# Metatranscriptomic and thermodynamic insights into medium-chain fatty acid production using an anaerobic microbiome

*Matthew J. Scarborough[1,2], Christopher E. Lawson[2], Joshua J. Hamilton[3,4], Timothy J. Donohue[1,3], and Daniel R. Noguera[1,2*]*

Running Title: MCFA Producing Anaerobic Microbiome

[1]The Great Lakes Bioenergy Research Center
University of Wisconsin – Madison
Madison, WI
[2]Department of Civil and Environmental Engineering
University of Wisconsin – Madison
Madison, WI
[3]Department of Bacteriology
University of Wisconsin – Madison
Madison, WI
[4]Department of Biochemistry
University of Wisconsin – Madison
*Corresponding author e-mail: dnoguera@wisc.edu

**Abstract**

Biomanufacturing from renewable feedstocks represents a sustainable strategy to offset fossil fuel-based chemical production. One potential biomanufacturing strategy is production of medium chain fatty acids (MCFA) from organic feedstocks using either pure cultures or microbiomes. While the set of microbes in a microbiome can often metabolize more diverse organic materials than a single species, little is known about the carbon and energy flow within and between organisms in these communities. Here, we integrate metagenomic, metatranscriptomic, and thermodynamic analyses to predict and characterize the metabolic network of an anaerobic microbiome producing MCFA from organic matter derived from lignocellulosic ethanol fermentation conversion residue. A total of 37 high quality metagenome-assembled genomes (MAGs) were recovered from the microbiome and metabolic reconstruction of the 10 most abundant MAGs, representing 91% of the total metagenomic reads, was performed. Metabolic reconstruction combined with metatranscriptomic analysis predicted that organisms affiliated with *Lactobacillus* and Coriobacteriaceae degraded carbohydrates and fermented sugars to lactate and acetate. Lachnospiraceae and Eubacteriaceae affiliated organisms were predicted to transform these fermentation products to MCFA. Additionally, a previously unknown role of $H_2$ production and consumption on MCFA production is predicted from our transcriptomic and thermodynamic analyses of this microbiome. From these findings, we propose new strategies for the rational development of microbiomes that produce MCFA from lignocellulosic-based feedstocks and possibly other mixed organic nutrient streams.

46 **Importance**

47 Mixed communities of microbes play important roles in health, the environment,
48 agriculture, and biotechnology. While tapping the combined activity of organisms within
49 microbiomes may have advantages over the use of pure cultures for biomanufacturing,
50 harnessing the metabolism of these mixed cultures remains a major challenge. Here, we
51 predict metabolic functions of bacteria in a microbiome that produces medium-chain fatty
52 acids from a renewable feedstock. Our approach and findings can be used to engineer and
53 control microbiomes for improved biomanufacturing; to build synthetic mixtures of
54 microbes that produce valuable chemicals from renewable resources; and to better
55 understand microbial communities that contribute to health, agriculture, and the
56 environment.

57 **Introduction**

58 Biological production of chemicals from renewable resources is an important step to
59 reduce societal dependence on fossil fuels. One approach that shows potential for the
60 biological production of chemicals from renewable resources, the carboxylate platform (1,
61 2), uses anaerobic microbial communities. Members of these microbial communities
62 hydrolyze complex organics to products that can be fermented and subsequently converted
63 to carboxylic acids, including medium-chain fatty acids (MCFA). MCFA, such as hexanoate
64 (a six carbon monocarboxylate, C6) and octanoate (an eight carbon monocarboxylate, C8),
65 are used in large quantities for the production of pharmaceuticals, antimicrobials, and
66 industrial materials, and can be processed to chemicals currently derived from fossil fuels
67 (3, 4). While there is growing interest in harnessing microbiomes for renewable chemical
68 manufacturing, little is known about the specific biochemical transformations carried out
69 by members of these microbiomes. We are interested in obtaining the knowledge needed
70 for the rational development and implementation of strategies to improve yields of MCFA
71 or other products from renewable carbon sources using engineered microbiomes.

72 To obtain this knowledge, we dissected the flow of substrates, intermediates, and
73 products in a carboxylate platform microbiome. Previous applications of the carboxylate
74 platform have focused on converting organics from undistilled corn beer (5, 6), food (7, 8)
75 or winery residue (9), and lignocellulose-derived materials (10-12) to MCFA. In addition,
76 production of MCFA from conversion residues using anaerobic microbiomes can improve
77 the economic sustainability of lignocellulosic biofuel production (4). In each of these
78 studies, it is proposed that MCFA precursors, resulting from hydrolysis and fermentation of
79 available organic substrates, are elongated via one or more chain-elongating

80 microorganisms (1). In addition, MCFA producing bioreactors contain diverse communities
81 with members affiliated with the Firmicutes, particularly Clostridia and *Lactobacilli*, having
82 high abundance (4, 5, 11). However, the specific roles of the abundant microbiome
83 members are not well understood.

84 Here, we utilize a combination of metagenomic, metatranscriptomic , pathway
85 modeling, and thermodynamic analyses to reconstruct the combined metabolic activity of a
86 microbiome converting lignocellulosic conversion residues to C6 and C8. In total, 37 high
87 quality draft metagenome-assembled genomes (MAGs) were assembled, and the gene
88 expression patterns of those representing the ten most abundant community members
89 during steady-state reactor operation were analyzed. Our results identify several MAGs in
90 the community that expressed genes predicted to be involved in complex carbohydrate
91 degradation and in the subsequent fermentation of degradation products to lactate and
92 acetate. Genes encoding enzymes capable of producing C6 and C8 from these fermentation
93 products were expressed by two abundant MAGs affiliated with the class Clostridia. Based
94 on a thermodynamic analysis of the proposed MCFA-producing pathways, we predict that
95 individual Clostridial MAGs use different substrates for MCFA production (lactate, versus a
96 combination of xylose, $H_2$, and acetate). We also describe how new insights into MCFA
97 production could improve the biomanufacturing productivity of microbiomes.

98 **Results**

99 **Microbiome characterization**

100 We previously described the establishment of a microbiome that produces MCFA in
101 a bioreactor that is continuously fed with the residues from lignocellulosic ethanol
102 production (4). The reactor feed contained high amounts of xylose, carbohydrate

103 oligomers, and uncharacterized organic matter (Fig. 1). Following an initial acclimation
104 period (Day 0 – Day 48), the quantity of produced organic acids stabilized (Fig. 1). To gain
105 insight into the microbial activities that were associated with this MCFA-producing
106 microbiome, samples were collected for metagenomic analysis at five different times (Days
107 12, 48, 84, 96, and 120), and RNA was prepared for metatranscriptomic analysis at Day 96.
108 At the time of metatranscriptomic sampling, the bioreactor converted 16.5% of the organic
109 matter (measured as chemical oxygen demand (COD)) in conversion residue to C6 and C8.
110 From the metagenomic samples, a total of 219 million DNA reads were assembled
111 and binned, resulting in 37 high quality MAGs (Supplement 1). For the Day 96 sample, 91%
112 of the DNA reads mapped to the ten most abundant MAGs (Supplement 2). Based on
113 analysis with CheckM, these abundant MAGs had high completion with low contamination
114 (Table S1)(13). Relative abundance of the top 10 MAGs was calculated from the total DNA
115 reads mapped to the MAGs and normalized to their predicted genome length (Fig. 2;
116 Supplement 2). The most abundant MAGs include a Lachnospiraceae (LCO1, 50%), a
117 *Lactobacillus* (LAC1, 30%), a Coriobacteriacea (COR1, 6.3%), and a Eubacteriaceae (EUB1,
118 6.0%). Four additional *Lactobacilli* and two additional Coriobacteriacea are also predicted
119 to be within the 10 most abundant MAGs (Fig. 2).
120 The metatranscriptome data, obtained from the Day 96 sample, contained 87
121 million cDNA reads. After quality checking and removal of rRNA sequences, 82.6 million
122 predicted transcript reads were used for mapping to MAGs. Of these, 85.3% of the
123 predicted transcripts (hereafter referred to as transcripts or mRNA) mapped back to the 10
124 most abundant MAGs (Supplement 2). Relative expression was calculated from the total
125 filtered mRNA mapped to the MAGs and normalized to the predicted genome length of

126 these bacteria (Fig. 2). MAGs with the highest levels of transcripts included LAC1 (60%),
127 EUB1 (12%), LCO1 (11%), and COR1 (6.3%), which also displayed high abundance in the
128 metagenome. Whereas LCO1 was most abundant based on DNA reads, LAC1 appeared to
129 have the highest activity based on transcript levels.

130 A phylogenetic tree of the ten most abundant MAGs was constructed based on
131 concatenated amino acid sequences of 37 single-copy marker genes (Fig. 3). All Bacilli
132 MAGs (LAC1, LAC2, LAC3, LAC4, and LAC5) clustered with *Lactobacillus.* Average
133 nucleotide identity (ANI) calculations with other *Lactobacillus* genomes were above the 95-
134 96% ANI suggested for species demarcation (14), indicating that these MAGs represent
135 strains of established *Lactobacillus* species (Supplement 3). Clostridia EUB1 clustered with
136 *Pseudoramibacter alactolyticus.* The ANI calculations for this MAG with *P. alactolyticus* and
137 related members of the Eubacteriaceae indicated that EUB1 could represent a new genus
138 within the Eubacterium (Supplement 3). The Clostridia LCO1 did not cluster with a specific
139 genus; ANIs of the LCO1 MAG and related organisms suggest this MAG could represent a
140 novel genus within the Lachnospiraceae (Supplement 3). The three Actinobacteria MAGs,
141 COR1, COR2, and COR3, clustered within the Coriobacteriaceae family. One of these MAGs
142 (COR2) clustered with *Olsenella umbonata* (15), but the ANI calculation did not support this
143 MAG being a representative of the *Olsenella* genus (Supplement 3). The other two MAGs
144 (COR1 and COR3) formed their own cluster within the Coriobacteriacea, but were
145 sufficiently different in ANI calculations to suggest that each represents new genera within
146 the Coriobacteriaceae.

147 **Genomic predictions of chemical transformations in the microbiome**

148 A prediction of metabolic networks in the microbiome was performed by mapping
149 mRNA reads to the open reading frames (ORFs) within each of the ten most abundant
150 bacteria. Metabolic reconstruction was performed with automated prediction algorithms
151 (16) and manual curation, particularly of proposed sugar utilization, fermentation and
152 chain elongation pathways (Supplement 4) (1). This analysis identified a set of genes that
153 also had high expression levels in the metatranscriptome that could be used to model the
154 metabolic potential of the microbiome to [1] degrade complex carbohydrates remaining in
155 ethanol conversion residue; [2] transform simple sugars into the fermentation products
156 acetate, lactate, and ethanol; and [3] produce butyrate (C4), C6, and C8 from sugars and
157 fermentation products. The predictions for each of these processes are summarized below.
158 ***Degradation of complex carbohydrates.*** Carbohydrates were a large portion of the
159 organic substrates present in the ethanol conversion residue fed to the bioreactor, and
160 uncharacterized carbohydrates . Quantitative analyses indicated that xylose was the most
161 abundant monosaccharide in the residue, accounting for 22% of the organic matter.
162 Glucose was undetected in most samples or a minor component, and other carbohydrates
163 corresponded to 20% of the organic matter in the residue (Fig. 1). Based on the
164 characterization of the bioreactor performance throughout the operational period,
165 approximately 40% of the uncharacterized carbohydrates were being degraded at the time
166 the metatranscriptomic samples were obtained (Day 96; Fig. 1).
167 To investigate the expression of genes related to degradation of complex
168 carbohydrates, we analyzed the predicted MAG ORFs using the carbohydrate-active
169 enzyme (CAZyme) database (17). Of particular interest was production of predicted
170 extracellular enzymes that hydrolyze glycosidic bonds in complex carbohydrates, as these

171 may release sugars that can be subsequently metabolized by community members that do
172 not express complex carbohydrate degrading enzymes. The subcellular localization
173 software, CELLO, was used to predict whether individual CAZyme proteins were located
174 within the cytoplasm or targeted to the extracellular space (Supplement 5) (18).
175 This analysis showed that transcripts encoding genes for several types of glycoside
176 hydrolases (GHs) were abundant in several MAGs in the microbiome (Fig. 4). All LAC MAGs
177 expressed genes encoding extracellular CAZymes that cleave glycosidic bonds between
178 hexose and pentose moieties in xylans. In particular, LAC1 LAC2, and LAC4 expressed
179 genes that encode several extracellular exo-β-xylosidases that could remove terminal
180 xylose molecules from xylans present in the conversion residue (GH43 and GH120; Fig. 4;
181 Supplement 5). LAC2 also had high levels of transcripts for an exo-α-L-1,5-arabinanase
182 (GH93), predicted to release other pentose sugars from arabinan, which accounts for 3% of
183 the sugar polymers in switchgrass (19, 20). In addition, the COR1, COR3 and LAC4
184 members of the community had high transcript levels for three extracellular CAZymes
185 (GH13) that are predicted to degrade a variety of glucans that may be remaining in
186 switchgrass conversion residue (21). In sum, this analysis predicts that glucans are
187 degraded by *Lactobacillus* and Coriobacteriaceae MAGs, where LAC MAGs may also degrade
188 xylans and arabinans. It further suggests that this microbiome is capable of releasing
189 oligosaccharides and sugar monomers from glucans, xylans, and arabinans, the primary
190 components of switchgrass and other plant biomass.
191 Bacterial oligosaccharide hydrolysis can also occur in the cytoplasm. All MAGs in
192 this microbiome contained predicted cytoplasmic GH13 enzymes, which are known to
193 degrade hexose oligosaccharides. The microbiome also contained high transcript levels for

194 genes encoding predicted cytoplasmic CAZYmes that degrade maltose (GH4, GH65), a
195 glucose dimer that may result from extracellular breakdown of glucans (Fig. 4). Transcripts
196 encoding known or predicted cytoplasmic β-glucosidases (GH1, GH3) and β-galactosidases
197 (GH2) are found across the MAGs (Fig. 4). In addition, transcripts that encode β-xylosidases
198 (GH1, GH3) and α-L-arabinofuranosidases (GH2), are found in all the LAC MAGs except
199 LAC3 (Fig. 4). Based on the metatranscriptomic analysis, other cytoplasmic CAZymes
200 predicted to hydrolyze pentose-containing oligosaccharides are predicted to be expressed
201 by the LAC1, LAC2, LAC4, and LAC5 members of this microbiome (Fig. 4).
202 ***Transport and production of simple fermentation products from sugars.*** Simple sugars
203 are abundant in ethanol conversion residue and produced during complex carbohydrate
204 hydrolysis. Sugars are therefore expected to be a major substrate for the microbiome.
205 Despite the use of a yeast strain that was engineered for improved xylose utilization in the
206 ethanol fermentation, xylose was the major abundant monosaccharide present in the
207 remaining conversion residue (Fig. 1). As discussed above, the relative transcript levels of
208 genes encoding extracellular GHs (Fig. 4) by several MAGs in the microbiome predict that
209 additional pentoses and hexoses may be released through degradation of complex
210 carbohydrates.
211 We therefore analyzed the genomic potential of the community to transport sugars, and
212 metabolize them to fermentation products, particularly the known MCFA precursors
213 lactate, acetate, and ethanol. To investigate the ability of the community to transport
214 sugars, MAG ORFs were annotated with the Transporter Classification Database
215 (Supplement 6). Expression of genes associated with the pentose phosphate pathway,

216 phosphoketolase pathways, and glycolysis (Fig. 5a) was analyzed to predict the potential
217 for sugar metabolism within individual MAGs.
218 This analysis found that transcripts from genes encoding predicted carbohydrate
219 transporters were among the most highly abundant mRNAs across the microbiome. These
220 putative transporter proteins belonged to a variety of families, including many associated
221 with the ATP-binding cassette (ABC) superfamily, and the phosphotransferase system
222 (PTS) family (TD 4.A.-) (Fig. S1). LCO1, LAC1, LAC2, and LAC3 are predicted to contain
223 xylose transporters (XylT) (Fig. 5a), while glucose (GluT), fructose (FruT), and other hexose
224 transporters were expressed across the LAC, COR, and LCO MAGs (Fig. 5a; Fig. S1). EUB1
225 only encoded transcripts encoding carbohydrate transporters for uptake of fructose and
226 sucrose (Fig. S1). Overall, this analysis predicts that all MAGs have the potential to utilize
227 hexose sugars, while gene expression patterns observed for the LCO1 and the *Lactobacillus*
228 MAGs (excluding LAC3) predicted that they play a major role in pentose utilization in this
229 microbiome.
230 We also analyzed the metatranscriptomic data to investigate potential routes for sugar
231 metabolism. Once transported to the cytoplasm, glucose can be phosphorylated with
232 hexokinase (HK) and converted to fructose-6-phosphate (F-6-P) by glucose-6-phosphate
233 isomerase (GI). Transcripts encoding predicted HK and GI enzymes are abundant for all
234 MAGs within the microbiome (Fig. 5a), except LAC5, which does not encode known
235 homologues of these proteins. Fructose utilization starts with phosphorylation during
236 transport (Fig. 5a). Fructose-6-phosphate (F-6-P) is either phosphorylated to fructose-1,6-
237 bisphosphate (F-1,6-BP) by phosphofructokinase (PFK) in glycolysis or is cleaved to acetyl-
238 P (Ac-P) and erythrose-4-P (E-4-P) by phosphoketolase (PK). While LAC1, LAC2, LAC4,

239 LAC5 and COR3 all lack homologues of genes encoding PFK (a highly-conserved glycolysis
240 enzyme known to be a major target for regulatory control in hexose utilization) (22), they
241 all contain transcripts for homologues of PK (Fig. 5a). In sum, these analyses predict that all
242 of the abundant MAGs in this microbiome can utilize hexoses that may be produced during
243 hydrolysis of complex oligosaccharides.

244 Transcripts predicted to encode enzymes to convert xylose to xylulose-5-phosphate,
245 xylose isomerase (XI) and xylulose kinase (XK) (23), were abundant in most of the
246 *Lactobacillus* MAGs and LCO1, and either absent or showed very low abundance in LAC3,
247 EUB1 and the COR MAGs (Fig. 5a). Once produced, xylulose-5-P can be degraded through
248 either the phosphoketolase pathway or the pentose phosphate pathway. Transcripts from a
249 gene predicted to encode the diagnostic enzyme of the phosphoketolase pathway,
250 phosphoketolase (PK), which splits xylulose-5-P (X-5-P) into acetyl-P (Ac-P) and
251 glyceraldehyde-3-P (G-3-P), were among the most abundant mRNAs in the *Lactobacillus*
252 MAGs and is also present at high levels in LCO1 (Fig. 5a; Supplement 4). LCO1 and LAC1
253 also contained transcripts from homologues of all of the genes needed for the pentose
254 phosphate pathway (R5PE, R5PI, TA, TK in Fig. 5a). Overall, this analysis predicted that
255 multiple routes of pentose utilization could be utilized by the MAGs in this microbiome.

256 The predicted routes for both hexose and xylose metabolism in this microbiome lead to
257 pyruvate production (Fig. 5a), so we also analyzed how this and other fermentation
258 products might lead to MCFA production in this community. All MAGs contained transcripts
259 encoding lactate dehydrogenase homologues (LDH) (Fig. 5a), an enzyme which reduces
260 pyruvate to lactate. Transcript analysis also predicts that all of the MAGs (except LAC3) can
261 oxidize pyruvate to acetyl-CoA, utilizing either pyruvate dehydrogenase (PDH) or pyruvate

262 flavodoxin oxidoreductase (PFOR) (Fig. 5a). All MAGs (except EUB1) contain transcripts
263 encoding homologues of acetate kinase (ACK), which converts acetyl-phosphate (Ac-P) to
264 acetate while producing ATP (Fig. 5a). Based on predictions of the gene expression data,
265 the COR and LAC MAGs are also able to convert acetyl-CoA (Ac-CoA) to ethanol with
266 aldehyde dehydrogenase (ADA) and alcohol dehydrogenase (ADH). In summary, analysis of
267 the gene expression patterns in the conversion residue microbiome predict that the MAGs
268 in the LCO, LAC and COR ferment sugars to acetate and lactate, while the LAC and COR
269 members produce ethanol as an additional fermentation product.
270 ***Elongation of fermentation products to MCFAs.*** Based on the above findings, we analyzed
271 the microbiome gene expression data to predict which members of the microbiome had the
272 potential for conversion of predicted fermentation products to MCFA. The Clostridia (LCO1
273 and EUB1) are the only MAGs that contained genes encoding homologues of genes known
274 to catalyze chain elongation reactions in the reverse β-oxidation pathway (Fig. 5b). Thus,
275 the subsequent analysis is based on the prediction that only LCO1 and EUB1 are the major
276 producers of MCFA in this microbiome.
277 Acetate, lactate and ethanol, are all fermentation products that would require
278 transformation to acetyl-CoA before being used as a substrate for elongation by the reverse
279 β-oxidation pathway. Acetate could be converted to acetyl-CoA utilizing ATP via acetyl-CoA
280 synthase (ACS) or the ACK and phosphate acetyltransferase (PTA) route (Fig. 5b).
281 Alternatively, acetate can be converted to acetyl-CoA with a CoA transferase (CoAT) which
282 transfers a CoA from one carboxylic acid to another (e.g., from butyryl-CoA to acetate,
283 producing butyrate and acetyl-CoA) (Fig. 5b). Genes encoding homologues of ACS and ACK
284 were not found in EUB1, but LCO1 contained abundant transcripts that encoded

285 homologues of both ACK and PTA (Fig. 5a). Both MAGs also contained transcripts predicted
286 to encode CoAT enzymes (Fig. 5b). Taken together, this analysis predicts that acetate may
287 be used as a substrate for MCFA production by LCO1 and EUB1.

288 Lactate has been proposed as a key intermediate in other microbiomes producing
289 MCFA (11). While transcripts encoding genes for lactate production are abundant in the
290 microbiome (Fig. 5a), lactate does not accumulate to detectable levels during steady
291 operation, but transiently accumulates when the bioreactor receives a higher load of
292 conversion residue (4). Transcripts for a gene encoding a predicted lactate transporter
293 (LacT) were abundant in EUB1 but not in LCO1, suggesting that only EUB1 may utilize the
294 lactate produced by other MAGs. Neither EUB1 nor LCO1 accumulated transcripts encoding
295 a predicted ADA homologue, which would be required for conversion of acetaldehyde to
296 acetyl-CoA during utilization of ethanol (Fig. 5b). This indicates that if ethanol is produced
297 in this microbiome, it is not used as a significant substrate for MCFA production. Moreover,
298 since ethanol did not accumulate in the reactor during either steady state operation (Fig. 1)
299 or after a high load of conversion residue (4), we predict that ethanol is not a substrate for
300 MCFA production in this microbiome. Rather, based on the predicted activity of LAC and
301 COR MAGs producing lactate and that of EUB1 consuming lactate, we predict that lactate is
302 a key fermentation intermediate for MCFA production.

303 Within the reverse β-oxidation pathway (Fig. 5b), a key enzyme is an electron-
304 bifurcating acyl-CoA dehydrogenase (ACD) containing two electron transfer flavoproteins
305 (EtfA, EtfB) that pass electrons from NADH to ferredoxin (Fig. 5b) (24). Transcripts for
306 genes encoding a homologue of the acyl-CoA dehydrogenase complex (ACD, EtfA, EtfB)
307 were abundant in both LCO1 and EUB1, as are those from other genes predicted to be

308 involved in this pathway (Fig. 5b). Chain elongation by the reverse β-oxidation pathway
309 conserves energy by increasing the ratio of reduced ferredoxin (a highly electropositive
310 electron carrier) to the less electropositive NADH (1). In organisms that use this pathway,
311 oxidation of ferredoxin by the Rnf complex generates an ion motive force, and ATP
312 synthase utilizes the ion motive force to produce ATP (25). We found that transcripts for
313 genes encoding homologues of all six subunits of the Rnf complex were abundant in both
314 EUB1 and LCO1 (RnfABCDEG, Fig. 5b). To maintain cytoplasmic redox balance, reduced
315 ferredoxin could transfer electrons to $H^+$ via hydrogenase, generating $H_2$. LCO1 and EUB1,
316 along with the COR MAGs, contained abundant transcripts for genes predicted to produce
317 ferredoxin hydrogenase (H2ase, Fig. 5b), supporting the hypothesis that $H_2$ production
318 plays a role within this MCFA-producing microbiome. We also looked for two additional
319 hydrogenases known to conserve energy either through the translocation of protons
320 (EchABCDEF Fig. 5b,) or by electron confurcation, utilizing electrons from both NADH and
321 reduced ferredoxin (HydABC, Fig. 5b) (25). It does not appear that these systems play a
322 major role in $H_2$ production in this microbiome since none of the MAGs contained genes
323 encoding homologues of the known components for either of these enzyme complexes (Fig
324 5b).

325 **Thermodynamic analysis of MCFA production in the microbiome**

326 The above analysis predicted several potential routes for MCFA production by LCO1 and
327 EUB1 in this microbiome. To evaluate the implications of these potential chain elongation
328 routes, we used thermodynamic analysis to investigate the energetics of the predicted
329 transformations. For this, we reconstructed metabolic pathways for xylose and lactate
330 conversion, as well as ATP yields based on the data obtained from gene expression

331 analyses (Tables S2-S3, Supplement 7). We used these reconstructions to calculate the free
332 energy changes of the overall biochemical reactions assuming an intracellular pH of 7.0, a
333 temperature of 35°C, and $H_2$ partial pressures of 1.0 x $10^{-6}$, 1.0, and 6.8 atm for low,
334 standard, and high $H_2$ partial pressure, respectively. The low value is the approximate
335 concentration of $H_2$ in water that is in equilibrium with the atmosphere and the high value
336 is an expected maximum in a pressurized mixed culture fermentation system (26). We also
337 compared the efficiency of ATP production to an expected maximum yield of 1 ATP per -60
338 kJ energy generated by the overall chemical transformation (27).

339 Metabolic reconstructions considered xylose (Table S2) and lactate (Table S3) as major
340 substrates for synthesis of C4, C6, and C8 products.. In addition, both LCO1 and EUB1 have
341 the potential to use a CoAT or a thioesterase (TE) as the terminal enzyme of the reverse β-
342 oxidation pathway (Fig. 5b), so we considered both possibilities in the thermodynamic
343 analysis.

344 The use of xylose as the substrate (Table S2, Eqs. 3-8) is possible for LCO1 but not
345 EUB1, since the later MAG lacks genes to transport and activate xylose to xylulose-5-P
346 (XylT, XI, XK, Fig. 5a). Our analysis predicts that, with a pathway containing a terminal
347 CoAT enzyme, the ATP yield (mol ATP $mol^{-1}$ xylose) does not increase if longer chain
348 MCFAs are produced. However, if TE is used for the terminal step of reverse β-oxidation
349 (Eqs. 6-8), the overall ATP yield is lower, but it increases with increasing product length,
350 and C8 production provides a 17% increase in ATP yield versus production of C4. This
351 suggests that LCO1 has no energetic benefit for producing C6 or C8 solely from xylose
352 unless TE is used as the terminal enzyme of reverse β-oxidation. Additionally, the higher
353 ATP yield of xylose conversion to C4 (Eqs. 3 and 6), in comparison to xylose conversion to

354 lactate and acetate by other members of the microbiome (Eq. 2), may explain why LCO1
355 reached higher abundance in the microbiome compared to the other less abundant MAGs
356 (LAC) that are predicted to ferment xylose to lactate and acetate (Fig. 2). In production of
357 C4 and C8, no $H_2$ is predicted to be formed if a CoAT is utilized (Eqs. 3 and 5). On the other
358 hand, if a TE terminal enzyme is utilized for the reverse β-oxidation, $H_2$ is predicted to be
359 produced for all carboxylic acid products.

360 Additional metabolic reconstructions analyzed the co-utilization of xylose with a
361 monocarboxylic acid (Table S2, Eq 9-18). This analysis predicted that co-metabolism of
362 these substrates could provide an energetic advantage (i.e., higher mol ATP per mol of
363 xylose) if $H_2$ is utilized as an electron donor. This suggests that $H_2$, produced by either
364 EUB1 or COR MAGs (H2ase, Fig5a), may be utilized by LCO1 to support MCFA production. If
365 TE is used as the terminal enzyme of reverse β-oxidation (Eqs. 14-18), there is no increase
366 in ATP yield versus utilization of xylose as the sole carbon source (Eqs. 6-8).

367 We also modeled MCFA production from lactate by EUB1, since the gene expression
368 data suggested that EUB1 could transform lactate to MCFA. In models utilizing CoAT (Table
369 S3, Eq. 19-21) as a final step in MCFA production, the ATP yield increases as longer chain
370 MCFA are produced, but the free energy released is near the expected limit for ATP
371 production (27) under conditions of low $H_2$ partial pressure and below this limit at high $H_2$
372 partial pressures (Table S3). If TE is utilized as a final step in MCFA production by EUB1
373 (Eq. 22-24), lower ATP yields are predicted, and in that case the production of longer chain
374 MCFAs has a more pronounced effect on the ATP generated per mol of lactate consumed.
375 For instance, production of C6 results in a 100% increase in the ATP yield compared to
376 producing C4. However, each elongation step reduces the amount of energy released per

377 mol ATP produced, such that production of C8 from lactate results in the release of -58 kJ
378 per ATP produced under high $H_2$ conditions, which is near the expected limits for a cell to
379 conserve chemical energy as ATP. Nevertheless, these models predict that chain elongation
380 by EUB1 using lactate as the electron donor would result in $H_2$ production and that there is
381 an energetic gain if TE is used as the terminal enzyme in reverse β-oxidation.

382 When modeling scenarios utilizing lactate plus carboxylic acids as growth substrates
383 (Table S3, Eqs. 25-36), their elongation by EUB1 would increase the amount of ATP it could
384 produce compared to using lactate only if using a terminal CoAT (Eqs. 19-21). $H_2$
385 production or consumption is not predicted in these scenarios, and the calculated free
386 energy released per mol of ATP produced (-50 to -53 kJ $mol^{-1}$ ATP) is low, near the
387 physiological limit of -60 kJ $mol^{-1}$ ATP for energy conservation by the cell. Models with TE
388 as the terminal enzyme in reverse β-oxidation were also analyzed (Eqs. 31-36) even though
389 EUB1 is not predicted to have this ability as it lacks ACS and ACK needed to utilize acetate
390 (Fig. 5b). In such models, producing C6 and C8 from lactate plus acetate (Eqs. 32-33) is
391 energetically favorable, whereas C4 production is not (Eq. 31).

392

393 **Discussion**

394 The goal of this work was to generate knowledge needed to understand how the combined
395 metabolic activity of predominant members of a microbiome could allow it to produce
396 valuable products from a renewable resource. In this study, we illustrate how combining
397 genomic, computational and thermodynamic predictions can illustrate how a microbial
398 community can convert organic substrates in lignocellulosic conversion residues into
399 MCFA (Fig. 6). Specifically, this approach predicts that the coordinated and step-wise

400 metabolic activity of different members of this microbiome allow cleavage of complex five-
401 and six-carbon containing polysaccharides; conversion of sugars into simple fermentation
402 products; and utilization of sugars and intermediate fermentation products for MFCA
403 production. This approach further predicts the role of intracellular and extracellular
404 reductants in these processes. Below, we illustrate the new insight that has been gained on
405 the activity of a MCFA producing microbiome and how this might relate to other systems.
406 Our data suggest that the contribution of *Lactobacillus* in this microbiome is in
407 extracellular carbohydrate degradation and subsequent metabolism of pentose-containing
408 carbohydrates, while Coriobacteriaceae are predicted to metabolize hexose-containing
409 carbohydrates. The combined metabolic activities of these two MAGs would produce
410 oligosaccharides and monomeric sugars that would become available to these and other
411 members of the microbiome. Metabolic reconstruction combined with microbiome
412 transcript levels also suggest that the *Lactobacillus* and Coriobacteriaceae MAGs produce
413 fermentation end products, primarily lactate and acetate, from these carbohydrates. In
414 addition, microbiome gene expression patterns indicate that two MAGs, EUB1 and LCO1,
415 produce MCFA via reverse β-oxidation. LCO1 is predicted to consume xylose based on gene
416 expression analysis, whereas RNA abundance measurements indicate that EUB1 consumes
417 lactate. Our thermodynamic analysis predicts that the most energetically advantageous
418 metabolism for LCO1 is the consumption of xylose, $H_2$ and carboxylates to produce C4, C6,
419 and C8. While xylose is a major component of conversion residue (Fig. 1), $H_2$ is expected to
420 be produced by Coriobacteriaceae MAGs and EUB1. For EUB1, which is expected to utilize
421 lactate, our analysis predicts that production of MCFA produces higher amounts of ATP,
422 with C6 resulting in a 2-fold increase in ATP production versus producing C2.

423 Predictions from our thermodynamic modeling indicate that C8 production from
424 lactate is energetically advantageous. However, this is at odds with C6 being produced from
425 conversion residue at higher concentrations than C8 (Fig. 1). It is known that C8 is a
426 biocide, so it may be that higher production of this MFCA would inhibit growth of one or
427 more members of the community (11). It has also been shown that removal of C8 allows for
428 higher productivities of carboxylate platform systems (1).

429 $H_2$ production and interspecies $H_2$ transfer are known to have significant impacts on
430 the metabolism of microbial communities (28). Our analysis predicts a previously
431 undescribed role of $H_2$ in supporting chain elongation in a carboxylate platform
432 microbiome. While high $H_2$ partial pressures are proposed to inhibit production of acetate
433 and other carboxylic acids (29, 30), organisms that use the phosphoketolase pathway (the
434 *Lactobacillus* and LCO1 MAGs identified in this study) can produce C2, C4 and C8 without
435 producing $H_2$ (Table S2, Eqs. 2, 3 and 5). While conversion of lactate to MCFA (Table S3,
436 Eqs. 19-24) is predicted to produce $H_2$, other processes such as co-utilization of xylose and
437 a monocarboxylic acid for MCFA production (Table S2, Eq. 9-18) would consume $H_2$.
438 Therefore, $H_2$ accumulation is not expected to limit production of MCFA, although $H_2$ partial
439 pressures may influence the metabolic routes utilized by the microbiome.

440 In considering how to further improve the production of MCFA with a microbiome,
441 additional work is needed to characterize and engineer reverse β-oxidation proteins from
442 the Firmicutes in order to improve production of organic acids longer than C4. Further, our
443 data predict that the terminal enzyme of reverse β-oxidation can influence production of
444 MCFA. While a CoAT enzyme conserves energy, a TE makes production of MCFA more

445 energetically advantageous (Tables S2-S3). Therefore, engineering chain-elongating
446 organisms to only have a TE rather than CoAT may improve production of MCFA.
447 Our metabolic reconstructions predict that lactate was a key fermentation product
448 that supports MCFA production. Therefore, strategies to enhance lactate production and
449 minimize other fermentation products (fermentation of carbohydrates to lactate, rather
450 than acetate or butyrate in this example), could improve production of desired end
451 products. Moreover, designing strategies to enrich a community that produces a critical
452 intermediate like lactate by one pathway (e.g., homofermentative lactate-producing
453 *Lactobacilli*, rather than hetero-fermenters producing both lactate and C2) could improve
454 performance of the microbiome. However, the principles controlling the presence or
455 dominance of heterofermentative versus homofermentative organisms in microbial
456 communities remain largely unexplored. Alternatively, higher production of a desired
457 product, C8, could be achieved by adjusting the abundance or establishing a defined co-
458 culture containing a lactic acid bacterium capable of complex carbohydrate degradation,
459 such as LAC1, and a lactate-elongating organism, such as EUB1. The ability to establish
460 defined synthetic communities, to adjust the abundance of microbiome members or to
461 regulate the metabolic routes within the microbiome may allow more control over the
462 function of a microbiome for production of MCFA or to optimize other traits.
463 In summary, this work demonstrates that one can dissect and model the
464 composition of microbiomes as a way to understand the contribution of different
465 community members to its function. In the case of an anaerobic carboxylate platform
466 microbiome fed lignocellulosic ethanol conversion residue, two Clostridia-related
467 organisms (EUB1 and LCO1) are predicted to be responsible for production of MCFA via

468 reverse β-oxidation. This provides a genome-centric rationale for the previously
469 established correlation between Clostridia-related abundance and MCFA production noted
470 in carboxylate platform systems (4, 11). This study further predicts that the terminal
471 enzyme in product synthesis and the fermentation end products produced by other
472 community members can play a role in determining predominant products of this
473 microbiome. These approaches, concepts and insights should be useful in predicting and
474 controlling MCFA production by reactor microbiomes and in analyzing the metabolic,
475 genomic and thermodynamic factors influencing the function of other microbiomes of
476 health, environmental, agronomic or biotechnological importance.

477 **Methods**

478 **Production of conversion residue.** Switchgrass used to generate conversion residue was
479 treated by ammonia fiber expansion and enzymatically treated with Cellic CTec3® and
480 Cellic HTec3® (Novozymes) to digest cellulose and hemicellulose (to glucose and xylose,
481 primarily) (31). Hydrolysate was fermented with *Saccharomyces cerevisiae* Y128, a strain
482 with improved xylose utilization (32). Fermentation media was distilled to remove ethanol
483 (4).

484 **Bioreactor operation.** The bioreactor was seeded with acid digester sludge from the Nine
485 Springs Wastewater treatment plant in Madison, WI. The retention time of the semi-
486 continuous reactor was maintained at six days by pumping conversion residue into the
487 reactor, pumping reactor effluent from the reactor once per hour, and maintaining a liquid
488 volume of 150mL in the reactor. The reactor was mixed with a magnetic stir bar. The
489 temperature of the reactor was controlled at 35 °C using a water bath, and the pH of the

490 reactor was maintained at 5.5 by feeding 5M KOH through a pump connected to a pH
491 controller. This reactor sustained MCFA production for 252 days (4).
492 **Metabolite analysis.** Samples from the bioreactor and conversion residue were collected
493 for metabolite analysis. All samples were filtered using 0.22 μm syringe filters
494 (ThermoFisher Scientific SLGP033RS, Waltham, MA, USA). Chemical oxygen demand (COD)
495 analysis was performed using High Range COD Digestion Vials (Hach 2125915, Loveland,
496 CO, USA) per standard methods (33). Soluble carbohydrates were measured with the
497 anthrone method (34). Glucose, xylose, acetic acid, formic acid, lactic acid, succinic acid,
498 pyruvic acid, glycerol and xylitol were analyzed with high performance liquid
499 chromatography and quantified with an Agilent 1260 Infinity refractive index detector
500 (Agilent Technologies, Inc. Palo Alto, CA) using a 300 x 7.8 mm Bio-Rad Aminex HPX-87H
501 column with Cation-H guard (BioRad, Inc., Hercules, CA). Acetamide, ethanol, n-propionic
502 acid, n-butyric acid, iso-butyric acid, n-pentanoic acid, iso-pentanoic acid, n-hexanoic acid,
503 iso-hexanoic acid, n-heptanoic acid, and n-octanoic acid were analyzed with tandem gas
504 chromatography-mass spectrometry. An Agilent 7890A GC system (Agilent Technologies,
505 Inc. Palo Alto, CA) with a 0.25 mm Restek Stabilwax DA 30 column (Restek 11008,
506 Belefonte, PA) was used. The GC-MS system was equipped with a Gerstel MPS2 (Gerstel,
507 Inc. Baltimore, MD) auto sampler and a solid-phase micro-extraction gray hub fiber
508 assembly (Supelco, Bellefonte, PA). The MS detector was a Pegasus 4D TOF-MS (Leco Corp.,
509 Saint Joseph, MI). Stable isotope labeled internal standards were used for each of the
510 analytes measured with GC-MS.
511 **DNA and RNA sequencing.** Biomass samples, consisting of centrifuged and decanted 2 mL
512 aliquots, were collected at Day 12, Day 48, Day 84, Day 96, and Day 120 of reactor

513 operations from initial start-up. Samples were also taken at 96 days and flash-frozen in
514 liquid nitrogen for RNA extraction. For DNA sequencing, cells were lysed by incubating in a
515 lysis solution (1.5M sodium chloride, 100mM trisaminomethane, 100mM ethylenediamine
516 (EDTA), 75mM sodium phosphate, 1% cetyltrimethylammonium bromide, and 2% sodium
517 dodecyl sulfate(SDS)), lysozyme (Thermo Fisher Scientific, MA, USA), and proteinase K
518 (New England Biolabs,MA, USA). For RNA sequencing, cells were lysed by incubating in a
519 lysis solution (20 mM sodium acetate, 1 mM EDTA, and 0.5% SDS prepared in
520 diethylpyrocarbonate-treated water (Invitrogen, CA, USA)) and TRIzol (Invitrogen, CA,
521 USA). The treated cells were subjected to bead beating with Lysing Matrix A (MP
522 Biomedicals, CA, USA). After this step, successive phase separations with
523 phenol:chloroform were used to separate nucleic acids from additional cell material.
524 Nucleic acids were precipitated with a sodium acetate solution and washed with ethanol.
525 RNA was further purified with an RNEasy Mini Kit (Qiagen, Hilden, Germany) and on-
526 column DNAse 1 (Qiagen, Hilden, Germany) treatment. After re-suspending the nucleic
527 acids, quantity, purity, and quality were assessed with a Qubit 4 Fluorometer (Thermo
528 Fisher Scientific, MA, USA), a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific,
529 MA, USA), and gel electrophoresis. RNA samples were submitted to the University of
530 Wisconsin Gene Expression Center for quality control with a Bioanalyzer (Agilent, CA,
531 USA), ribosomal RNA reduction with a RiboZero-Bacteria rRNA Removal Kit (Illumina, CA,
532 USA) with a 1μg RNA input. Strand-specific cDNA libraries were prepared with a TruSeq
533 RNA Library Prep Kit (Illumina, CA, USA). DNA and RNA were sequenced with the Illumina
534 HiSeq 2500 platform (Illumina, CA, USA). For DNA, an average insert size of 550 bp was
535 used and 2 x 250 bp reads were generated. For RNA, 1 x 100 bp reads were generated. Raw

536 DNA and cDNA reads can be found on the National Center for Biotechnology Information
537 (NCBI) website under BioProject PRJNA418244.
538 **Metagenomic assembly, binning, and quality control.** DNA sequencing reads were
539 filtered with Sickle using a minimum quality score of 20 and a minimum sequence length of
540 100 (35). Reads from all five samples were then co-assembled using metaspades and kmer
541 values of 21, 33, 55, 77, 99, and 127 (36). Binning of assembled contigs was performed with
542 MaxBin v2.2.1 (37). The quality, completeness, and contamination of each bin was analyzed
543 with CheckM v1.0.3 (13). Read mapping was performed with BBMAP v35.92
544 (https://sourceforge.net/projects/bbmap) to estimate the relative abundance of each bin.
545 Relative abundance was calculated by normalizing the number of mapped reads to the
546 genome size.
547 **Phylogenetic analysis.** Phylogeny of the draft genomes was assessed using 37 universal
548 single-copy marker genes with Phylosift v1.0.1 (38). In addition to the draft genomes, 62
549 publically-available genomes of related organisms were used to construct a phylogenetic
550 tree. Concatenated amino acid sequences of the marker genes were aligned with Phylosift,
551 and a maximum likelihood phylogenetic tree was constructed with RAxML v8.2.4 with the
552 PROTGAMMAAUTO model and 100 bootstraps (39). ANI calculations were performed
553 using JSpecies (40).
554 **Genome annotations.** Draft genomes were annotated with MetaPathways v2.5.(16) Open
555 reading frames (ORFs) were predicted using Prodigal v2.0 (41), and the ORFs were
556 annotated with the following databases: SEED (accessed March 2013), Clusters of
557 Orthologous Groups (COG, accessed December 2013), RefSeq (accessed January 2017),
558 Metacyc (accessed October 2011), and KEGG (accessed January 2017). The LAST algorithim

559 was used for assigning functional annotations (42). Functional annotations for each MAG
560 are provided in Supplement 4. Draft genomes were further annotated with the CAZY
561 database (17). CELLO was used to determine the subcellular location of the CAZYs (18).
562 Transporters were identified using the Transporter Classification Database.
563 **Transcript analysis.** Analysis of transcript data was performed as described in Lawson, et
564 al. (43). cDNA reads were quality filtered as described above for DNA. SortMeRNA was used
565 to remove rRNA sequences using multiple databases for RNA sequences (44). The
566 remaining non-rRNA sequences were then mapped back to the draft genomes using BBMap
567 v35.92 (https://sourceforge.net/projects/bbmap) with the minimum sequence identity set
568 to 0.95. Ambiguous reads with multiple top-hit mapping locations were assigned to a
569 random ORF. The number of RNA reads mapping to each ORF was calculated with htseq-
570 count v0.6.1 with the "intersection-strict" parameter (45). Relative gene expression
571 (RPKM) was calculated for each ORF by normalizing the number of mapped RNA reads for
572 each ORF to the ORF length and the total number of RNA reads mapping back to the
573 genome. The relative RPKM (relPKM) was then calculated as the ratio of the RPKM for the
574 ORF to the median RPKM across the draft genome. Finally, the $\log_2$(relRPKM) was
575 calculated to determine the log-fold difference. As such, a positive number corresponds to
576 greater than median expression levels and a negative number to expression below median
577 levels.

578 **Acknowledgements**

579 This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER
580 Office of Science DE-FC02-07ER64494 and DE-SC0018409). M.J.S. is supported by the
581 National Science Foundation Graduate Research Fellowship Program under grant No. DGE-

582 1256259. C.E.L is supported by a Postgraduate Scholarship–Doctoral (PGS-D) award from
583 the National Sciences and Engineering Research Council of Canada (NSERC) and a
584 Wisconsin Distinguished Graduate Fellowship. JJH was supported by the National Institute
585 of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-
586 67012-24709.

587 The authors thank Francisco Moya for assistance with RNA extraction and the
588 University of Wisconsin Biotechnology Center DNA Sequencing Facility and Gene
589 Expression Center for library preparation and Illumina sequencing services.

590 **References**

591

592 1. **Angenent LT, Richter H, Buckel W, Spirito CM, Steinbusch KJJ, Plugge CM, Strik**
593 **D, Grootscholten TIM, Buisman CJN, Hamelers HVM.** 2016. Chain Elongation with
594 Reactor Microbiomes: Open-Culture Biotechnology To Produce Biochemicals.
595 Environ Sci Technol **50:**2796-2810.
596 2. **Holtzapple MT, Granda CB.** 2009. Carboxylate platform: the MixAlco process part
597 1: comparison of three biomass conversion platforms. Appl Biochem Biotechnol
598 **156:**95-106.
599 3. **Sarria S, Kruyer NS, Peralta-Yahya P.** 2017. Microbial synthesis of medium-chain
600 chemicals from renewables. Nat Biotechnol **35:**1158-1166.
601 4. **Scarborough MJ, Lynch G, Dickson M, McGee M, Donohue TJ, Noguera DR.** 2018.
602 Increasing the economic value of lignocellulosic stillage through medium-chain fatty
603 acid production. Biotechnology for Biofuels **11:**200.
604 5. **Agler MT, Spirito CM, Usack JG, Werner JJ, Angenent LT.** 2012. Chain elongation
605 with reactor microbiomes: upgrading dilute ethanol to medium-chain carboxylates.
606 Energy Environ Sci **5:**8189.
607 6. **Ge S, Usack JG, Spirito CM, Angenent LT.** 2015. Long-Term n-Caproic Acid
608 Production from Yeast-Fermentation Beer in an Anaerobic Bioreactor with
609 Continuous Product Extraction. Environ Sci Technol **49:**8012-8021.
610 7. **Grootscholten TIM, Strik DPBTB, Steinbusch KJJ, Buisman CJN, Hamelers HVM.**
611 2014. Two-stage medium chain fatty acid (MCFA) production from municipal solid
612 waste and ethanol. Applied Energy **116:**223-229.
613 8. **Grootscholten TIM, dal Borgo FK, Hamelers HVM, Buisman CJN.** 2013.
614 Promoting chain elongation in mixed culture acidification reactors by addition of
615 ethanol. Biomass & Bioenergy **48:**10-16.

616 9. **Kucek LA, Xu JJ, Nguyen M, Angenent LT.** 2016. Waste Conversion into n-
617 Caprylate and n-Caproate: Resource Recovery from Wine Lees Using Anaerobic
618 Reactor Microbiomes and In-line Extraction. Front Microbiol **7**.
619 10. **Xiong BY, Richard TL, Kumar M.** 2015. Integrated acidogenic digestion and
620 carboxylic acid separation by nanofiltration membranes for the lignocellulosic
621 carboxylate platform. J Membr Sci **489:**275-283.
622 11. **Andersen SJ, De Groof V, Khor WC, Roume H, Props R, Coma M, Rabaey K.** 2017.
623 A Clostridium Group IV Species Dominates and Suppresses a Mixed Culture
624 Fermentation by Tolerance to Medium Chain Fatty Acids Products. Front Bioeng
625 Biotechnol **5:**8.
626 12. **Kenealy WR, Cao Y, Weimer PJ.** 1995. Production of caproic acid by cocultures of
627 ruminal cellulolytic bacteria and Clostridium kluyveri grown on cellulose and
628 ethanol. Appl Microbiol Biotechnol **44:**507-513.
629 13. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM:
630 assessing the quality of microbial genomes recovered from isolates, single cells, and
631 metagenomes. Genome Res **25:**1043-1055.
632 14. **Kim M, Oh HS, Park SC, Chun J.** 2014. Towards a taxonomic coherence between
633 average nucleotide identity and 16S rRNA gene sequence similarity for species
634 demarcation of prokaryotes (vol 64, pg 346, 2014). Int J Syst Evol Microbiol
635 **64:**1825-1825.
636 15. **Kraatz M, Wallace RJ, Svensson L.** 2011. Olsenella umbonata sp. nov., a
637 microaerotolerant anaerobic lactic acid bacterium from the sheep rumen and pig
638 jejunum, and emended descriptions of Olsenella, Olsenella uli and Olsenella profusa.
639 Int J Syst Evol Microbiol **61:**795-803.
640 16. **Konwar KM, Hanson NW, Bhatia MP, Kim D, Wu SJ, Hahn AS, Morgan-Lang C,**
641 **Cheung HK, Hallam SJ.** 2015. MetaPathways v2.5: quantitative functional,
642 taxonomic and usability improvements. Bioinformatics **31:**3345-3347.
643 17. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B.** 2014. The
644 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42:**D490-
645 495.
646 18. **Yu CS, Chen YC, Lu CH, Hwang JK.** 2006. Prediction of protein subcellular
647 localization. Proteins **64:**643-651.
648 19. **Zurawski JV, Khatibi PA, Akinosho HO, Straub CT, Compton SH, Conway JM, Lee**
649 **LL, Ragauskas AJ, Davison BH, Adams MWW, Kelly RM.** 2017. Bioavailability of
650 Carbohydrate Content in Natural and Transgenic Switchgrasses for the Extreme
651 Thermophile Caldicellulosiruptor bescii. Appl Environ Microbiol **83**.
652 20. **Sakamoto T, Thibault JF.** 2001. Exo-arabinanase of Penicillium chrysogenum able
653 to release arabinobiose from alpha-1,5-L-arabinan. Appl Environ Microbiol
654 **67:**3319-3321.
655 21. **Abdullah M, French D.** 1970. Substrate Specificity of Pullulanase. Arch Biochem
656 Biophys **137:**483-&.
657 22. **Ronimus RS, Morgan HW.** 2001. The biochemical properties and phylogenies of
658 phosphofructokinases from extremophiles. Extremophiles **5:**357-373.
659 23. **Lawlis VB, Dennis MS, Chen EY, Smith DH, Henner DJ.** 1984. Cloning and
660 sequencing of the xylose isomerase and xylulose kinase genes of Escherichia coli.
661 Appl Environ Microbiol **47:**15-21.

662 24. **Demmer JK, Pal Chowdhury N, Selmer T, Ermler U, Buckel W.** 2017. The
663 semiquinone swing in the bifurcating electron transferring flavoprotein/butyryl-
664 CoA dehydrogenase complex from Clostridium difficile. Nat Commun **8:**1577.
665 25. **Buckel W, Thauer RK.** 2013. Energy conservation via electron bifurcating
666 ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. Biochim
667 Biophys Acta **1827:**94-113.
668 26. **de Kok S, Meijer J, van Loosdrecht MC, Kleerebezem R.** 2013. Impact of dissolved
669 hydrogen partial pressure on mixed culture fermentations. Appl Microbiol
670 Biotechnol **97:**2617-2625.
671 27. **Buckel W, Thauer RK.** 2013. Energy conservation via electron bifurcating
672 ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation.
673 Biochimica Et Biophysica Acta-Bioenergetics **1827:**94-113.
674 28. **Stams AJ, Plugge CM.** 2009. Electron transfer in syntrophic communities of
675 anaerobic bacteria and archaea. Nat Rev Microbiol **7:**568-577.
676 29. **Zhang F, Zhang Y, Chen M, van Loosdrecht MCM, Zeng RJ.** 2013. A modified
677 metabolic model for mixed culture fermentation with energy conserving electron
678 bifurcation reaction and metabolite transport energy. Biotechnol Bioeng **110:**1884-
679 1894.
680 30. **Rodriguez J, Kleerebezem R, Lema JM, van Loosdrecht MCM.** 2006. Modeling
681 product formation in anaerobic mixed culture fermentations. Biotechnol Bioeng
682 **93:**592-606.
683 31. **Balan V, Bals B, Chundawat SPS, Marshall D, Dale BE.** 2009. Lignocellulosic
684 Biomass Pretreatment Using AFEX. Biofuels: Methods and Protocols **581:**61-77.
685 32. **Parreiras LS, Breuer RJ, Avanasi Narasimhan R, Higbee AJ, La Reau A,**
686 **Tremaine M, Qin L, Willis LB, Bice BD, Bonfert BL, Pinhancos RC, Balloon AJ,**
687 **Uppugundla N, Liu T, Li C, Tanjore D, Ong IM, Li H, Pohlmann EL, Serate J,**
688 **Withers ST, Simmons BA, Hodge DB, Westphall MS, Coon JJ, Dale BE, Balan V,**
689 **Keating DH, Zhang Y, Landick R, Gasch AP, Sato TK.** 2014. Engineering and two-
690 stage evolution of a lignocellulosic hydrolysate-tolerant Saccharomyces cerevisiae
691 strain for anaerobic fermentation of xylose from AFEX pretreated corn stover. PLoS
692 One **9:**e107499.
693 33. **American Public Health Association WEF, and American Water Works**
694 **Association** 2005. Standard Methods for the Examination of Water and
695 Wastewater: 21st Edition. American Public Health Association.
696 34. **Yemm EW, Willis AJ.** 1954. The estimation of carbohydrates in plant extracts by
697 anthrone. Biochem J **57:**508-514.
698 35. **Joshi NA, Fass JN.** 2011. Sickle: A sliding-window, adaptive, quality-based trimming
699 tool for FastQ files (Version 1.33)
700 36. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new
701 versatile metagenomic assembler. Genome Res **27:**824-834.
702 37. **Wu YW, Simmons BA, Singer SW.** 2016. MaxBin 2.0: an automated binning
703 algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics
704 **32:**605-607.
705 38. **Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM, Eisen JA.** 2014. PhyloSift:
706 phylogenetic analysis of genomes and metagenomes. PeerJ **2:**e243.

707 39. **Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-
708 analysis of large phylogenies. Bioinformatics **30:**1312-1313.
709 40. **Richter M, Rossello-Mora R, Oliver Glockner F, Peplies J.** 2016. JSpeciesWS: a
710 web server for prokaryotic species circumscription based on pairwise genome
711 comparison. Bioinformatics **32:**929-931.
712 41. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal:
713 prokaryotic gene recognition and translation initiation site identification. BMC
714 Bioinformatics **11:**119.
715 42. **Kielbasa SM, Wan R, Sato K, Horton P, Frith MC.** 2011. Adaptive seeds tame
716 genomic sequence comparison. Genome Res **21:**487-493.
717 43. **Lawson CE, Wu S, Bhattacharjee AS, Hamilton JJ, McMahon KD, Goel R, Noguera**
718 **DR.** 2017. Metabolic network analysis reveals microbial community interactions in
719 anammox granules. Nat Commun **8:**15416.
720 44. **Kopylova E, Noe L, Touzet H.** 2012. SortMeRNA: fast and accurate filtering of
721 ribosomal RNAs in metatranscriptomic data. Bioinformatics **28:**3211-3217.
722 45. **Anders S, Pyl PT, Huber W.** 2015. HTSeq--a Python framework to work with high-
723 throughput sequencing data. Bioinformatics **31:**166-169.
724

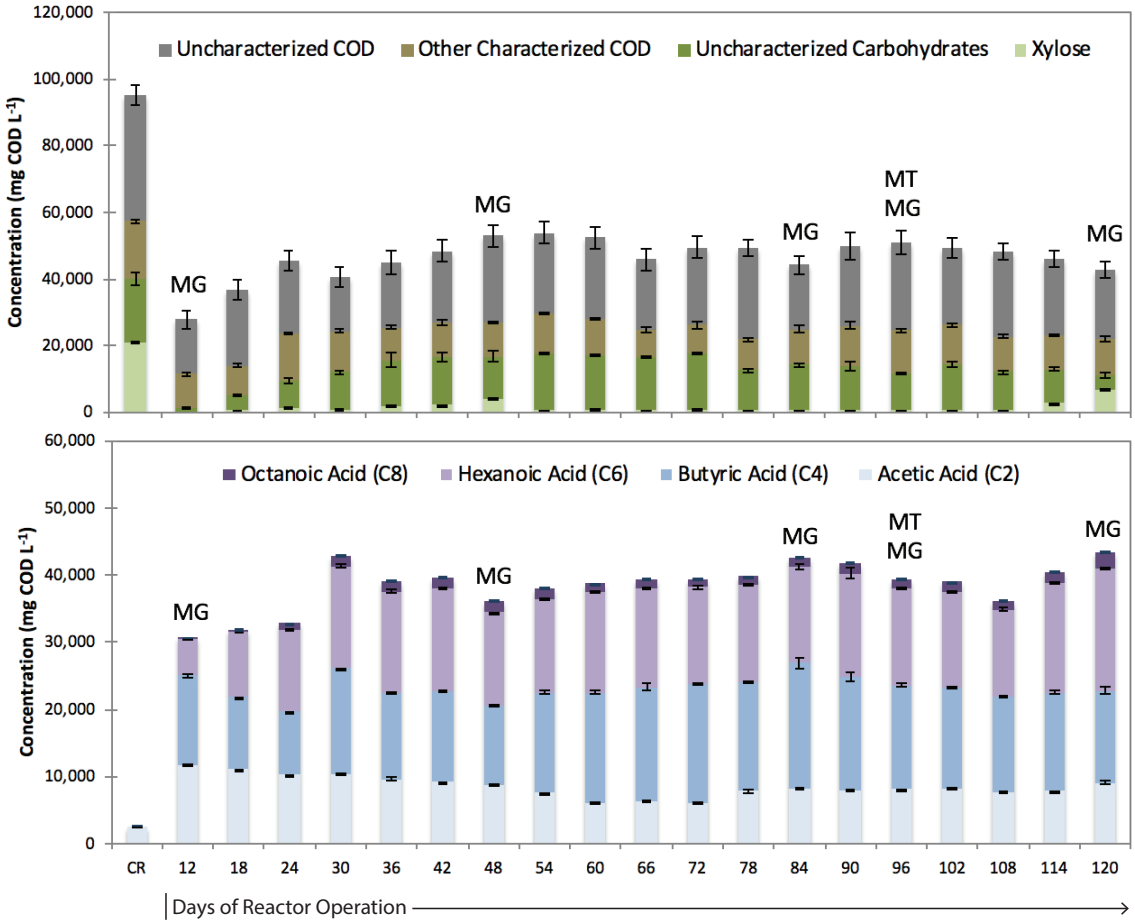


**Fig. 1. Transformation of materials in lignocellulosic ethanol conversion residue by an anaerobic microbiome.** During 120 days of reactor operation, compounds in conversion residue (CR) were converted to medium-chain fatty acids. Samples were taken for metagenomic (MG) analysis from five timepoints (Day 12, Day 48, Day 84, Day 96, and Day 120 and for metatranscriptomic analysis (MT) from one time point (Day 96). Overall, the bioreactor transformed xylose, uncharacterized carbohydrates and uncharacterized COD to acetic (C2), butyric (C4), hexanoic (C6) and octanoic (C8) acids. This figure is generated with data from Scarborough, et al. 2018 **(4).**

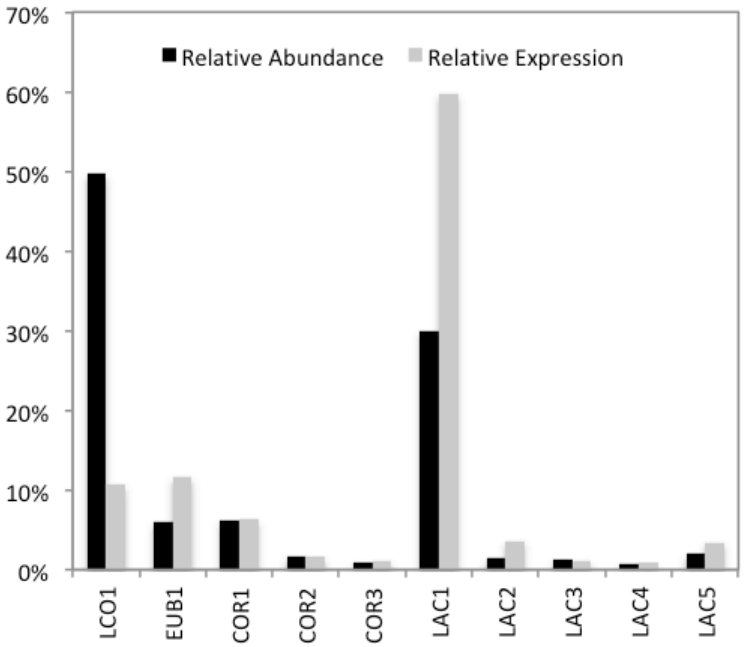


**Fig. 2. Relative abundance and expression of the 10 most abundant MAGs in the bioreactor at Day 96.** Relative abundance was determined by mapping DNA sequencing reads to the MAG and normalizing to the length of the MAG genome. Relative transcript abundance (expression) was determined by mapping c-DNA sequencing reads to the MAG and normalizing to the length of the MAG genome.

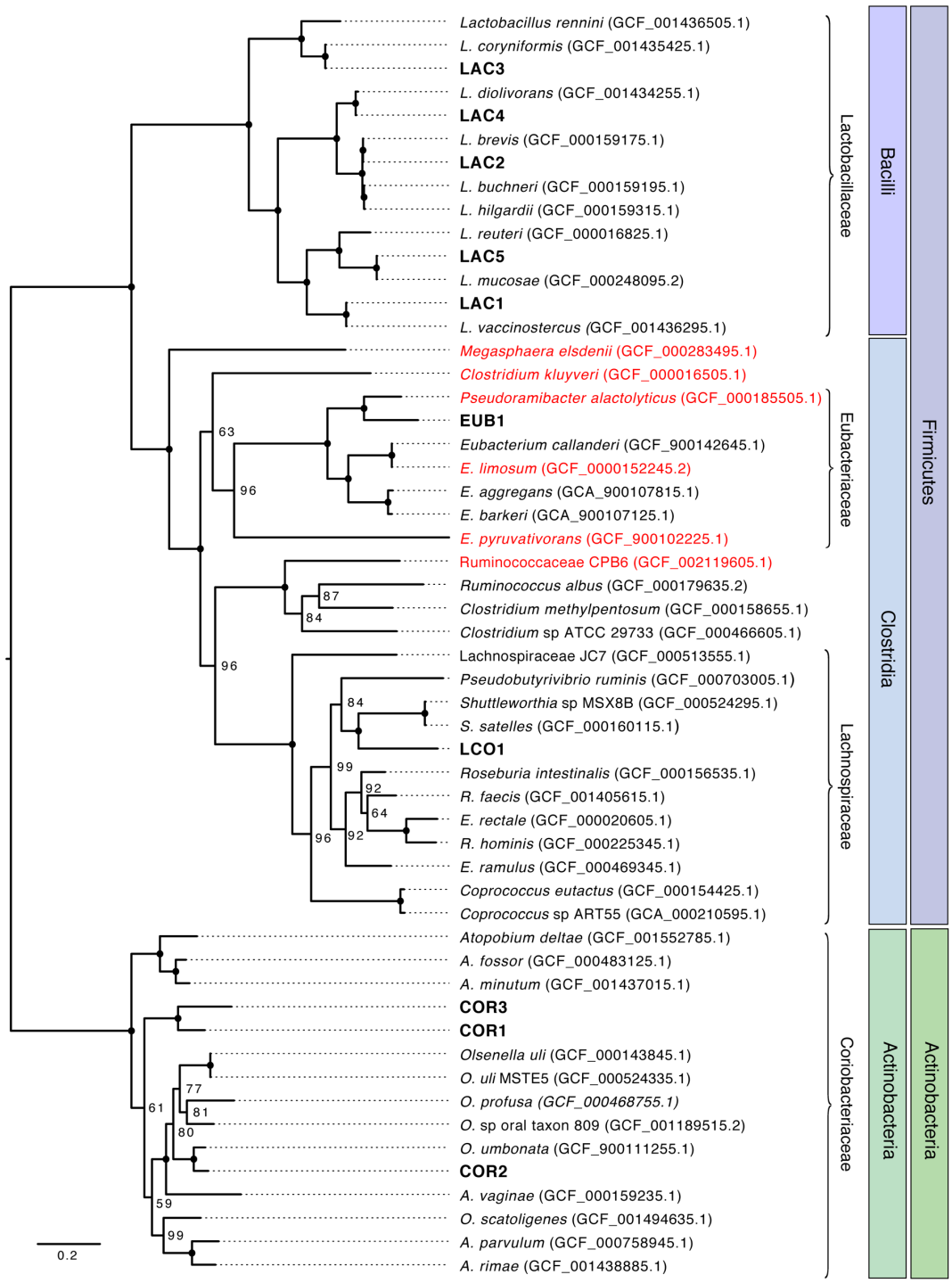


**Fig. 3. Phylogenetic analysis for ten MAGs obtained from reactor biomass.** Draft genomes from this study are shown in bold text. Red text indicates an organism that has been shown to produce MCFA. National Center for Biotechnology Information assembly accession numbers are shown in parentheses. Node labels represent bootstrap support values with solid circles representing a bootstrap support value of 100. The phyla and class of genomes are shown in shaded boxes and families are indicated by brackets For Actinobacteria genomes, Actinobacteria is both the phylum and class.

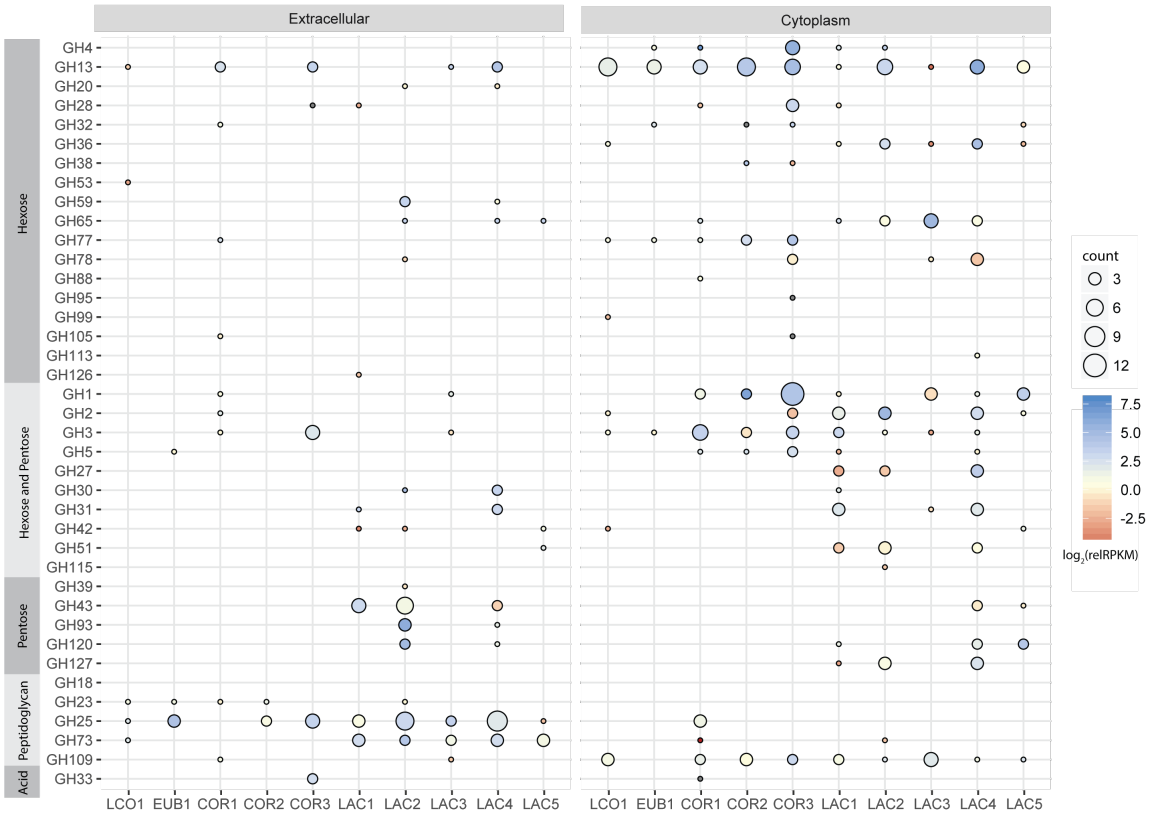


**Fig. 4. Relative expression of extracellular and cytoplasmic glycoside hydrolase enzymes.** Circle size indicates the number of ORFs annotated within each bin as the indicated CAZyme family. CAZyme families are grouped according to the target substrate moieties, being either bonds between hexoses only, bonds between both hexoses and pentoses, bonds between pentoses only, peptidoglycan-specific moieties, or acid moieties. The color intensity represents the relative gene expression, with red color intensity indicating expression below median levels and blue intensity indicating expression above median levels. The represented expression is the highest relative expression for an individual gene within the indicated family. Descriptions of each family are available through the CAZY database (http://www.cazy.org/Glycoside-Hydrolases.html).

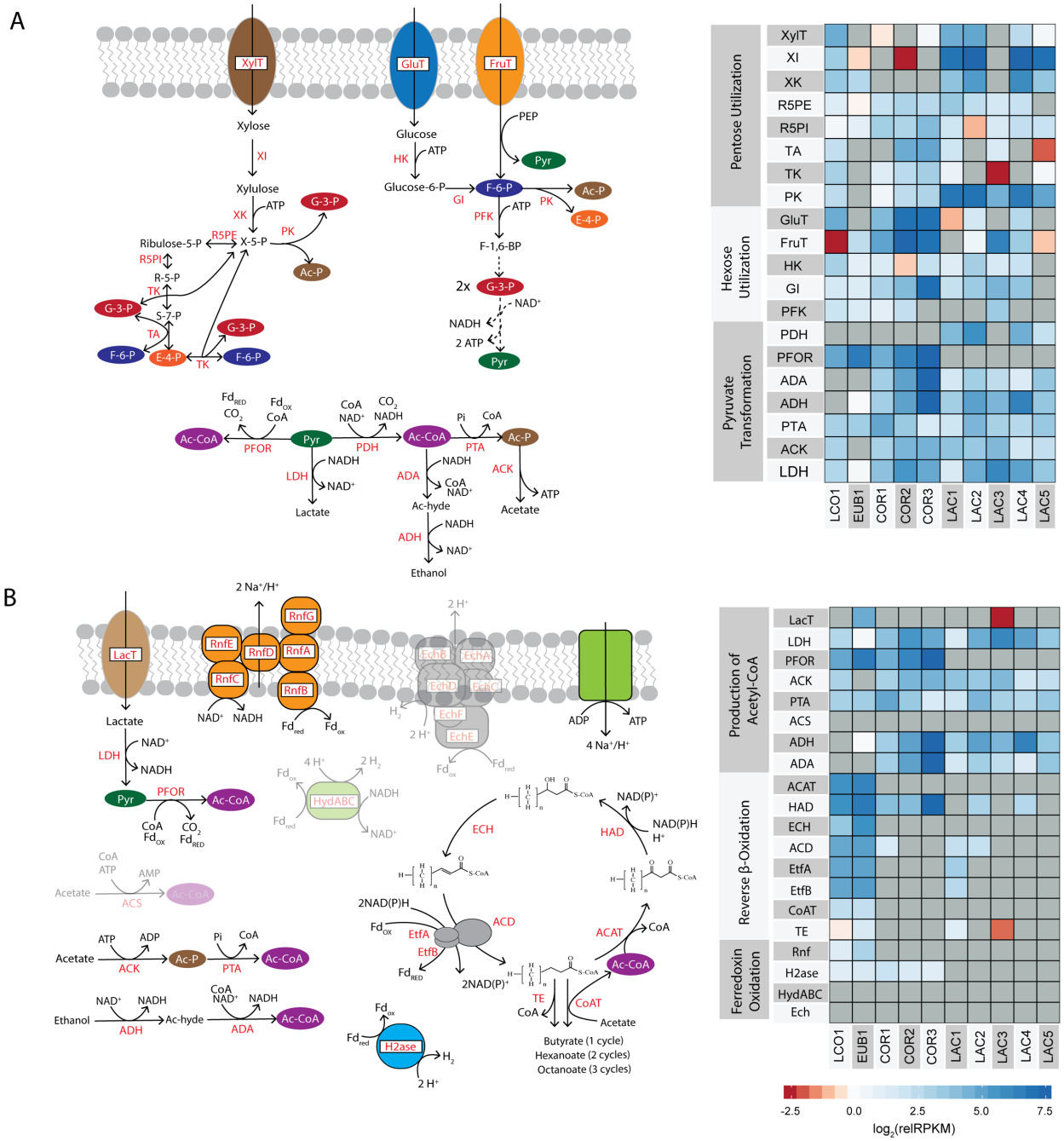


**Fig. 5. Relative expression of genes involved in the conversion of xylose and glycerol to MCFA**. Expression of key enzymes involved in (**A**) the utilization of xylose, and **(B)** acyl chain elongation. Dashed lines represent the existence of multiple enzyme reactions between the indicated molecules. The color intensities of the heatmaps represent the relative gene expression, with red color intensity indicating expression below median levels and blue intensity indicating expression above median levels. Grey indicates that the gene is absent from the genome. For genes that are not predicted to be expressed by any MAGs, the associated enzyme product is greyed out in the pathway figure. Key pathway intermediates include xylulose, xylulose-5-phosphate (X-5-P), ribose-5-phosphate (R-5-P), sedoheptulose-7-phosphate (S-7-P), erythyrose-4-phosphate (E-4-P), glyceraldehyde-3-phosphate (G-3-P), fructose-6-phosphate (F-6-P), acetate (Ac), acetyl-phosphate (Ac-P), acetyl-CoA (Ac-CoA), lactate and ethanol. Enzyme abbreviations are as follows: **(A)** XylT = xylose transporter, XI = xylose isomerase (EC 5.3.1.5); XK = xylulose kinase (EC 2.7.1.17), R5PE = ribulose-5-phosphate epimerase

(EC 5.1.3.1), R5PI = ribose-5-phosphate isomerase (EC 5.3.1.6), TA = transaldolase (EC 2.2.1.2), TK = transketolase (EC 2.2.1.1), PK = D-xylulose 5-phosphate/ D-fructose 6-phosphate phosphoketolase (EC 4.1.2.9), GluT = glucose transporter, FruT = fructose PTS transporter, HK = hexokinase (EC 2.7.1.1, EC 2.7.1.2), G6PI =glucose-6-phoshphate isomerase (EC 5.3.1.9), PFK = phosphofructokinase (EC 2.7.1.11), PDH = pyruvate dehydrogenase complex (EC 1.2.4.1, EC 2.3.1.12, EC 1.8.1.4), PFOR = pyruvate flavodoxin oxidoreductase (EC 1.2.7.-), ADA = acetaldehyde dehydrogenase (EC 1.2.1.10), AD = alcohol dehydrogenase (EC 1.1.1.1), PTA = phosphate acetyltransferase (EC 2.3.1.8), ACK = acetate kinase (EC 2.7.2.1), LDH = lactate dehydrogenase (EC 1.1.1.27), **(B)** LacT = lactate permease, ACS = acetyl-CoA synthetase (EC 6.2.1.1), ACAT = Acetyl-CoA C-acyltransferase (EC 2.3.1.16, EC 2.3.1.9), HAD = 3-Hydroxyacyl-CoA Dehydrogenase (EC 1.1.1.157, 1.1.1.35), ECH = Enoyl-CoA Hydratase (EC 4.2.1.55, EC 4.2.1.17), ACDH = Acyl-CoA Dehydrogenase (EC 1.3.99.2, EC 1.3.99.-), EtfA = electron transfer flavoprotein A, EtfB = electron transfer flavoprotein B, TE = thioesterase (EC 3.1.2.20), CoAT = 4-Hydroxybutyrate CoA Transferase (EC 2.8.3.-), Rnf = ferredoxin: $NAD^+$ oxidoreductase - Na+ translocating (EC 1.18.1.8), H2ase = ferredoxin hydrogenase (EC 1.12.7.2), HydABC = bifurcating [Fe-Fe] hydorgenase (EC 1.12.1.4), Ech = energy conserving hydrogenase (EchABCDEF)

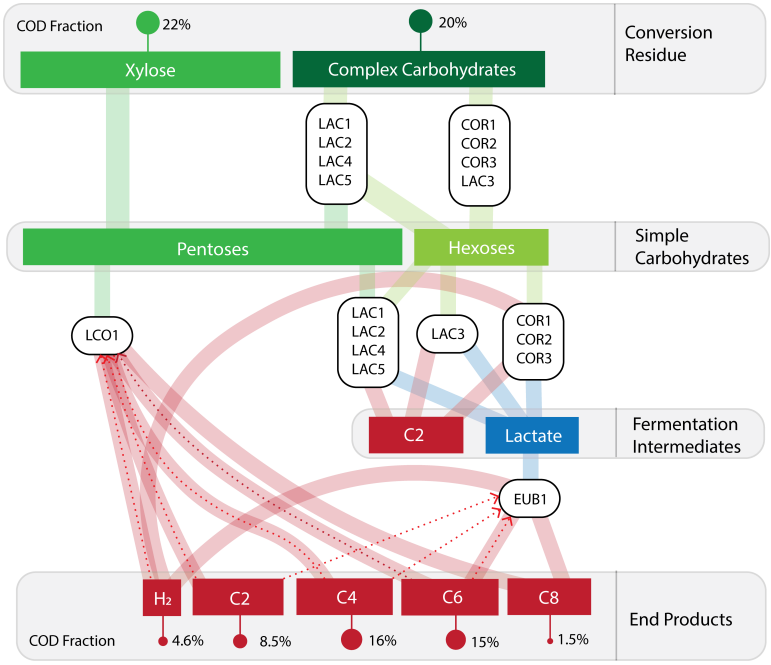


**Fig. 6. Predicted transformations of major substrates in conversion residue to MCFA by this anaerobic microbiome.** The microbes in the LAC and COR bins are predicted to produce sugars from complex carbohydrates. Simple carbohydrates, including xylose remaining in conversion residue, are converted to lactate and acetate by *Lactobacillus* (LAC) and Coriobacteriaceae (COR) MAGs. The Lachnospiraceae (LCO1) MAG converts pentoses directly to butyric acid (C4). The Eubacteriaceae (EUB1) produces hexnaoic acid (C6) and octanoic acid (C8) from lactate. Further, LCO1 may utilize hydrogen to elongate C2 and C4 to MCFA, as represented by dashed lines. Additionally, EUB1 may elongate C2 and C4 and C8.

795 **Table S1. Summary of MAGs obtained from DNA sequence analysis of the reactor microbiome.**
796 Taxonomy, completeness, and contamination were estimated with CheckM. Draft genomes were assembled
797 from five independent reactor samples. These MAGs represent the ten most abundant MAGs at Day 96 (Fig. 1).
798

| Bin ID | Taxonomy | Completeness (%) | Contamination (%) | Genome size (bp) | # Scaffolds | N50 | GC (%) | Predicted Genes |
|---|---|---|---|---|---|---|---|---|
| LCO1 | Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Shuttleworthia | 95.4 | 0.0 | 2,106,912 | 44 | 103,964 | 45.7 | 1,900 |
| EUB1 | Firmicutes; Clostridia; Clostridiales; Eubacteriaceae; Pseudoramibacter | 97.8 | 0.2 | 2,002,609 | 35 | 142,846 | 51.2 | 1,857 |
| COR1 | Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Olsenella | 99.2 | 0.8 | 2,512,349 | 225 | 22,880 | 59.0 | 2,358 |
| COR2 | Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Olsenella | 100 | 1.6 | 2,422,853 | 155 | 34,678 | 64.8 | 2,185 |
| COR3 | Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Olsenella | 98.4 | 7.4 | 3,647,413 | 533 | 13,368 | 55.5 | 4,068 |
| LAC1 | Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus | 99.5 | 1.1 | 2,633,889 | 18 | 640,122 | 43.6 | 2,567 |
| LAC2 | Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus | 99.4 | 1.6 | 3,179,174 | 79 | 122,889 | 40.5 | 2,989 |
| LAC3 | Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus | 99.2 | 1.4 | 2,704,063 | 174 | 29,509 | 43.0 | 2,731 |
| LAC4 | Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus | 98.9 | 1.3 | 3,335,227 | 95 | 86,779 | 40.2 | 3,150 |
| LAC5 | Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus | 80.1 | 0.8 | 1,487,044 | 181 | 12,363 | 46.1 | 1,524 |

799
800

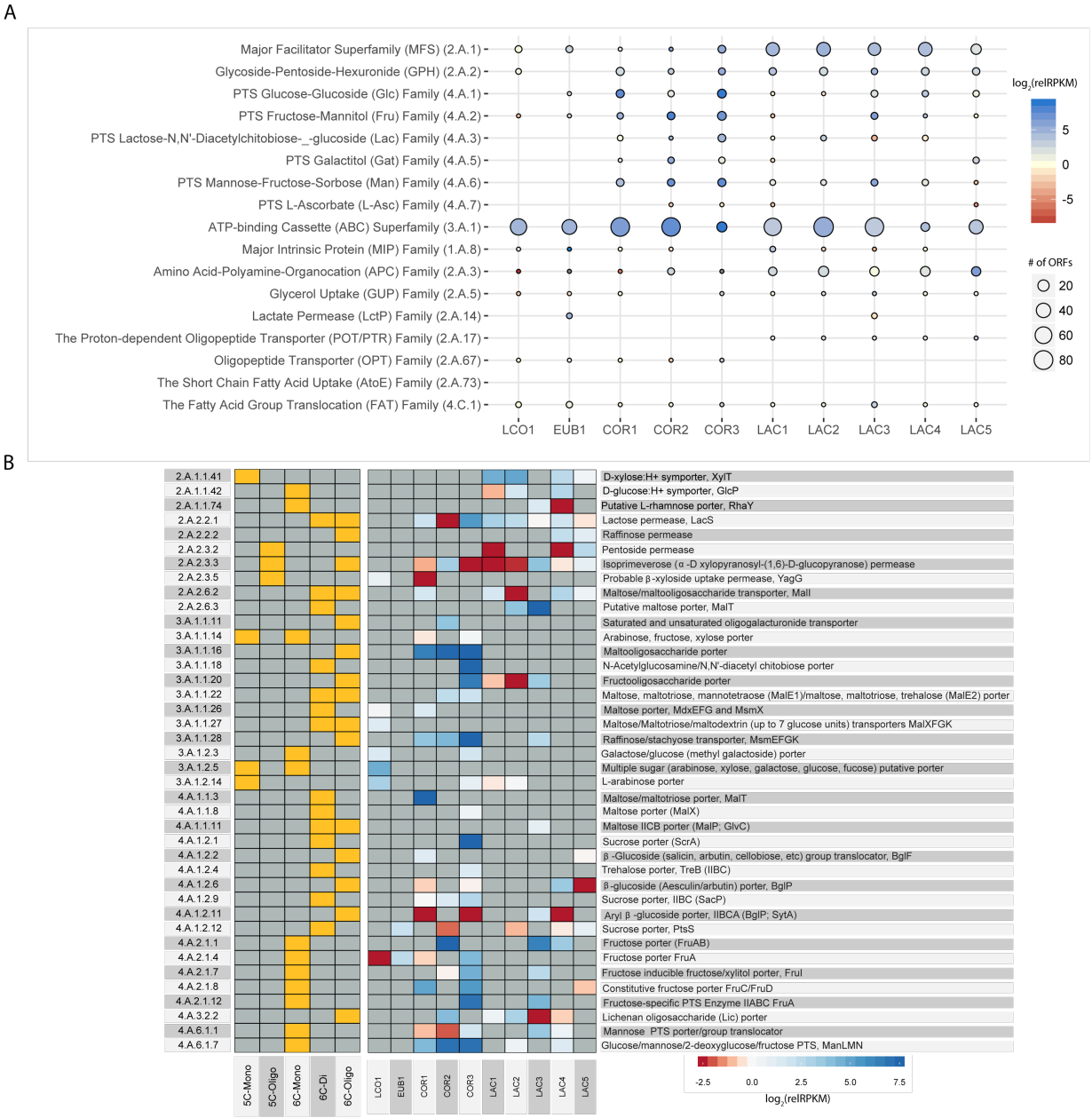


**Fig. S1. Expression of known or predicted transporters by microbiome members.** Transporters were annotated with the Transporter Classification Database. **(A)** The number of genes and relative expression for each analyzed MAG are shown for Transporter Database families. Circle size indicates the number of ORFs annotated within each bin as the indicated CAZY class or family. The color intensity represents the relative gene expression, with brown color intensity indicating expression below median levels and purple intensity indicating expression above median levels. The represented expression is the highest relative expression for an individual gene within the indicated transporter family. (B) The expected substrates, either a pentose (5C-Mono), pentose-containing oligosaccharide (5C-Oligo), hexose (6C-mono), hexose disaccharide (6C-Di), or hexose-containing oligosaccharide (6C-Oligo) are indicated by orange squares. The color intensity represents the relative gene expression, with red color intensity indicating expression below median levels and blue intensity indicating expression above median levels. The transporter family ID and descriptions are shown. More detailed descriptions of transporters are available from the Transporter Database (http://www.tcdb.org).

**Table S2. Thermodynamics of biochemical reactions involved in conversion of xylose to butyrate, hexanoate and octanoate.** Free energies of formation for all chemical compounds were obtained from Kbase (www.kbase.us). The ATP yield was determined based on biochemical models presented in Supplement 7 and is indicated as mol ATP produced per mol xylose or lactate consumed. The terminal enzyme of reverse β-oxidation, either a CoA transferase (CoAT) or thioesterase (TE), is also indicated.

| Eq No. | Equation | Associated MAG(s) | ΔG per mol substrate[1] (kJ mol$^{-1}$) | | | | $\Delta G^{0'}$available per ATP produced[2,3] (kJ mol$^{-1}$ ATP) | Terminal Enzyme |
|---|---|---|---|---|---|---|---|---|
| | | | T = 35 °C $P_{H2}$ =10-6 atm | T = 35 °C $P_{H2}$ = 1 atm | T = 35 °C $P_{H2}$ = 6.8 atm | $Y_{ATP}$ (mol ATP mol$^{-1}$ Substrate)[1,2] | | |
| **Xylose Simple Fermentation** | | | | | | | | |
| Eq. 1. | $3\ C_5H_{10}O_5 \rightarrow 5\ C_3H_5O_3^- + 5\ H^+$ | LAC1, LAC2, LAC4, LAC5 | -174 | -174 | -174 | 1.67 | -104 to -104 | - |
| Eq. 2. | $3\ C_5H_{10}O_5 \rightarrow 3\ C_3H_5O_3^- + 3\ C_2H_3O_2^- + 6\ H^+$ | LAC1, LAC2, LAC4, LAC5 | -214 | -214 | -214 | 2.00 | -107 to -107 | - |
| **Xylose Elongation** | | | | | | | | |
| Eq. 3. | $3\ C_5H_{10}O_5 \rightarrow 3\ C_4H_7O_2^- + 3\ CO_2 + 3\ H_2O + 3\ H^+$ | LCO1 | -264 | -264 | -264 | 3.00 | -88 to -88 | CoAT |
| Eq. 4. | $3\ C_5H_{10}O_5 \rightarrow 1\ C_6H_{11}O_2^- + 3\ C_2H_3O_2^- + 3\ CO_2 + 4\ H^+ + 2\ H_2$ | LCO1 | -272 | -248 | -245 | 2.83 | -87 to -96 | CoAT |
| Eq. 5. | $3\ C_5H_{10}O_5 \rightarrow 1\ C_8H_{15}O_2^- + 2\ C_2H_3O_2^- + 3\ CO_2 + 3\ H_2O + 3\ H^+$ | LCO1 | -265 | -265 | -265 | 3.00 | -88 to -88 | CoAT |
| Eq. 6. | $2\ C_5H_{10}O_5 \rightarrow 1\ C_4H_7O_2^- + 2\ C_2H_3O_2^- + 2\ CO_2 + 3\ H^+ + 2\ H_2$ | LCO1 | -276 | -240 | -235 | 2.25 | -105 to -123 | TE |
| Eq. 7. | $3\ C_5H_{10}O_5 \rightarrow 1\ C_6H_{11}O_2^- + 3\ C_2H_3O_2^- + 3\ CO_2 + 1\ H_2O + 4\ H^+ + 2\ H_2$ | LCO1 | -272 | -248 | -245 | 2.50 | -98 to -109 | TE |
| Eq. 8 | $4\ C_5H_{10}O_5 \rightarrow 1\ C_8H_{15}O_2^- + 4\ C_2H_3O_2^- + 4\ CO_2 + 2\ H_2O + 5\ H^+ + 2\ H_2$ | LCO1 | -270 | -253 | -250 | 2.63 | -95 to -103 | TE |
| **Xylose and C2/ C4/ C6 Elongation** | | | | | | | | |
| Eq. 9 | $1\ C_5H_{10}O_5 + 2\ C_2H_3O_2^- + 2\ H_2 \rightarrow 2\ C_4H_5O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -240 | -311 | -320 | 3.50 | -69 to -92 | CoAT |
| Eq. 10 | $1\ C_5H_{10}O_5 + 1\ C_2H_3O_2^- + 2\ H_2 \rightarrow 1\ C_6H_{11}O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -240 | -311 | -320 | 3.50 | -69 to -92 | CoAT |
| Eq. 11 | $1\ C_5H_{10}O_5 + 1\ C_4H_7O_2^- + 2\ H_2 \rightarrow 1\ C_8H_{15}O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -264 | -264 | -264 | 3.50 | -75 to -75 | CoAT |
| Eq. 12 | $1\ C_5H_{10}O_5 + 1\ C_4H_7O_2^- \rightarrow 1\ C_6H_{11}O_2^- + 1\ C_2H_3O_2^- + 1\ CO_2 + 1\ H_2O + 1\ H^+$ | LCO1 | -243 | -314 | -324 | 3.00 | -81 to -108 | CoAT |
| Eq. 13 | $1\ C_5H_{10}O_5 + 1\ C_6H_{11}O_2^- \rightarrow 1\ C_8H_{15}O_2^- + 1\ C_2H_3O_2^- + 1\ CO_2 + 1\ H_2O + 1\ H^+$ | LCO1 | -267 | -267 | -267 | 3.00 | -89 to -89 | CoAT |
| Eq. 14 | $1\ C_5H_{10}O_5 + 2\ C_2H_3O_2^- + 2\ H_2 \rightarrow 2\ C_4H_5O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -240 | -311 | -320 | 1.50 | -160 to -214 | TE |
| Eq. 15 | $1\ C_5H_{10}O_5 + 1\ C_2H_3O_2^- + 2\ H_2 \rightarrow 1\ C_6H_{11}O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -240 | -311 | -320 | 2.50 | -96 to -128 | TE |
| Eq. 16 | $1\ C_5H_{10}O_5 + 1\ C_4H_7O_2^- + 2\ H_2 \rightarrow 1\ C_8H_{15}O_2^- + 1\ CO_2 + 3\ H_2O$ | LCO1 | -264 | -264 | -264 | 2.50 | -105 to -105 | TE |
| Eq. 17 | $1\ C_5H_{10}O_5 + 1\ C_4H_7O_2^- \rightarrow 1\ C_6H_{11}O_2^- + 1\ C_2H_3O_2^- + 1\ CO_2 + 1\ H_2O + 1\ H^+$ | LCO1 | -243 | -314 | -324 | 2.00 | -122 to -162 | TE |
| Eq. 18 | $1\ C_5H_{10}O_5 + 1\ C_6H_{11}O_2^- \rightarrow 1\ C_8H_{15}O_2^- + 1\ C_2H_3O_2^- + 1\ CO_2 + 1\ H_2O + 1\ H^+$ | LCO1 | -267 | -267 | -267 | 2.00 | -133 to -133 | TE |

1. ΔG values and expected ATP yields are normalized to mol of xylose, mol of lactate, or mol of glycerol
2. Pathway reconstructions shown in Supplement 7 were used to determine expected ATP yields.
3. The minimum expected DG0' released per mol ATP produced is 60 kJ. Values less than this are indicated by red text and indicate the predicted ATP yield exceeds what is phyiologically feasible

**Table S3. Thermodynamics of biochemical reactions involved in conversion of lactate to butyrate, hexanoate and octanoate.**

| Eq No. | Equation | Associated MAG(s) | ΔG per mol substrate[1] (kJ mol$^{-1}$) | | | | $\Delta G^{0'}$ available per ATP produced[2,3] (kJ mol$^{-1}$ ATP) | Terminal Enzyme |
|---|---|---|---|---|---|---|---|---|
| | | | T = 35 °C $P_{H2}$ =10-6 atm | T = 35 °C $P_{H2}$ = 1 atm | T = 35 °C $P_{H2}$ = 6.8 atm | $Y_{ATP}$ (mol ATP mol$^{-1}$ Substrate)[1,2] | | |
| **Lactate Elongation** | | | | | | | | |
| Eq. 19 | $2\ C_3H_5O_3^- + 1\ H^+ \rightarrow 1\ C_4H_7O_2^- + 2\ CO_2 + 2\ H_2$ | EUB1 | -62 | -26 | -21 | 0.75 | -29 to -82 | CoAT |
| Eq. 20 | $3\ C_3H_5O_3^- + 2\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 3\ CO_2 + 2\ H_2 + 1\ H_2O$ | EUB1 | -58 | -34 | -31 | 0.83 | -37 to -70 | CoAT |
| Eq. 21 | $4\ C_3H_5O_3^- + 3\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 4\ CO_2 + 2\ H_2 + 2\ H_2O$ | EUB1 | -57 | -39 | -36 | 0.88 | -41 to -64 | CoAT |
| Eq. 22 | $2\ C_3H_5O_3^- + 1\ H^+ \rightarrow 1\ C_4H_7O_2^- + 2\ CO_2 + 2\ H_2$ | EUB1 | -62 | -26 | -21 | 0.25 | -86 to -247 | TE |
| Eq. 23 | $3\ C_3H_5O_3^- + 2\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 3\ CO_2 + 2\ H_2 + 1\ H_2O$ | EUB1 | -58 | -34 | -31 | 0.50 | -62 to -116 | TE |
| Eq. 24 | $4\ C_3H_5O_3^- + 3\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 4\ CO_2 + 2\ H_2 + 2\ H_2O$ | EUB1 | -57 | -39 | -36 | 0.63 | -58 to -90 | TE |
| **Lactate and C2/ C4/ C6 Elongation** | | | | | | | | |
| Eq. 25 | $1\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 1\ H^+ \rightarrow 1\ C_4H_7O_2^- + 1\ CO_2 + 1\ H_2O$ | EUB1 | -50 | -50 | -50 | 1.00 | -50 to -50 | CoAT |
| Eq. 26 | $2\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 2\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 2\ CO_2 + 2\ H_2O$ | EUB1 | -50 | -50 | -50 | 1.00 | -50 to -50 | CoAT |
| Eq. 27 | $3\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 3\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 3\ CO_2 + 3\ H_2O$ | EUB1 | -51 | -51 | -51 | 1.00 | -51 to -51 | CoAT |
| Eq. 28 | $1\ C_3H_5O_3^- + 1\ C_4H_7O_2^- + 1\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 1\ CO_2 + 1\ H_2O$ | EUB1 | -50 | -50 | -50 | 1.00 | -50 to -50 | CoAT |
| Eq. 29 | $2\ C_3H_5O_3^- + 1\ C_4H_7O_2^- + 2\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 2\ CO_2 + 2\ H_2O$ | EUB1 | -51 | -51 | -51 | 1.00 | -51 to -51 | CoAT |
| Eq. 30 | $1\ C_3H_5O_3^- + 1\ C_6H_{11}O_2^- + 1\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 1\ CO_2 + 1\ H_2O$ | EUB1 | -53 | -53 | -53 | 1.00 | -53 to -53 | CoAT |
| Eq. 31 | $1\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 1\ H^+ \rightarrow 1\ C_4H_7O_2^- + 1\ CO_2 + 1\ H_2O$ | None[4] | -50 | -50 | -50 | 0.00 | NA[5] | TE |
| Eq. 32 | $2\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 2\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 2\ CO_2 + 2\ H_2O$ | None[4] | -50 | -50 | -50 | 0.50 | -100 to -100 | TE |
| Eq. 33 | $3\ C_3H_5O_3^- + 1\ C_2H_3O_2^- + 3\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 3\ CO_2 + 3\ H_2O$ | None[4] | -51 | -51 | -51 | 0.67 | -76 to -76 | TE |
| Eq. 34 | $1\ C_3H_5O_3^- + 1\ C_4H_7O_2^- + 1\ H^+ \rightarrow 1\ C_6H_{11}O_2^- + 1\ CO_2 + 1\ H_2O$ | None[4] | -50 | -50 | -50 | 0.00 | NA[5] | TE |
| Eq. 35 | $2\ C_3H_5O_3^- + 1\ C_4H_7O_2^- + 2\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 2\ CO_2 + 2\ H_2O$ | None[4] | -51 | -51 | -51 | 0.50 | -103 to -103 | TE |
| Eq. 36 | $1\ C_3H_5O_3^- + 1\ C_6H_{11}O_2^- + 1\ H^+ \rightarrow 1\ C_8H_{15}O_2^- + 1\ CO_2 + 1\ H_2O$ | None[4] | -53 | -53 | -53 | 0.00 | NA[5] | TE |

1. ΔG values and expected ATP yields are normalized to mol of xylose, mol of lactate, or mol of glycerol
2. Pathway reconstructions shown in Supplement 7 were used to determine expected ATP yields.
3. The minimum expected ΔG0' released per mol ATP produced is 60 kJ. Values less than this are indicated by red text and indicate the predicted ATP yield exceeds what is phyiologically feasible
4.The proposed model requires ACK for either incorporation of a carboxylate or for ATP generation from acetyl-CoA. EUB1 is not predicted to produce this enzyme.
5. No net ATP production is predicted for this model.