls to release the sugars and phenolics in biomass is of great current interest for their subsequent conversion to biofuels and bio-based chemicals [2–4]. For biological deconstruction, microorganisms, including filamentous fungi, bacteria, and protists, employ an armamentarium of enzymes that systematically deconstruct the plant cell wall [5–7]. These include hydrolytic and oxidative enzymes that deconstruct the polysaccharides and radical-based oxidative enzymes that deconstruct lignin [8–10].

Though most understanding of biological cell wall deconstruction has been obtained from isolates, microbiomes that break down cell walls have emerged as new sources of microbes and enzymes [11–15]. These microbiomes feature successive structures that are linked to the mechanism of depolymerization in the cell wall [16]. Microbiomes that degrade plant cell walls are readily cultivated from inocula rich in deconstructive microbes, like compost and rumen [17, 18]. These cultivations yield microbiomes formed by different microorganisms interacting with each other through processes such as competition or cross-feeding of resources, division of labor, and niche construction [19]. In the case of biomass deconstructing microbiomes, these interactions result in a community with reproducible composition and dynamics that can be traced allowing the enzymatic activity of individual microbial populations to be linked to the plant polymers they deconstruct [17]. However, the development of parallel consortia from heterogeneous inocula, like an environmental sample, can lead to variations, often referred to as founder effects, in microbiome structure that may influence microbiome performance [20, 21]. Therefore, the cultivation of parallel consortia from environmental samples is a promising strategy to link the structure and dynamics of biomass-deconstructing microbiomes to the identification of key contributors to lignocellulose decomposition and provide insight into the genes correlated with lignocellulolytic activity. Identifying key functional groups among the cultivated consortia can then guide the design of synthetic communities for biomass-to-biofuel/bioproduct production.

Here, compost-derived parallel microbiomes with different community structures were cultivated with sorghum biomass as the sole carbon source (Fig. 1). The performance of these distinct microbiomes was investigated by comparing community composition dynamics and activity on wild-type forage sorghum and the double sorghum mutant *bmr6 x bmr12* characterized by having the last two steps of monolignol synthesis impaired [22]. The performance of the identified microbial communities was linked to specific populations by network and functional analyses of metagenomics and time-resolved metatranscriptomics. Our results show that the microbiomes proceeded through successive stages linked to the expression of the genes encoding enzymes likely involved in the deconstruction of plant cell wall polymers. It was also observed that low-abundance microbial populations differentiate the cell wall-degrading capacity of the adapted communities.
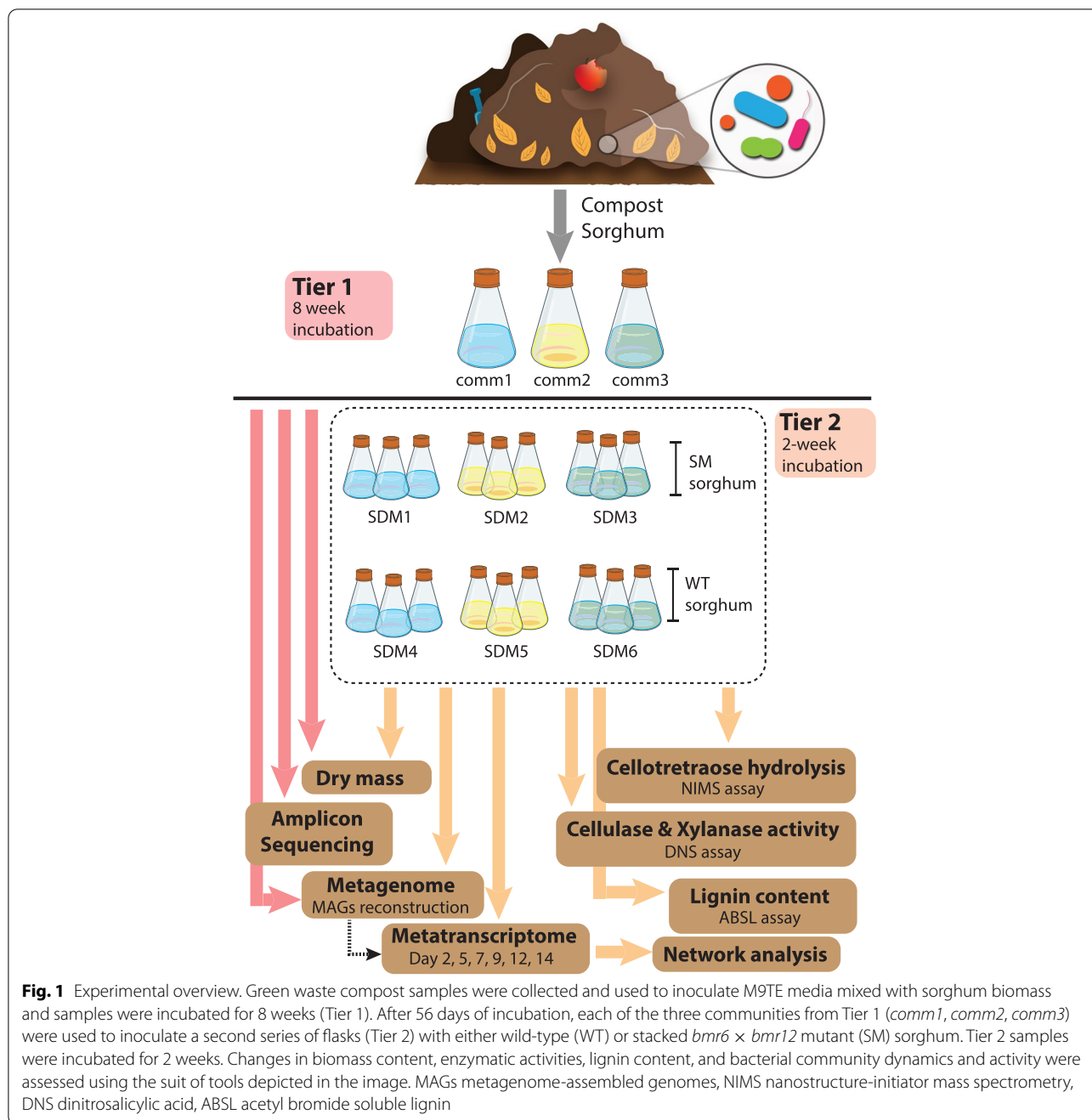
Materials and methods

Sample collection and biomass preparation

Green waste compost was collected from the City of Berkeley (https://www.cityofberkeley.info/freecompost/) and transported to the lab at room temperature. The compost was sieved and stored at 4℃ prior to use. Untreated sorghum biomass samples were provided by Dr. Scott Sattler (USDA-ARS, Lincoln, NE). Wild-type (WT) and *bmr6 x bmr12* sorghum (here referred to as mutant) plants were grown in the field and harvested at physiological maturity, and their panicles were carefully removed as previously described [22]. Samples consisting of a mixture of stems and leaves were ground using a Model 4 Wiley Mill equipped with a 2-mm mesh (Thomas Scientific, Swedesboro, NJ). The forage sorghum samples were washed, autoclaved, and dried in an oven at 50℃ prior to their use in subsequent experiments.

Enrichment/priming (Tier 1)

Green waste compost (0.1 g), 50 mL M9 media supplemented with vitamins and 50 mM of 2-(N-morpholino)ethanesulfonic acid (MES) at a final pH of 6.5 (M9TE media) [23], and 0.5 g of sorghum were inoculated in 250-mL baffled Erlenmeyer flasks (Fig. 1). Three parallel incubations, along with a negative control without inoculant, were incubated at 50°C at 200 rpm and adjusted for evaporation using filter-sterilized deionized water every 2–3 days. Passages were conducted every 2 weeks (day 14, 28, 42, and 56) by transferring 2 mL of culture to a new set of flasks. At the end of each passage, pH was measured and 500 μL aliquots were collected and centrifuged to separate pellet and supernatant fractions. DNA was extracted from the pellet fraction and sent for 16S rRNA gene and metagenomic sequencing. Additionally, for the final passage (day 56), 3,5-dinitrosalicylic acid (DNS) assays [24] and nanostructure-initiator mass spectrometry (NIMS) [25] were performed on the supernatant fraction and the remaining material was filtered using Miracloth (Millipore Sigma, Burlington, MA, USA) and dried at 50°C to determine the biomass dry mass. The DNS assay was performed as described in detail elsewhere [26]. The supernatant was added

Fig. 1 Experimental overview. Green waste compost samples were collected and used to inoculate M9TE media mixed with sorghum biomass and samples were incubated for 8 weeks (Tier 1). After 56 days of incubation, each of the three communities from Tier 1 (*comm1*, *comm2*, *comm3*) were used to inoculate a second series of flasks (Tier 2) with either wild-type (WT) or stacked *bmr6* × *bmr12* mutant (SM) sorghum. Tier 2 samples were incubated for 2 weeks. Changes in biomass content, enzymatic activities, lignin content, and bacterial community dynamics and activity were assessed using the suit of tools depicted in the image. MAGs metagenome-assembled genomes, NIMS nanostructure-initiator mass spectrometry, DNS dinitrosalicylic acid, ABSL acetyl bromide soluble lignin

to 2% carboxymethylcellulose (CMCase) or 2% beech-wood xylan (xylanase) in 100-mM sodium acetate pH 5.0 for a total volume of 80 μL and incubated for 30 min at 50°C. DNS (80 μL) was then added to each sample and heated to 95°C for 5 min. The absorbance at 540 nm was measured for each sample. NIMS analysis was performed as described in detail elsewhere [27]. Briefly, a 2-μL aliquot of the supernatant was transferred into a vial containing 6 μL of 100-mM

glycine acetate, pH 1.2, 0.5 μL of a 5.0-mM aqueous solution of $[U]$-^{13}C glucose, 2 μL of CH_3CN, 1 μL of methanol, 1 μL of solution probe (100 mM in 1:1 (v/v) H_2O: methanol), and 0.1 μL of aniline. The mixture was incubated at room temperature for 16 h. NIMS analysis was performed using a Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer. In each case, 0.2 μL of the quenched reaction sample was spotted onto the NIMS surface and removed after 30 s. Signal

intensities were identified for the ions of the tagging products, and ~4000 laser shots were collected.

Dynamics (Tier 2)

On day 56, each of the three communities from Tier 1 (*comm1*, *comm2*, *comm3*) was used to inoculate a second series of flasks, the Tier 2 incubations (Fig. 1). Two milliliters of each sorghum-deconstructing microbiome (SDM) from Passage 4 (day 56) was used to inoculate triplicate flasks containing 50 mL M9TE (pH 6.5), and 0.5 g of either the parent forage sorghum (WT) or *bmr6 × bmr12* stacked mutant (SM) [22]. Triplicate flasks along with a control were incubated at 50°C, 200 rpm for 2 weeks. At each timepoint (days 2, 5, 7, 9, 12, and 14), flasks were adjusted for evaporation, measured for pH, and sampled for nucleic acid extraction. Five hundred-microliter samples were centrifuged for 5 min at 14,000 × g, and pellets were used for DNA/RNA co-extraction. After 14 days, 500 μL of media was centrifuged, and the supernatant was used for DNS assays and NIMS analysis. Residual biomass was filtered through Miracloth, and a subsample of 100 mg was used for lignin quantification using the Acetyl Bromide Soluble Lignin (ABSL) assay [28] and the rest was dried to determine dry mass. DNA from day 14 was used for metagenome sequencing, while RNA from each sampling point was submitted for metatranscriptome sequencing as described below.

DNA/RNA extraction for metagenomics and metatranscriptomics

DNA and RNA were co-extracted from 500 uL of SDM pellets as previously described [29] using a modified CTAB extraction buffer consisting of equal volumes of 0.5 M phosphate buffer (pH 8) in 1 M NaCl and 10% hexadecyltrimethylammonium bromide (CTAB) in 1 M NaCl. Briefly, tubes containing a bead lysis matrix (Lysing Matrix E Tubes, MP Biomedicals, Solon, OH, USA) received 500 μL of SDM pellet, 0.5 mL of modified CTAB extraction buffer, 50 μL of 0.1 M ammonium aluminum sulfate, and 0.5 mL of phenol to chloroform to isoamyl alcohol (25:24:1) were bead-beaten at 5.5 m/s for 45 s in a FastPrep instrument (MP Biomedicals, Solon, OH, USA). Following bead-beating, tubes were centrifuged at 16,000 × g for 5 min at 4°C. The supernatant was transferred to a new tube containing an equal volume of chloroform to isoamyl alcohol (24:1), vortexed, and centrifuged again. The supernatant was transferred into a new tube containing 1 mL of polyethylene glycol 6000 solutions and 1 μL of linear acrylamide and incubated at room temperature for 2 h. Each sample was extracted a second time by adding 0.5 mL of modified CTAB extraction buffer

to the original Lysing Matrix E tubes and repeating the steps from bead-beating onwards. The first and second extractions were centrifuged at 16,000 × g for 10 min at 4°C. The pellets (two per sample) were washed with 0.5 mL of cold 70% ethanol, dried and combined in 50 μL of RNase-Free water. Purification was carried out using the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. DNA and RNA were eluted in 60 μL and 30 μL of RNase-Free water, respectively. Concentrations were measured by Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA), and the quality was assessed by BioAnalyzer (Agilent, Santa Clara, CA, USA).

Characterization of bacterial communities with amplicon sequencing

Amplicon libraries were prepared in triplicate using 3 ng of DNA per reaction, and the primers 515F [30] and 806R [31] were modified with Illumina sequencing adapters and barcodes [32]. Libraries were pooled in equimolar concentrations and sequenced on the Illumina MiSeq platform using the Illumina Miseq Reagent kit v3. Sequences were demultiplexed based on their unique barcodes and trimmed to the same length. Sequences were dereplicated and sorted by decreasing abundance using USEARCH v11 [33]. The dereplicated sequences were denoised, de novo chimera-filtered, and zero-radius OTUs (ZOTU) generated using unoise3 from USEARCH v11. The resulting ZOTUs, which are a form of amplicon sequence variants (ASVs), were taxonomically characterized against the Greengenes database gg_16s_13.5 [34] using Sintax (USEARCH v11) with a cutoff of 0.8 and genus as the maximum taxonomic level. Total sequences were mapped against the ZOTUs at a 97% identity, and an abundance table was generated that was subsequently transformed into a biom table. ZOTUs were aligned using ClustalW [35], and the alignment was used to generate a phylogenetic tree with IQ-TREE 2 [36] using the model TIM3+F+I+G4 (identified using model finder) and ultrafast bootstrap approximation (UFBoot) with 1000 replicates. The abundance table, mapping file, and phylogenetic tree were imported to the R software using the Phyloseq package [37] (version 1.12.2). For community composition analyses (beta-diversity), data was VST-normalized using the DESeq2 package [38] (version 1.34.0) using a mean fit that was used to calculate a weighted Unifrac distance matrix. The obtained distance matrix was ordinated using multidimensional scaling in Phyloseq. The samples were categorized based on the passage and its effect on data variation tested with Adonis (nonparametric permutation multivariate analysis of variance), performed with 1000 permutations.

Metagenomic sequencing and analysis

Twenty-one DNA samples, 3 from Tier 1 day 56 and 18 from Tier 2 day 14, were submitted to the Joint Genome Institute (JGI, https://jgi.doe.gov/) for sequencing using the Illumina Novaseq platform (150 bp × 2). Individual reads were filtered using JGI's standard metagenomic analysis pipeline (version 3.4.7 from BBtools version 38.24), corrected using bbcms (version 38.34), and co-assembled using metaSPAdes [39] (version 3.13.0). Open Reading Frames were predicted from the assembled contigs using MetaGeneMark [40]. Protein domain annotations were predicted using the pfamA-30 [41] and dbCAN-V8 [42] Hidden Markov Model protein domain databases using an e-value of 1×10^{-5}. Protein categories of interest were screened against the National Center for Biotechnology Information database using BLASTp and dbCAN2's CAZy database for DIAMOND [43] (version 0.9.21.122) with an $e-$value 1×10^{-5}. The metagenome co-assembly was binned using MaxBin 2.0 [44] (version 2.2.5) with default parameters, yielding 103 Metagenome Assembled Genomes (MAGs). The most likely taxonomy was predicted using a custom script (get-Taxon.pl), which searched the predicted proteins of the individual bins against the NCBI non-redundant (NR) database using DIAMOND (version 0.9.21.122) and processed the hits using the least common ancestor (LCA) algorithm proposed by MEGAN Community edition (version 6.11.0) [45]. Completeness and contamination rates for all MAGs were assessed using CheckM (version 1.0.12). MAGs (and associated genes) with at least 30% completeness and less than 10% contamination were retained for the rest of the analyses. Coverage information for the scaffolds of each MAG was extracted from the calculated coverage data TPM normalized data for each scaffold in the metagenome, and MAG abundances in each replicated sample were calculated as the average TPM coverage value over all the scaffolds in a MAG. The compositional variation of each enriched community was analyzed by quantifying their Local Contribution to Beta Diversity (LCBD) using the R package *adespatial* with the Hellinger dissimilarity coefficient and *p* value correction using the Holm method. A phylogenetic tree for the MAGs was reconstructed in KBase [46] based on universal genes defined by the Cluster of Orthologous Groups using maximum likelihood. Average nucleotide identity between taxonomically related MAGs (genus level) was quantified also in KBase.

Metatranscriptomic sequencing and analysis

Fifty-four RNA samples, from each of the treatments and time points of the Tier 2 experiment, were also submitted to JGI for metatranscriptomic sequencing using the Illumina Novaseq platform (150bp × 2).

Sequenced samples represented triplicated RNA samples from adapted communities incubated with stacked mutant (SM) and wild-type (WT) sorghum. The filtered reads were assessed using FastQC (version 0.11.8, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and mapped to the metagenome co-assembly using Bowtie2 [47] (version 2.3.4.3). Gene counts were generated using featureCounts [48] (version 1.6.3) and normalized for both gene length and library size by transcripts per million (TPM), using a custom R script. For metatranscriptome ordination analyses a Bray-Curtis dissimilarity matrix was calculated using R's Vegan on the raw feature counts table that was first filtered to retain only those genes appearing in at least 5 samples (out of the total 54 samples) and a mean count of 10. The resulting table was VST-normalized with DESeq2. The samples were categorized based on time (day), type of biomass (WT and SM), and categorical effects on data variation tested with Adonis (nonparametric permutation multivariate analysis of variance), performed with 1000 permutations. Average transcriptome abundances per selected MAG were calculated on the TPM-normalized data. For differential expression analyses, the feature count data was filtered using the parameters used for the transcriptome ordination analysis, retaining genes appearing in at least 5 samples with a mean count of 10. Differential expression analyses were carried out using DESeq2 using a parametric fit. The results filtered for a corrected *p* value < 0.01 and an absolute $\log_2$ fold change > 1. Heatmaps showing normalized expression levels for relevant genes were calculated on the DESeq-2 VST-normalized data using R's *pheatmap* package, and rows were arranged based on a Bray-Curtis dissimilarity matrix.

Network reconstruction

A network was constructed for the transcriptome data based on centered log ratio-transformed feature counts data [49]. Prior to normalization, the data was subset to include genes detected in at least 50% of the total number of samples. Network reconstruction was conducted with the Molecular Ecological Network Analyses pipeline (MENAP, http://ieg4.rccc.ou.edu/mena/) with the following settings: for missing data, fill blanks with 0.01 if data have paired values; do not take logarithm as the data was already CLR normalized; use Spearman correlation similarity matrix; and calculate by decreasing cutoff from the top. Random Matrix Theory (RMT) was used to automatically identify the appropriate similarity threshold for network reconstruction [50, 51]. The network was visualized in Cytoscape [52] (version 3.9.0) using force-directed graph drawing and colored

based on the taxonomic identity of the included MAGs. Potential interactions between MAGs were defined by the detection of direct linkages between pairs of networked MAGs.
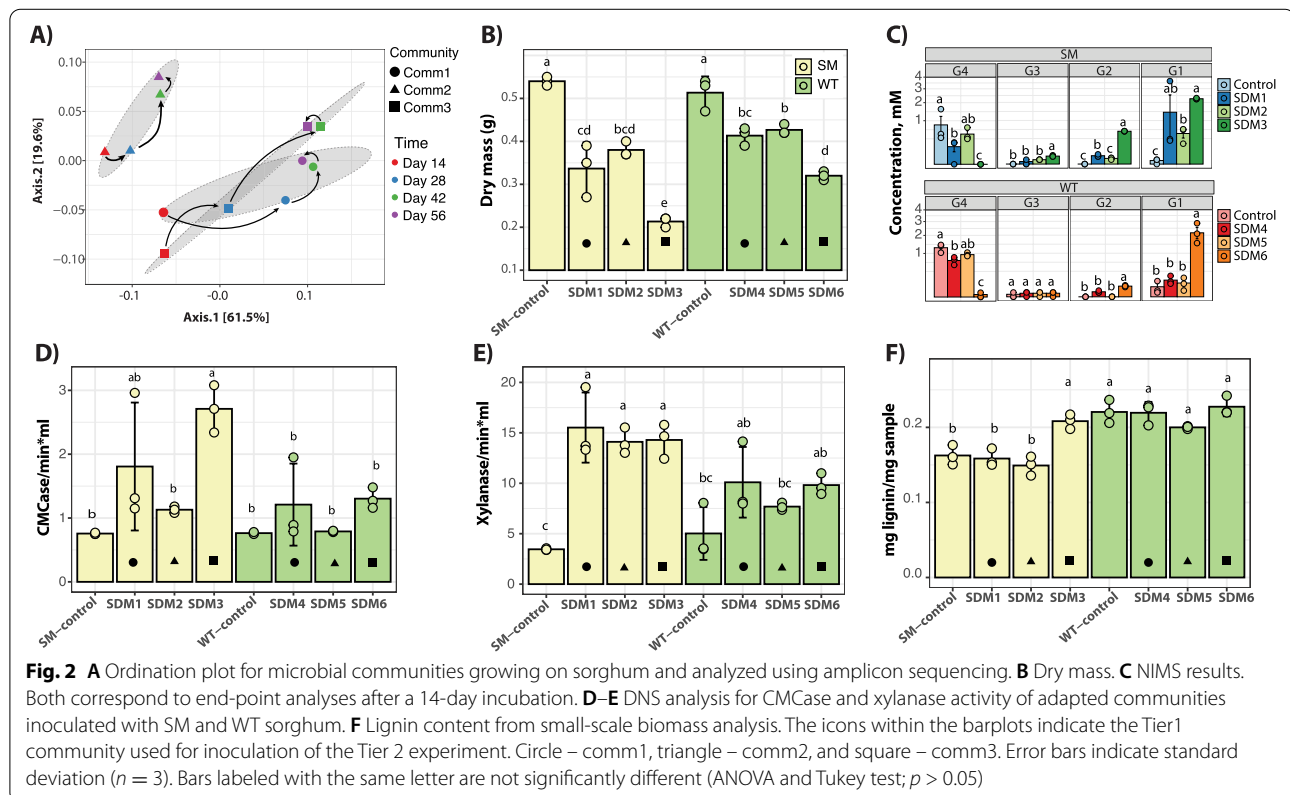
Results

Microbial community adaptation to grow on sorghum

Green waste compost was used to inoculate three parallel microbiomes which were adapted to grow on sorghum biomass as the sole carbon source for 56 days (Fig. 1). Since compost is a heterogeneous inoculum, we anticipated that this cultivation would result in microbiomes with distinct populations. Measurement of residual sorghum biomass by day 56 showed that *comm1* and *comm2* had a 40% reduction in biomass content and *comm3* had a 57% biomass reduction (Supplementary Fig. 1). Amplicon sequencing demonstrated that these microbiomes differentiated into individual communities (*comm1*, *comm2*, and *comm3*). Analysis of community composition showed that the individual microbiomes did not group over time (PERMANOVA: $df = 3$, $F = 1.59$, $p = 0.21$) but rather varied by the community (PERMANOVA: $df = 2$, $F = 4.93$, $p = 0.003$, $r^2 = 52.3\%$) with each following a different trajectory (Fig. 2A). The microbiomes *comm1* and *comm3* were more closely related to each other than *comm2*, which

was separated at a considerable distance from the other microbiomes in the ordination plot. The trajectories of these microbiomes suggest that they possess distinct metabolic capabilities and that by day 56 the community composition had stabilized.

Comparative deconstruction of sorghum

The emergence of three distinct microbiomes from the initial green-waste compost inoculum provided an opportunity to compare the performance of parallel microbiomes with different community compositions. We compared the deconstructive abilities of these communities on sorghum varieties with different lignin content and monomeric composition to examine the effect of lignin on microbiome performance. Multiple mutants from the lignin biosynthetic pathway have been developed in sorghum, and the *bmr-6x12* double mutant was chosen for the parallel experiments [22]. This stacked mutant, in which mutations that affect both the lignin biosynthetic gene cinnamyl alcohol dehydrogenase (*bmr-6*) and caffeic acid O-methyltransferase (*bmr-12*), has lower lignin content and is more easily deconstructed compared to the native sorghum line [53]. Therefore, we compared the corresponding non-mutant sorghum hybrid, referred to as wild type (WT), and the *bmr-6x12* line, referred to as stacked mutant (SM). Microbiomes

Fig. 2 A Ordination plot for microbial communities growing on sorghum and analyzed using amplicon sequencing. **B** Dry mass. **C** NIMS results. Both correspond to end-point analyses after a 14-day incubation. **D**–**E** DNS analysis for CMCase and xylanase activity of adapted communities inoculated with SM and WT sorghum. **F** Lignin content from small-scale biomass analysis. The icons within the barplots indicate the Tier1 community used for inoculation of the Tier 2 experiment. Circle – comm1, triangle – comm2, and square – comm3. Error bars indicate standard deviation ($n = 3$). Bars labeled with the same letter are not significantly different (ANOVA and Tukey test; $p > 0.05$)

cultivated for 56 days on forage sorghum were individually inoculated into triplicate cultures containing SM sorghum (SDM1-3) or the WT sorghum (SDM4-6) and cultivated for 14 days (Fig. 1). Endpoint measurements of residual dry mass biomass demonstrated that the communities cultivated on the SM sorghum exhibited the greatest biomass loss (Fig. 2B). Among the SM communities, SDM3 had a significantly higher average biomass reduction (75% vs control), while SDM6, inoculated using WT-sorghum, exhibited 54% biomass loss compared to the control (Fig. 2B). SDM3 (SM-treatment) and SDM6 (WT-treatment) are derived from the same *comm3*, and the levels of biomass consumption showed that SDM3 was significantly higher than SDM6 (ANOVA and Tukey test, $p < 0.01$) (Fig. 2B). An analysis of cellotetraose hydrolysis showed that regardless of the type of biomass used as substrate (WT or SM), treatments inoculated with *comm3*-derived microbiome released the highest levels of glucose with SDM3 (SM sorghum) compared to SDM6 (WT sorghum) (SDM3: 2.25 mM, $\sigma = 0.04$; SDM6: 2.15 mM, $\sigma = 0.6$) (Fig. 2C). Cellulase and xylanase activity were further investigated using DNS assays and showed the highest enzymatic activity in the *comm1* and *comm3*-derived treatments (SDM1/SDM4 and SDM3/SDM6, respectively). The results also indicated a higher cellulase and xylanase activity in the SM-sorghum treatments compared to the WT-sorghum treatment (Fig. 2D, E).
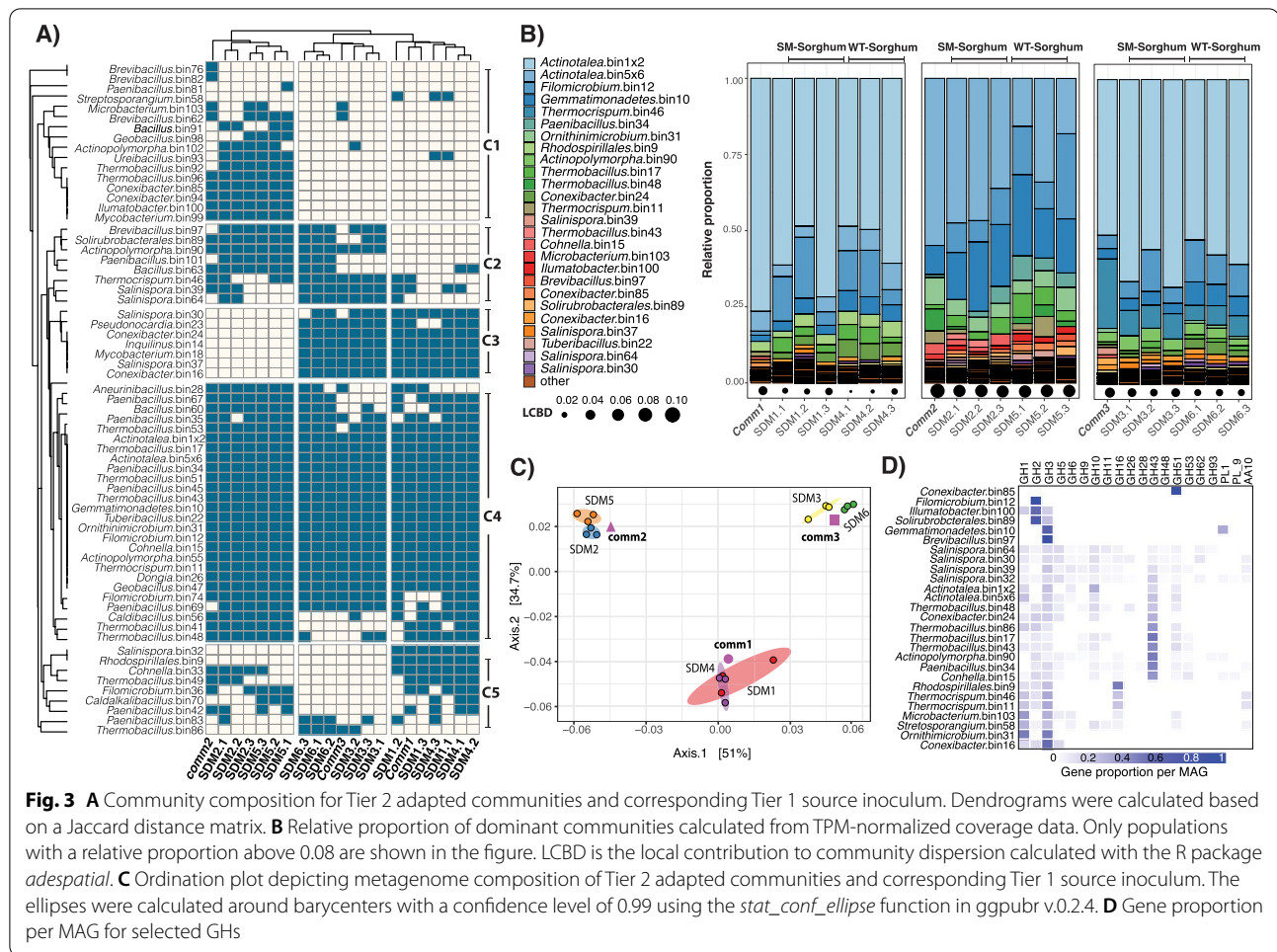
The changes in biomass composition were further analyzed by measuring relative lignin content compared to uninoculated controls. The residual biomass from the SM communities had a significantly lower lignin content than its WT counterpart. This result can be explained by the difference in lignin composition between the SM and WT plants (Fig. 2F). An increase in lignin amount in residual biomass was statistically significant only for SDM3 treatment, whereas all inoculated treatments had a lower lignin content than their controls and the calculations were not statistically relevant (ANOVA and Tukey test, $p > 0.05$) (Fig. 2F).

Metagenomic analyses reveal the metabolic potential for biomass transformation

A total of 103 metagenome-assembled genomes (MAGs) were reconstructed, but only 66 that had a completeness above 30% and contamination lower than 10% were considered for downstream analysis. These selected MAGs and their phylogenetic relationships based on universal genes are shown in Supplementary Figure 2. Figure 3A shows the shared and unique reconstructed MAGs in each of the Tier 2 samples and their community sources (*comm1*, *comm2*, and *comm3*). Inspection of the clustering patterns showed that the composition of Tier 2

samples clustered according to their community sources, as also observed in amplicon-based analysis (Fig. 2A). The MAGs are separated into five clusters (C1–C5, Fig. 3A). Cluster 1 (C1) represented the communities mostly unique to *comm2*-derived samples (SDM2/SDM5) and included *Actinopolymorpha* bin102, *Bacillus* bin91, *Brevibacillus* bin76, 82, and 62 (average nucleotide identity (ANI) = 76%), *Conexibacter* bin85 and 94 (ANI = 78%), *Geobacillus* bin98, *Illumatobacter* bin100, *Microbacterium* bin103, *Mycobacterium* bin99, *Paenibacillus* bin81, *Streptosporangium* bin58, *Thermobacillus* bin92 and 96 (ANI = 77%), and *Ureibacillus* bin93. Cluster 2 (C2) contained bacterial populations shared between *comm2* and *comm3*-derived samples (SDM2/SDM5 and SDM3/SDM6). Cluster 2 included *Actinopolymorpha* bin90, *Bacillus* bin63, *Brevibacillus* bin97, *Paenibacillus* bin101, *Salinispora* bin39 and 64 (ANI = 77%), *Solirubrobacterales* bin89, and *Thermocrispum* bin46. Cluster 3 (C3) represented the populations exclusively shared between *comm3* and *comm1*-derived samples (SDM6/SDM3, and SDM1/SDM4). Cluster 3 populations included *Conexibacter* bin16 and 24 (ANI = 79%), *Inquilinus* bin14, *Mycobacterium* bin18, *Pseudoncardia* bin23, *Salinispora* bin30, and 37 (ANI = 77%). Cluster 4 (C4) represented the core populations among all samples and included *Actinopolymorpha* bin55; *Actinotalea* bin1 and 5 (ANI = 86%); *Aneuribacillus* bin28; *Bacillus* bin60; *Caldibacillus* bin56; *Conhella* bin15; *Dongia* bin26; *Filomicrobium* bin12 and 24 (ANI < 70%); *Gemmatimonadetes* bin10; *Geobacillus* bin47; *Ornithimicrobium* bin31; *Paenibacillus* bin34, 35, 45, 67, and 69 (ANI = 76–78%); *Thermobacillus* bin17, 41, 43, 48, 51, and 53 (ANI = 77–89%); *Thermocrispum* bin11; and *Tuberibacillus* bin22. Finally, cluster 5 (C5) included some populations such as the *Rhodospirillales* bin9 and *Salinispora* bin32 which were unique to SDM1/SDM4 and *Thermobacillus* bin96 that was unique to SDM1/SDM6. Other populations in this cluster included *Cohnella* bin32, *Thermobacillus* bin49, *Filomicrobium* bin36, *Caldalkalibacillus* bin70, and *Paenibacillus* bin42 and 35 (ANI < 70%), all of which were shared between SDM2/SDM5 and SDM1/SDM4.

According to the analysis of coverage distribution of the binned genomes (Fig. 3B), Tier 2 communities were dominated by the *Actinotalea* genome populations (*Actinotalea* bin1 and bin5). *Actinotalea* bin1 contigs accounted for more than 70% of the total contig coverage in SDM1/SDM3 and SMD4/SDM6, while *Actinotalea* bin5 accounted for 41% of the total contig coverage in SDM2 and 24% in SDM5. Highly prevalent MAGs, also identified as part of cluster 4 (core populations), included populations of *Filomicrobium* bin12, *Gemmatimonadetes* bin10, *Paenibacillus* bin34, *Ornithiumicrobium* bin31, *Thermobacillus* bin17, 43 and 48, *Thermocrispum*

Fig. 3 **A** Community composition for Tier 2 adapted communities and corresponding Tier 1 source inoculum. Dendrograms were calculated based on a Jaccard distance matrix. **B** Relative proportion of dominant communities calculated from TPM-normalized coverage data. Only populations with a relative proportion above 0.08 are shown in the figure. LCBD is the local contribution to community dispersion calculated with the R package *adespatial*. **C** Ordination plot depicting metagenome composition of Tier 2 adapted communities and corresponding Tier 1 source inoculum. The ellipses were calculated around barycenters with a confidence level of 0.99 using the *stat_conf_ellipse* function in ggpubr v.0.2.4. **D** Gene proportion per MAG for selected GHs

bin46 and 11, and *Cohnella* bin15. Other highly prevalent MAGs were *Brevibacillus* bin97 (C2), *Conexibacter* bin16, 24 (C3), 85 (C1), *Ilumatobacter* bin100 (C1), *Microbacterium* bin103 (C1), *Rhodospirillales* bin9 (C5), *Salinispora* bin37 (C3), 39 (C2), 64 (C2), *Solirubrobacterales* bin89 (C2), and *Thermocrispum* bin46 (C2). Analysis of local contribution to beta diversity (LCBD) showed no significant variation (Holm corrected *p* values > 0.05) in the composition of the enriched communities when comparing the composition of the Tier2 enrichments and their Tier1 source inoculum (Fig. 3B).

An ordination analysis of the normalized coverage for the contigs of the selected bins (Fig. 3C) showed that the different samples clustered together based on their inoculum regardless of biomass type (SM or WT). Furthermore, a permutational analysis of variance showed that the type of inoculum (PERMANOVA: $df = 2$, $F = 54.9$, $p = 9.9 \times 10^{-5}$) and type of biomass (PERMANOVA: $df = 1$, $F = 5.8$, $p = 9.9 \times 10^{-5}$) had significant effects on metagenomic clustering and explained 84.2% and 4.4% of the observed patterns (Fig. 2C).

Prediction and annotation of genes identified within each MAG showed that the abundant *Actinotalea* bins contained some genes coding for putative glycoside hydrolases (GHs) relevant for the degradation of polysaccharides. Assignments of putative activities were made by reference to the CAzy database (www.cazy.org) and represent predicted activities of defined protein families. *Actinotalea* bin1 contained genes encoding GH6 and GH10, while *Actinotalea_bin5* had genes encoding for GH5, 6, 10, 43, and 51 genes. On the other hand, other abundant MAGs such as *Actinopolymorpha* bin90, *Conhella* bin15, *Paenibacillus* bin34, and *Thermobacillus* bin17 and bin48 contained more of the genes encoding GHs possibly involved with pectin, hemicellulose, and cellulose degradation (Fig. 3D). Supplementary Figure 3 shows the distribution of relevant GHs among the selected MAGs.

Sequential degradation of sorghum biomass follows two distinct trajectories

The SDM1 and SDM3 treatments had the same most abundant population (*Actinotalea* bin1) and the highest

activities among the Tier 2 microbiomes. Therefore, we performed an in-depth comparison of time-dependent gene expression patterns in these microbiomes to identify similarities and differences in expression patterns, focusing on genes for the deconstruction of plant polymers. We also performed a comparison between SDM3 and SDM6 to see if the sorghum substrate had any effect on gene expression patterns.

An ordination analysis of the metatranscriptome showed that the three selected enrichments (SDM1, SDM3, and SDM6) followed two distinct trajectories (Supplementary Fig. 4A). Similar to the metagenome analysis, the metatranscriptomes clustered based on their initial inoculum and shifted gradually over the course of 14 days. SDM3 and SDM6 followed a similar 2-week trajectory, despite being grown in SM and WT sorghum, respectively. SDM1 followed a different trajectory from SDM3 and SDM6, but also exhibited gradual shifts in overall activity, indicative of sequential changes in community structure. A permutational analysis of variance further indicated that type of inoculum (PERMANOVA: $df = 2$, $F=40.33$, $p = 9.9 \times 10^{-5}$) and time (PERMANOVA: $df = 5$, $F= 21.59$, $p = 9.9 \times 10^{-5}$) each had a significant effect on metatranscriptome trajectory, explaining 40.3% and 21.5% of the observed variation, respectively (Supplementary Fig. 4A). The analysis of variance also indicated that the type of biomass (WT and SM) did not have a significant effect on metatranscriptome trajectory. Based on these results, we chose to focus our analyses on the characterization of SDM1 and SDM3.

An analysis of the normalized abundance of transcriptomes for each reconstructed MAG indicated that the *Actinotalea*-bin1 was the most active organism in the enrichments across sampling times (Supplementary Fig. 4B). Other highly active bins included *Thermobacillus* bin17, *Filomicrobium* bin12, *Thermocrispum* bin46, *Gemmatimonadetes* bin10, *Thermobacillus* bin53 and 51, *Tuberibacillus* bin22, and *Geobacillus* bin47. Genome bins that were more active in SDM1 included *Actinotalea* bin5, *Rhodospirillales* bin9, and *Inquilinus* bin14, while *Actinopolymorpha* bin90, *Paenibacillus* bin42, and *Thermobacillus* bin49 were more active in SDM3 (Supplementary Fig. 4B).

Random matrix theory (RMT)-based network analysis was performed to define putative interactions among the networked populations and to further explore transcriptome dynamics [50, 51]. Figure 4A depicts the reconstructed network based on metatranscriptome expression profiles. Each MAG in the network is colored showing that bacterial populations identified as highly abundant in the metagenome and with high expression levels in the metatranscriptome formed highly connected clusters within the network. The reconstructed network

(Fig. 4A) consisted of 22,887 nodes (networked genes) and 5,018,619 links with correlation values between 0.9–1.0 and 164 large modules (>10 connected nodes). Isolation of nodes and links by reconstructed MAG with connected neighbors representing co-expression patterns were used to define pairs of interacting bacterial populations, this is, MAGs likely interacting with each other. These patterns showed that populations represented by *Actinotalea* bin1, *Actinotalea* bin5, *Filomicrobium* bin12, and *Gemmatimonadetes* bin10 were highly interconnected and likely interacting with each other (Fig. 4B). Because of their conservation in all the microbiomes, high level of abundance and activity (Figs. 2B and 3B), and the direct interconnections between these four MAGs (Fig. 4B), we defined these bins as key populations within the adapted communities. Mapping of differential expression ($\log_2$fold change for genes with $p < 0.01$) onto the network showed that *Actinotalea* bin5 was significantly more active in SDM1 during the 14-day incubation (Fig. 4B). We also observed that *Actinotalea* bin1, *Filomicrobium* bin12, and *Gemmatimonadetes* bin10 were more active in SDM3 than in SDM1 from day 5 to day 9. The significantly higher activity of these three central bins remained through the 14–day incubation for *Actinotalea* bin1 and declined first for *Gemmatimonadetes* bin10 by day 12 and then for *Filomicrobium* bin12 by day 14 (Fig. 4C).

One-to-one putative interactions between these four central MAGs with other members of the adapted community were also predicted from the network (Fig. 4A, B). *Actinotalea* bin5 and *Filomicrobium* bin12 had direct connections with a larger number of MAGs than *Actinotalea* bin1 and *Gemmatimonadetes* bin10. Populations directly linked with *Actinotalea* bin5 included *Thermocrispum* bin11, *Cohnella* bin29, *Salinispora* bin37, *Streptosporangium* bin38, *Thermobacillus* bin50, *Thermobacillus* bin53, and *Caldibacillus* bin56. *Filomicrobium* bin12 on the other hand had direct links with *Thermobacillus* bin17, *Ornithimicrobium* bin31, *Paenibacillus* bin45, *Thermobacillus* bin51, *Paenibacillus* bin61, *Filomicrobium* bin74, *Paenibacillus* bin83, and *Filomicrobium* bin88. Aside from their connections with the other central MAGs, *Actinotalea* bin1 was found as linked with *Conexibacter* populations bin16 and 24, while *Gemmatimonadetes* bin10 was linked to *Thermobacillus bin*53 (also connected with *Actinotalea* bin5) and with *Caldalkalibacillus* bin73.

We also explored the network associations of MAGs that represented likely key contributors to the process of cell wall degradation. We focused on those MAGs identified as containing genes coding for putative GHs reported as potentially involved in pectin, cellulose, and xylan degradation (Supplementary Fig. 3) and their
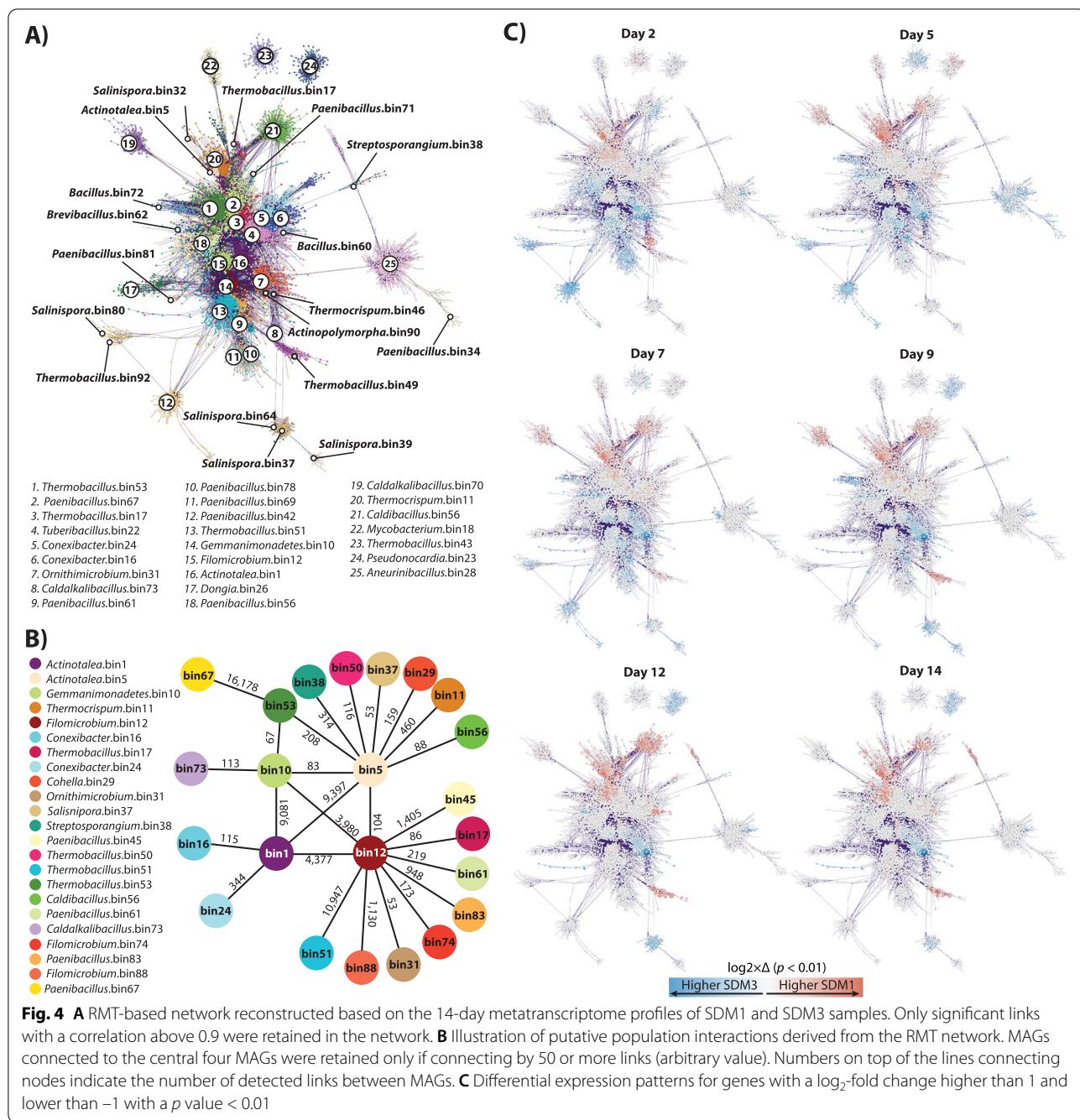
Fig. 4 A RMT-based network reconstructed based on the 14-day metatranscriptome profiles of SDM1 and SDM3 samples. Only significant links with a correlation above 0.9 were retained in the network. **B** Illustration of putative population interactions derived from the RMT network. MAGs connected to the central four MAGs were retained only if connecting by 50 or more links (arbitrary value). Numbers on top of the lines connecting nodes indicate the number of detected links between MAGs. **C** Differential expression patterns for genes with a log$_2$-fold change higher than 1 and lower than −1 with a *p* value < 0.01

expression levels (Supplementary Fig. 4B). *Paenibacillus* bin67 was another MAG of interest as it contains several putative genes encoding for GHs potentially contributing to the degradation of pectin (GH2, GH43) and hemicellulose (GH10, GH51). *Paenibacillus* bin67 was highly connected with *Thermobacillus* bin53, which contained genes encoding for a wide array of putative GHs including those from families 2, 5, 10, 16, 28, 43, 51, and 53;

the carbohydrate esterase (CE) family 8; and pectin lyase (PL) families 1 and 9 (Supplementary Fig. 3). *Thermobacillus* bin53 was also linked to with *Actinotalea* bin1 and bin5, likely acting as a connection between the dominant *Actinotalea* populations and the rest of the communities.

Another likely prominent group in the process of polysaccharide degradation was the *Salinispora* populations. Three of these MAGs (bin37, 39, and 64) were
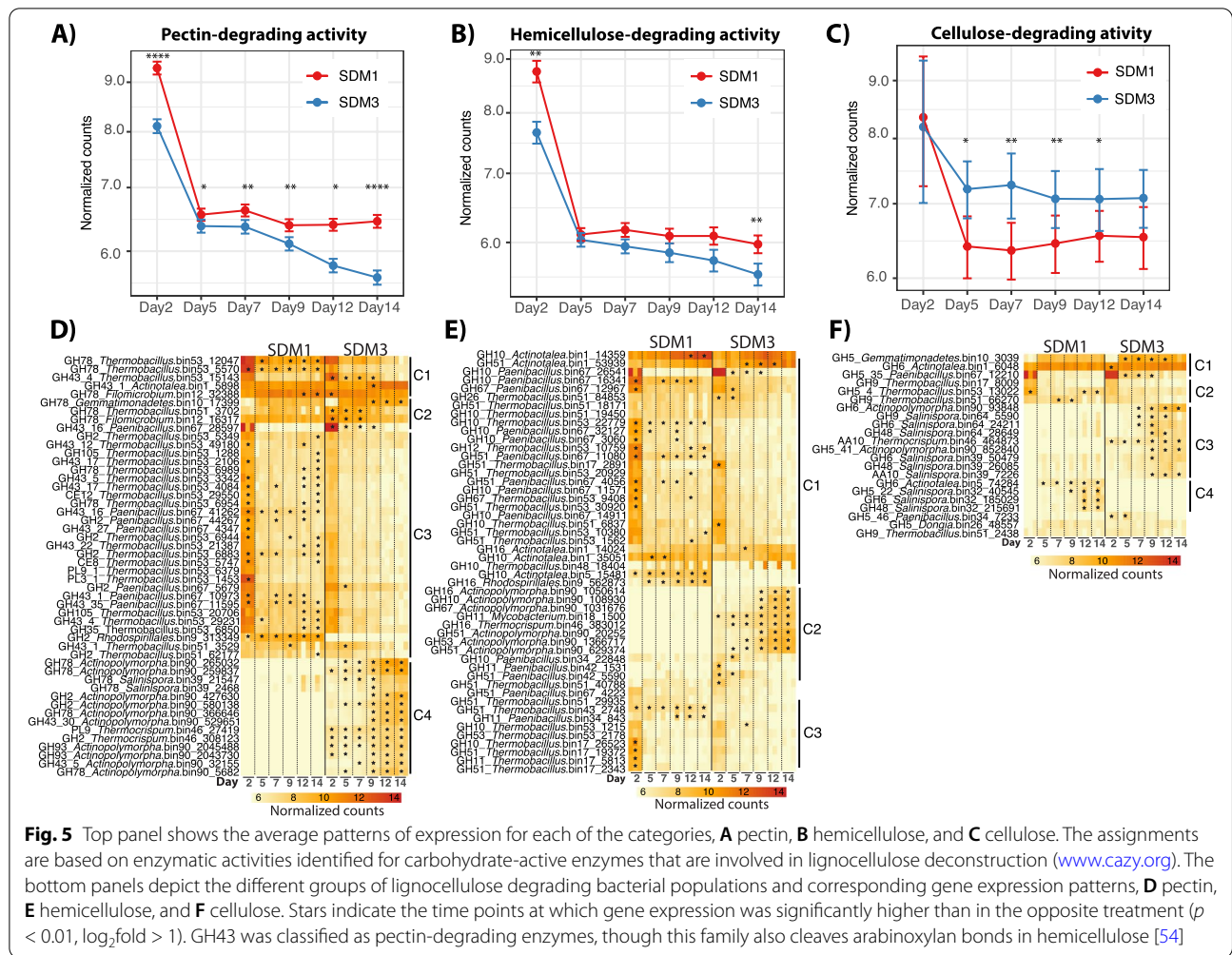
Fig. 5 Top panel shows the average patterns of expression for each of the categories, **A** pectin, **B** hemicellulose, and **C** cellulose. The assignments are based on enzymatic activities identified for carbohydrate-active enzymes that are involved in lignocellulose deconstruction (www.cazy.org). The bottom panels depict the different groups of lignocellulose degrading bacterial populations and corresponding gene expression patterns, **D** pectin, **E** hemicellulose, and **F** cellulose. Stars indicate the time points at which gene expression was significantly higher than in the opposite treatment (p < 0.01, $\log_2$fold > 1). GH43 was classified as pectin-degrading enzymes, though this family also cleaves arabinoxylan bonds in hemicellulose [54]

detected forming a discrete cluster showing high levels of transcriptomic activity in SDM3 from day 7 to day 14 (Fig. 4A, B). Among these three MAGs, *Salinispora* bin39 and 64 contained a wide arsenal of genes encoding for GHs including GH2, GH5, GH6, GH9, GH10, GH11, GH16, GH43, GH48, GH51, GH53, and PL9 (only bin39), and GH62 and GH93 (only in bin64) (Supplementary Fig. 3).

Detailed exploration of the normalized transcriptome expression profiles indicated that degradation of the primary cell wall was likely initiated by the activity of microorganisms producing enzymes for pectin degradation in a process that was significantly higher in SDM1 than in SDM3 (Wilcox pairwise comparison, p < 0.01) and that continued steadily over the 14 days of incubation (Fig. 5A). Pectin-degrading gene expression patterns were separated into four main clusters (Fig. 5D). Cluster 1 (C1) included pectin-degrading genes that were significantly highly expressed in SDM1 and SDM3 (p < 0.01, $\log_2$fold > 1), C2 and C4 composed by genes

highly expressed in SDM3, and C3 genes significantly highly expressed in SDM1. Based on the observed patterns of expression in these clusters, pectin degradation in both treatments was driven by the high levels of expression of genes encoding GH43 and GH78 from *Actinotalea* bin1 and *Filomicrobium* bin12, respectively. Two main populations of *Firmicutes* controlled pectin degradation at the start of the incubation in SDM1 including *Thermobacillus* bin53 and *Paenibacillus* bin67 through the expression of most of the genes shown in C1 and C3 at significantly higher levels than in SDM3 (p < 0.01, $\log_2$fold > 1). Other contributors to the process of pectin degradation in SDM3 were *Thermobacillus* bin51 and *Rhodospirillaes* bin9 that expressed genes encoding GH2 and GH43 (in bin51 only) through the whole incubation. Initial drivers of pectin degradation in SDM3 included *Thermobacillus* bin51 and *Filomicrobium* bin12 (C2) and *Salinispora* bin39 (C4) expressing a gene encoding GH78 and *Paenibacillus* bin67 (C2) expressing a gene encoding GH43. *Actinopolymorpha*

bin90 (C4) was also among the main contributors to pectin degradation in SDM3 through the expression ($p < 0.01$, $\log_2$fold > 1) of genes encoding GH2, GH43, GH78, and GH93 together with *Thermocrispum* bin46 expressing genes encoding PL9 and GH2.

Hemicellulose deconstruction gene expression dynamics resembled the expression pattern for pectin deconstruction genes as indicated by the higher levels of hemicellulose-deconstructing gene expression in SDM1 than in SDM3 (Fig. 5B). It is likely that *Actinotalea* bin1, the most abundant bacterial population, initiated and maintained the process of hemicellulose deconstruction in both treatments given the high expression levels of the genes encoding GHs from the families 10, 51, and 16 from this MAG (C1, Fig. 5E) together with the gene encoding GH10 from *Paenibacillus* bin67. Other pioneering populations in the hemicellulose deconstruction process were *Thermobacillus* bin17 (C3), 51, 53 (C1 and C3), and *Paenibacillus* bin67 (Fig. 5E, cluster 1) through the expression of gene encoding GHs from the families 10, 11, 16, 26, and 51 whose expression was significantly higher in SDM1 than in SDM3 ($p < 0.01$, $\log_2$fold > 1). The expression of these GHs was higher during day 2 and then declined but continued through the incubation period. GHs that contributed to the high hemicellulose-degrading activity in SDM1 were the genes encoding GHs 10 and 16 from *Actinotalea* bin5 and *Rhodospirillales* bin9, whose activity was detected since the beginning of the incubation and increased over time up to day 14 (C1, Fig. 5E). Significantly highly expressed GHs in SDM3 are shown in Fig. 3E cluster 2 and included genes encoding GH10, 11, 16, 51, 53, and 67 from *Actinopolymorpha* bin90, *Mycobacterium* bin18, *Thermocrispum* bin46, and *Paenibacillus* bin42. The expression of these genes encoding GHs increased over time with those from *Actinopolymorpha* bin90, *Mycobacterium* bin18, and *Thermocrispum* bin46 reaching higher levels from day 9 to day 14 likely indicating the critical roles of these populations for the progression of biomass decomposition in SDM3.

In contrast to pectin and hemicellulose, the expression of genes related to cellulose deconstruction was overall higher in SDM3 than in SDM1 (Fig. 5C). Expression patterns showed that the cellulose deconstruction commenced in both treatments (SDM1/SDM3) by the activity of *Actinotalea* bin1, *Gemmatimonadetes* bin10, *Paenibacillus* bin67, *Thermobacillus* bin53, and *Thermobacillus* bin51 expressing genes encoding GH5 and GH9 (C1, Fig. 5F). In SDM3, the cellulose degradation process was complemented by the significantly higher activity ($p < 0.01$, $\log_2$fold > 1) of *Actinopolymorpha* bin90 expressing genes encoding GH5 and GH9, *Thermocrispum* bin46 and *Salinispora* bin39 expressing genes encoding AA10 proteins,

which represents a family of lytic polysaccharide monooxygenases [55], together with *Salinispora* bin64 expressing genes encoding GH6, 9, and 48, all of which increased over time (C3, Fig. 5F). In SDM1, *Salinispora* bin32 was a key contributor to cellulose degradation through the expression of a genes encoding GH5, 6, and 48 that reached its highest from day 9 to day 12. Other bacterial populations likely contributing to cellulosic activity were *Paenibacillus* bin34, *Dongia* bin26 (C5), and *Thermobacillus* bin17 (C2) through the expression of genes encoding GH5 and GH9.

In comparison to bacterial polysaccharide deconstruction, bacterial lignin deconstruction is less understood [56]. Inspection of the metagenome and metatranscriptome identified a protein annotated as a multi-copper oxidase in the *Gemmatimonadetes* bin10. Despite this annotation, this protein is not related to characterized bacterial laccases [57]. A homolog of this protein in a closely related thermophilic *Gemmatimonadetes* population was identified by proteomics as one of the most abundant proteins in the supernatant of bacterial consortium growing on switchgrass at 60°C [58]. In addition, a homologous protein was identified in cultures of *Thermobifida fusca* growing on sugarcane bagasse [59]. The protein from *T. fusca* was isolated from the supernatant of the bagasse-grown culture and was shown to contain 1 Cu per protein, rather than 4 Cu atoms, as had been previously observed for laccases [58]. This protein improved sugar release from bagasse when mixed with endoglucanase and xylanase purified from the *T. fusca* supernatant. The protein also improved the deconstruction efficiency of an engineered cellulosome on the wheat straw when it was incorporated as a heterologous protein with a dockerin domain that bound to the artificial cellulosome [60]. In the sorghum cultures, the expression of this gene in *Gemmatimonadetes* bin10 expression was found to be significantly higher in SDM3 than in SDM1 from day 5 to day 7, reaching similar levels on day 9 (Fig. 6). In addition, the expression of a complete pathway for catabolism of aromatic compounds from 4-hydroxybenzoate transformation to protocatechuate and its conversion to succinyl-CoA and acetyl-CoA via the beta-ketopadipate pathway was observed in the *Filomicrobium* bin12. This pathway was detected at significantly higher levels in SDM3 compared to SDM1 from day 2 to day 7 (Fig. 6).

Discussion

The two-tier cultivation of compost-derived microbiomes on sorghum led to the establishment of microbiomes for which community structure and performance could be assessed. Initial inoculation and growth on biomass sorghum provided distinct microbiomes (*comm 1-3*) that traversed independent trajectories during 2
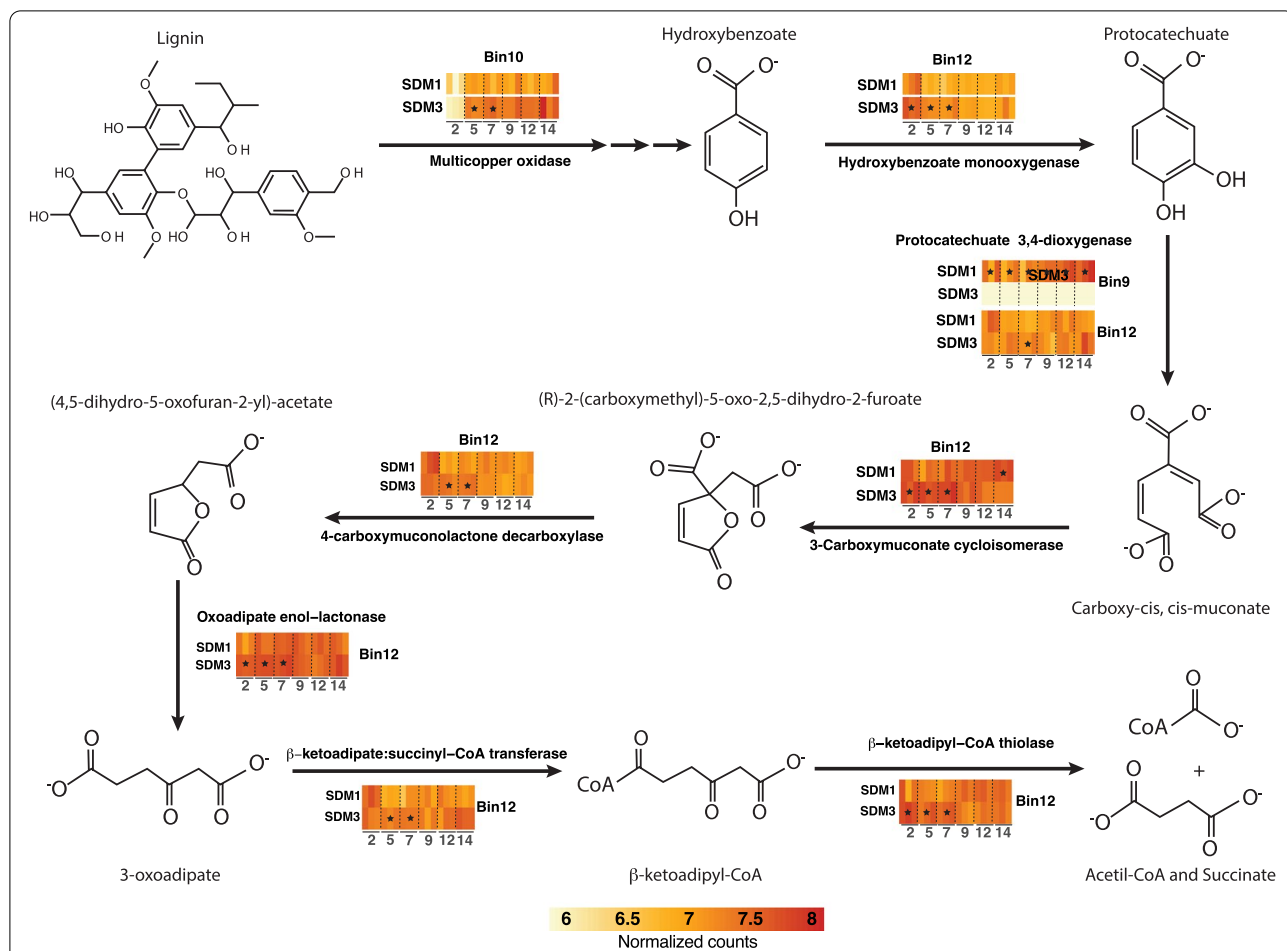
Fig. 6 Schematic representation of the expression patterns for aromatic-degrading genes. The heatmaps are colored based on normalized counts for the targeted genes. Stars indicate the time points at which gene expression was significantly higher than in the opposite treatment ($p < 0.01$, $\log_2$fold > 1)

months of adaptation (Fig. 1A). The development of distinctive microbiomes echoes parallel cultivation of microbiomes from *Sarracenia purpurea* pitcher plants grown on ground crickets [61]. The community structures of these parallel microbiomes also diverged during adaptation and the pitcher plant-derived consortia had variable activities in chitin deconstruction. The second-tier growth using the *comm 1-3* microbiomes as inoculum for growth on wild-type (SDM4-SDM6) and lignin-reduced sorghum varieties (SDM1-SDM3) demonstrated that the structure and deconstructive activities of these microbiomes are reproducible. This observation suggests that after adaptation the community structures are maintained, allowing detailed comparisons between microbiomes that are statistically robust. Furthermore, the analysis of variance between our compost-enriched microbiomes grown on wild-type sorghum (SDM6) compared to the *bmr-6x12* mutant (SDM3) provides persuasive evidence that community structure, rather

than plant cell wall structure, defines the trajectory of deconstruction. The increased digestibility of the *bmr-6x12* mutant is consistent with its reduced lignin content and resulting in lower recalcitrance [22, 53].

Genome-resolved metagenomics demonstrated the most abundant populations in the microbiomes were two closely related *Actinotalea* populations. The most abundant *Actinotalea* population in the *comm1* and *comm3*-derived microbiomes (*Actinotalea* bin1) possessed fewer deconstructive enzymes than the most abundant *Actinotalea* population (*Actinotalea* bin5) in *comm2*; however, the performance of the *comm2*-derived microbiomes, as measured by biomass loss, cellulase/xylanase activity, and lignin remaining in the residual biomass was generally lower compared to the *comm1* and *comm3*-derived microbiomes. The presence of *Gemmatimonadetes* bin10 and *Filomicrobium* bin12 in *comm1-3* and their daughter communities suggested their prominent role in biomass deconstruction. This hypothesis was confirmed by

both network analysis of gene expression, which demonstrated that gene expression in these populations was correlated, and functional analysis, which demonstrated that the *Gemmatimonadetes* and *Filomicrobium* populations were involved in lignin deconstruction, an essential function in the deconstruction of the secondary plant cell wall. In addition, *Paenibacillus* bin67 and *Thermobacillus* bin17, bin51 and bin53 are broadly distributed and demonstrated high, correlated expression of predicted pectinases and hemicellulases, especially early in the 2-week cultivation, that is consistent with the deconstruction of the primary cell wall. The contribution of these lower abundance populations to cell wall deconstruction is a phenomenon which has been observed in native microbiomes that deconstruct complex polysaccharide substrates like the human gut [62].

The microbiomes derived from *comm1* and *comm3* growing on the *bmr-6x12* sorghum mutant that was dominated by *Actinotalea* bin1 provided an opportunity to link the community performance, as measured by biomass loss and enzymatic activity, to detailed gene expression dynamics. Focusing on the genes for biomass deconstruction, the *comm1*-derived microbiome (SDM1) had higher levels of expression of the genes encoding pectin and hemicellulose deconstructing enzymes, with the peak of gene expression activity occurring during the initial time (day 2) and most genes being expressed by the *Firmicutes*. We interpret this pattern as an initial deconstruction of the primary cell wall, which continues throughout the 2-week cultivation. On day 5, there was increased expression of the predicted lignin-degrading gene from the *Gemmatimonadetes* bin10 population, consistent with the commencement of deconstruction of the secondary cell wall, and the relative expression level was higher in SDM3, the more active set of cultures. This increased expression was also mirrored in the aromatic catabolic genes expressed by *Filomicrobium* bin10, the majority of which were expressed from day 2 to day 7 at higher levels in SDM3. The cellulase genes, especially genes encoding GH6, GH9, GH48, and AA10, are expressed by *Actinobacteria* (*Salinospora*, *Actinopolymopha*, *Thermocrispum*) later in the cultivation (day 9 to day 14) and at higher levels in SDM3. SDM1 and SDM3 form two separate clusters of cellulase gene expression, indicating that these activities are distinct between the two communities. This distinction is also seen in the network analysis, where *Salinospora* bin32 (SDM1) and *Salinospora* bin64 (SDM3) are peripheral and divergent members of the network, suggesting the response to cellulose has less overlap between the two communities compared to the other cell wall polymers. The increases in gene expression are consistent with biochemical measurements which show SDM3 has higher cellulase activity.

The observation of higher cellulase activity, which arises from the actinobacterial populations, may explain the increased biomass deconstruction observed in SDM3 communities. The overall pattern of community dynamics, with *Firmicutes* being active at early timepoints and Actinobacteria active and later timepoints, mirrors the dynamics observed during composting [63].

Conclusions

The work reinforces the observation that subtle differences in community composition and the genomic content of strains may lead to significant differences in community performance [64]. These considerations should be accounted for in using microbiomes for biotechnology and building synthetic microbiomes [65].

Abbreviations
WT: Wild type; SM: Stacked mutant; Bmr: Brown midrib.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-022-01377-x.

> **Additional file 1: Supplementary Fig. 1.** Residual dry weight biomass after a 56-day incubation. The figure shows that three incubations diverge in terms of biomass degradation. **Supplementary Fig. 2.** Phylogenetic tree reconstructed based on 49 core universal genes defined by Clusters of Orthologous Groups (COGs) using maximum likelihood. The tree was reconstructed on KBase. **Supplementary Fig. 3.** Gene proportion per MAG for selected GHs. MAGs with a completeness above 30% and contamination level lower than 10% are included in the figure. **Supplementary Fig. 4.** A) Ordination biplot depicting the trajectory of metatranscriptomes for the adapted communities growing on SM and WT sorghum. The ellipses were calculated around barycenters with a confidence level of 0.99 using the *stat_conf_ellipse* function in ggpubr v.0.2.4 B) Average TPM-normalized transcriptome abundance per MAG over the 14-day incubation.
>
> **Additional file 2: Supplementary Table S1.** MG_FC_annotations. Table with data for contigs, predicted genes, and feature count values for genes in the metagenomes. MG_metadata. Metadata for sequenced metagenomes. MT_FC_annotations. Table with data for contigs, predicted genes, and feature count values for genes in the metatranscriptomes. MT_metadata. Metadata for sequenced metatranscriptomes.

Authors' contributions
S.W.S and J.C.-N. designed the experiments; L.M.T., M.A., K.D., Y.G., B.G.R., and A.E. performed the experiments; L.M.T., M.A., Y.-W.W, N.X., and J.C.-N. performed the data analysis; C.L., J.C.M., P.D.A., T.R.N., H.S., and B.A.S. supervised the research; L.M.T., M.A., J.C.-N., and S.W.S wrote the manuscript. The authors approved the final manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Joint BioEnergy Institute, Emeryville, CA, USA. [2] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3] Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 110, Taiwan. [4] Environmental Systems and Genome Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5] Institute of Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA. [6] College of Natural Sciences, University of Puerto Rico, Rio Piedras, Puerto Rico. [7] School of Agriculture, Food and Wine, & Waite Research Institute, University of Adelaide, Glen Osmond, SA, Australia. [8] Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [9] Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA. [10] Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA, USA. [11] Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA. [12] Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA.

References
1. Cosgrove DJ, Jarvis MC. Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci. 2012;3:204. https://doi.org/10.3389/fpls.2012.00204.
2. Peralta-Yahya PP, Zhang F, Del Cardayre SB, Keasling JD. Microbial engineering for the production of advanced biofuels. Nature. 2012;488:320–8. https://doi.org/10.1038/nature11478.
3. Stephanopoulos G. Challenges in engineering microbes for biofuels production. Science. 2007;315:801–4. https://doi.org/10.1126/science.1139612.
4. Kerr RA. Global warming is changing the world. Science. 2007;316:188–90. https://doi.org/10.1126/science.316.5822.188.
5. Arfi Y, Chevret D, Henrissat B, et al. Characterization of salt-adapted secreted lignocellulolytic enzymes from the mangrove fungus *Pestalotiopsis* sp. Nat Commun. 2013;4:1–9. https://doi.org/10.1038/ncomms2850.
6. Straub CT, Khatibi PA, Wang JP, et al. Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. Nat Commun. 2019;10:1–6. https://doi.org/10.1038/s41467-019-11376-6.
7. Blifernez-Klassen O, Klassen V, Doebbe A, et al. Cellulose degradation and assimilation by the unicellular phototrophic eukaryote *Chlamydomonas reinhardtii*. Nat Commun. 2012;3:1–9. https://doi.org/10.1038/ncomms2210.
8. Eibinger M, Sattelkow J, Ganner T, et al. Single-molecule study of oxidative enzymatic deconstruction of cellulose. Nat Commun. 2017;8:1–7. https://doi.org/10.1038/s41467-017-01028-y.
9. Himmel ME, Ding S-Y, Johnson DK, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315:804–7. https://doi.org/10.1126/science.1137016.
10. Tan TC, Kracher D, Gandini R, et al. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. Nat Commun. 2015;6:1–11. https://doi.org/10.1038/ncomms8542.
11. Ceja-Navarro JA, Karaoz U, Bill M, et al. Gut anatomical properties and microbial functional assembly promote lignocellulose deconstruction and colony subsistence of a wood-feeding beetle. Nat Microbiol. 2019;4. https://doi.org/10.1038/s41564-019-0384-y.
12. Gharechahi J, Vahidi MF, Bahram M, et al. Metagenomic analysis reveals a dynamic microbiome with diversified adaptive functions to utilize high lignocellulosic forages in the cattle rumen. ISME J. 2021;15:1108–20. https://doi.org/10.1038/s41396-020-00837-2.
13. Aulitto M, Tom LM, Ceja-Navarro JA, et al. Whole-Genome Sequence of Brevibacillus borstelensis SDM, Isolated from a Sorghum-Adapted Microbial Community. Microbiol Resour Announc. 2020;9:15–16. https://doi.org/10.1128/MRA.01046-20.
14. Romero Victorica M, Soria MA, Batista-García RA, et al. Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes. Sci Rep. 2020;10:3864. https://doi.org/10.1038/s41598-020-60850-5.
15. Silva PC, Ceja-Navarro JA, Azevedo F, et al. A novel d-xylose isomerase from the gut of the wood feeding beetle *Odontotaenius disjunctus* efficiently expressed in Saccharomyces cerevisiae. Sci Rep. 2021;11:4766. https://doi.org/10.1038/s41598-021-83937-z.
16. Jiménez DJ, Dini-Andreote F, DeAngelis KM, et al. Ecological insights into the dynamics of plant biomass-degrading microbial consortia. Trends Microbiol. 2017;25:788–96. https://doi.org/10.1016/j.tim.2017.05.012.
17. Kolinko S, Wu Y-W, Tachea F, et al. A bacterial pioneer produces cellulase complexes that persist through community succession. Nat Microbiol. 2018;3:99–107. https://doi.org/10.1038/s41564-017-0052-z.
18. Peng X, Wilken SE, Lankiewicz TS, et al. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. Nat Microbiol. 2021;6:499–511. https://doi.org/10.1038/s41564-020-00861-0.
19. Rodríguez Amor D, Dal Bello M. Bottom-up approaches to synthetic cooperation in microbial communities. Life. 2019;9:22. https://doi.org/10.3390/life9010022.
20. Santos J, Pascual M, Simões P, et al. From nature to the laboratory: the impact of founder effects on adaptation. J Evol Biol. 2012;25:2607–22. https://doi.org/10.1111/jeb.12008.
21. Carlström CI, Field CM, Bortfeld-Miller M, et al. Synthetic microbiota reveal priority effects and keystone strains in the *Arabidopsis* phyllosphere. Nat Ecol Evol. 2019;3:1445–54. https://doi.org/10.1038/s41559-019-0994-z.
22. Sattler SE, Funnell-Harris DL, Pedersen JF. Efficacy of singular and stacked brown midrib 6 and 12 in the modification of lignocellulose and grain chemistry. J Agric Food Chem. 2010;58:3611–6. https://doi.org/10.1021/jf903784j.
23. Eichorst SA, Varanasi P, Stavila V, et al. Community dynamics of cellulose-adapted thermophilic bacterial consortia. Environ Microbiol. 2013;15:2573–87. https://doi.org/10.1111/1462-2920.12159.
24. King BC, Donnelly MK, Bergstrom GC, et al. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. Biotechnol Bioeng. 2009;102:1033–44. https://doi.org/10.1002/bit.22151.
25. Ing N, Deng K, Chen Y, et al. A multiplexed nanostructure-initiator mass spectrometry (NIMS) assay for simultaneously detecting glycosyl hydrolase and lignin modifying enzyme activities. Sci Rep. 2021;11:1–9. https://doi.org/10.1038/s41598-021-91181-8.
26. Gladden JM, Allgaier M, Miller CS, et al. Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. Appl Environ Microbiol. 2011;77:5804–12. https://doi.org/10.1128/AEM.00032-11.
27. Deng K, Takasuka TE, Heins R, et al. Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). ACS Chem Biol. 2014;9:1470–9. https://doi.org/10.1021/CB5000289/SUPPL_FILE/CB5000289_SI_001.PDF.

28. Barnes W, Anderson C. Acetyl bromide soluble lignin (ABSL) assay for total lignin quantification from plant biomass. Bio Protoc. 2017;7. https://doi.org/10.21769/BioProtoc.2149.

29. DeAngelis KM, Gladden JM, Allgaier M, et al. Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulo-lytic microbial communities. Bioenergy Res. 2010;3:146–58. https://doi.org/10.1007/s12155-010-9089-z.

30. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016;18:1403–14. https://doi.org/10.1111/1462-2920.13023.

31. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterio-plankton. Aquat Microb Ecol. 2015;75:129–37. https://doi.org/10.3354/ame01753.

32. Ceja-Navarro JA, Vega FE, Karaoz U, et al. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. Nat Commun. 2015;6:7618. https://doi.org/10.1038/ncomms8618.

33. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461.

34. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72. https://doi.org/10.1128/AEM.03006-05.

35. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 2003;00. https://doi.org/10.1002/0471250953.bi0203s00.

36. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.

37. McMurdie PJ, Holmes S. phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217. https://doi.org/10.1371/journal.pone.0061217.

38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.

39. Prjibelski A, Antipov D, Meleshko D, et al. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70. https://doi.org/10.1002/cpbi.102.

40. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagen-omic sequences. Nucleic Acids Res. 2010;38:e132. https://doi.org/10.1093/nar/gkq275.

41. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30. https://doi.org/10.1093/nar/gkt1223.

42. Yin Y, Mao X, Yang J, et al. dbCAN: a web resource for automated carbo-hydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51. https://doi.org/10.1093/nar/gks479.

43. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366–8. https://doi.org/10.1038/s41592-021-01101-x.

44. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7. https://doi.org/10.1093/bioinformatics/btv638.

45. Huson DH, Beier S, Flade I, et al. MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol. 2016;12:e1004957. https://doi.org/10.1371/journal.pcbi.1004957.

46. Palumbo A, Baumohl J, Best A, et al. KBase: an integrated knowledgebase for predictive biology and environmental research. In: Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP); 2014. p. 1.

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

48. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformat-ics. 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656.

49. Aitchison J. Principles of compositional data analysis. Lect Notes-Mono-graph Ser. 1994:73–81. https://doi.org/10.1214/lnms/1215463786.

50. Zhou J, Deng Y, Luo F, et al. Functional molecular ecological networks. MBio. 2010;1. https://doi.org/10.1128/mBio.00169-10.

51. Zhou J, Deng Y, Luo F, et al. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. MBio. 2011;2. https://doi.org/10.1128/mBio.00122-11.

52. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504. https://doi.org/10.1101/gr.1239303.

53. Godin B, Nagle N, Sattler S, et al. Improved sugar yields from biomass sorghum feedstocks: comparing low-lignin mutants and pretreat-ment chemistries. Biotechnol Biofuels. 2016;9. https://doi.org/10.1186/s13068-016-0667-y.

54. Mewis K, Lenfant N, Lombard V, Henrissat B. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. Appl Environ Microbiol. 2016;82:1686–92. https://doi.org/10.1128/AEM.03453-15/FORMAT/EPUB.

55. Forsberg Z, Sørlie M, Petrović D, et al. Polysaccharide degradation by lytic polysaccharide monooxygenases. Curr Opin Struct Biol. 2019;59:54–64. https://doi.org/10.1016/j.sbi.2019.02.015.

56. Brown ME, Chang MC. Exploring bacterial lignin degradation. Curr Opin Chem Biol. 2014;19:1–7. https://doi.org/10.1016/j.cbpa.2013.11.015.

57. Akram F, Ashraf S, Haq IU, et al. Eminent industrial and biotechnologi-cal applications of laccases from bacterial source: a current overview. Appl Biochem Biotechnol. 2022;194:2336–56. https://doi.org/10.1007/s12010-021-03781-9.

58. D'haeseleer P, Gladden JM, Allgaier M, et al. Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. PLoS One. 2013;8:e68465. https://doi.org/10.1371/journal.pone.0068465.

59. Chen CY, Hsieh ZS, Cheepudom J, et al. A 24.7-kDa copper-containing oxidase, secreted by *Thermobifida fusca*, significantly increasing the xyla-nase/cellulase-catalyzed hydrolysis of sugarcane bagasse. Appl Microbiol Biotechnol. 2013;97:8977–86. https://doi.org/10.1007/s00253-013-4727-y.

60. Davidi L, Moraïs S, Artzi L, et al. Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulo-some. Proc Natl Acad Sci. 2016;113:10854–9. https://doi.org/10.1073/pnas.1608012113.

61. Bittleston LS, Gralka M, Leventhal GE, et al. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. Nat Commun. 2020;11:1–10. https://doi.org/10.1038/s41467-020-15169-0.

62. Flint HJ, Scott KP, Duncan SH, et al. Microbial degradation of complex carbohydrates in the gut. Gut Microbes. 2012;3:289–306. https://doi.org/10.4161/gmic.19897.

63. Antunes LP, Martins LF, Pereira RV, et al. Microbial community structure and dynamics in thermophilic composting viewed through metagenom-ics and metatranscriptomics. Sci Rep. 2016;6:38915. https://doi.org/10.1038/srep38915.

64. Thompson AW, Turkarslan S, Arens CE, et al. Robustness of a model micro-bial community emerges from population structure among single cells of a clonal population. Environ Microbiol. 2017;19:3059–69. https://doi.org/10.1111/1462-2920.13764.

65. Strous M, Sharp C. Designer microbiomes for environmental, energy and health biotechnology. Curr Opin Microbiol. 2018;43:117–23. https://doi.org/10.1016/j.mib.2017.12.007.

Publisher's Note