Supplementary Information for:

Chain elongation with reactor microbiomes: upgrading dilute ethanol to medium-chain carboxylates

Matthew T. Agler[1], Catherine M. Spirito[1], Joseph G. Usack[1], Jeffrey J. Werner[1,2], and Largus T. Angenent[1]*

[1] Department of Biological and Environmental Engineering, 214 Riley-Robb Hall, Cornell University, Ithaca NY 14853, USA.

[2] Chemistry Department, SUNY Cortland, Bowers Hall, PO Box 2000, Cortland, NY 13045, USA.

*Correspondence to: la249@cornell.edu

Supplementary Information:

Results and Discussion: ***Linking the 16S rRNA gene and shotgun metagenomic sequencing surveys*****.** Our analysis of the 16S rRNA gene sequencing effort found five abundant OTUs that were significantly correlated ($r > 0.8$ and $p < 0.05$) with increasing *n*-caproic acid production rates (Figure 2A and Table S1). The OTUs were of various taxonomies, indicating that a range of bacteria played important roles in conversion of yeast-fermentation beer to *n*-caproic acid. To more specifically determine the roles of the OTUs, and to provide more certainty as to which bacteria were responsible for chain elongation, we performed a shotgun metagenomic sequencing analysis of bioreactor samples. Taxonomic analysis of the seven genes most significantly correlated with production rates of *n*-caproic acid (Figure S4) suggested the importance of some of the same taxonomic groups as the 16S rRNA gene sequencing analysis (genus *Clostridium* and family Ruminococcaceae). Seven genera made up most of the reads assigned to the seven genes, and four of those (*Ethanoligenens* [family Ruminococcaceae],

Desulfitobacterium, *Clostridium*, and *Propionibacterium*) increased as *n*-caproic acid production rates increased. To determine the probable roles of the 7 important genera, we looked at the taxonomic breakdown of genes catalyzing what are likely to be the most important carbon metabolism pathways in the bioreactor (Figure S3). Compared to the taxonomic breakdown of all reads, *Ethanoligenens* and *Bifidobacterium* were relatively abundant in starch hydrolysis and xylan metabolism, respectively, indicating that they may have been important in retrieving carbon from complex substrate molecules. *Clostridium* spp. strongly dominated the chain-elongation gene pool, reflecting their important role in the terminal process for *n*-caproic acid formation. Even though a chain-elongating bacterium, such as *C. kluyveri*, can oxidize ethanol on its own, *Desulfitobacterium* was apparently a catalyst for this important step. We have not clarified if they provided: i. reducing equivalents from ethanol to some chain-elongating bacterium in a mutualistic relationship; ii. hydrogen to hydrogenotrophic methanogens (although their large abundance and the relatively small flux of carbon to methane would indicate that it should have had some other role as well); or iii. a drain of reducing equivalents from our system to another electron acceptor, such as organic material or sulfate (we did not find H_2S in biogas analysis, though). The first option seems the most likely participation and more research is necessary to substantiate.

Materials and Methods: *Bioreactor Setup and Operation*. We operated a 5-L glass bioreactor for over a year to convert yeast fermentation beer to *n*-caproic acid (Figure 1 and Figure S1). The yeast fermentation beer was received in one shipment from Western New York Energy in Medina, NY. The raw, undiluted beer had a total and volatile solids content of 125.73 ± 0.14 g L^{-1} (n = 6) and 117.19 ± 0.15 g L^{-1} (n = 6), respectively, a chemical oxygen demand (COD) of 450.50 ± 51.12 g L^{-1} (n = 6), and the ethanol content was 152.7 ± 3.4 g L^{-1} (~15%) (n = 4). We diluted the beer 6.6 times before feeding to the bioreactor; in a real industrial setup, recycled liquor would have been used instead of make-up water. The bioreactor included a heated water jacket connected to a water heater to maintain a temperature of 30^{o}C, an hourly automatic mixing system that worked by recirculating biogas with a peristaltic pump, and an automated pH control system, which pumped 5M NaOH or HCl during mixing to maintain a pH of 5.5 (range 5.4-5.6). To prevent under pressure in the bioreactor during effluent withdrawal and feeding, the headspace was connected to a device that equalized pressure with the room while preventing air

intrusion. We fed the bioreactor semi-continuously on a 48-hour schedule: substrate was first fed into the bioreactor (time 0), followed by a react period with pH control, hourly mixing, and continuous *n*-caproic acid extraction (hours 0-47), followed by a 1-h biomass settling period (hour 47), and rapid effluent removal of a volume equal to substrate volume with a peristaltic pump (time 0). We maintained a hydraulic retention time (HRT) of 15 days (666 mL of substrate fed per cycle to the 5-L bioreactor) and a substrate organic loading rate of 4.5 g COD L^{-1} d^{-1} (ethanol loading rate: 66.6 mmol C L^{-1} d^{-1}) through day 120. By day 120, chain-elongation reactions consumed nearly all available ethanol in the substrate (Figure S5), and therefore we decreased the HRT to 12 days (833 mL of substrate per cycle for the 5-L bioreactor), resulting in a higher substrate loading rate of 5.7 g COD L^{-1} d^{-1} (ethanol loading rate: 83.3 mmol C L^{-1} d^{-1}) for the remainder of the operating period. The bioreactor was inoculated from previously operating bioreactors optimized for *n*-butyric acid production from dilute-acid pretreated corn fiber. Originally the *n*-butyric acid-producing bioreactor was started with a natural microbiome from sheep rumen and a thermophilic anaerobic digester from the city of Duluth, Minnesota (Western Lake Superior Sanitary District, Duluth, MN). For the first 30 days of the current study, the bioreactor was fed dilute-acid pretreated corn fiber, which was supplemented with ethanol at an HRT of 15 days and a loading rate of 1.7 g COD L^{-1} d^{-1} (ethanol loading rate: 32.6 mmol C L^{-1} d^{-1}).

In-line Liquid/Liquid* n-*Caproic Acid Extraction. To prevent product inhibition and to recover the product, we continuously extracted *n*-caproic acid using a membrane-based liquid/liquid extraction system (Figure 1). The extraction system consisted of hollow-fiber hydrophobic membrane contactors that allowed a high surface area for contact between the aqueous and solvent phases. We used eight commercially available hydrophobic membranes with a contact area of 2.32 m^2 for both the bioreactor/solvent and solvent/stripping interfaces (four on each side) (1.5x5.5 MiniModule X50, Liqui-Cel, Membrana, Wuppertal, Germany). On day 300, we increased the membrane contact area for both the bioreactor/solvent and solvent/stripping interfaces to 8.1 m^2 to avoid limitations due to rate of product extraction (4x13 316L SS X50, Liqui-Cel, Membrana). Indeed, until the end of the study the substrate-feeding rate was limiting the production rate and not the extraction rate. The driving force for the extraction was two-fold: 1. We used a reactive solvent (3% tri-*n*-octylphosphineoxide in mineral oil, Sigma-Aldrich, St.

Louis, MO), which is more selective for hydrophobic molecules, such as *n*-caproic acid, compared to shorter-chain molecules, such as acetic acid [1]; and 2. We maintained a pH gradient to take advantage of the dissociation constant for *n*-carboxylic acids (pK_a = 4.8-4.9) to selectively extract undissociated acids from the bioreactor at pH 5.5 and recover them in the dissociated form in a pH 9 aqueous solution. First, the bioreactor supernatant was pumped into the lumen side (inside the fibers) of the hollow-fiber membrane units at 10 mL/min after being filtered to remove remaining large particles (as much of the particles as possible were returned to the bioreactor on a weekly basis). The solvent, which wet the hydrophobic membranes, was pumped at the same rate on the shell side (outside the fibers) of the membrane, counter-flow to the bioreactor liquid. The solvent was constantly recirculated between contact with the bioreactor liquid and the shell side of a second membrane unit, where it contacted an aqueous phase buffered with a 0.5 M borate solution at pH 9. The pH 9 solution was continuously recirculated from a 5-L reservoir where a pH controller maintained the pH by automatic addition of 5M NaOH.

Methanogenic Activity Test. We operated 35-mL batch fermentation vessels in 93-h fermentations to test whether bioreactor microbiomes produced methane from acetic acid or only from carbon dioxide with hydrogen or ethanol as the source of reducing equivalents (electrons). All batch reactions were carried out in triplicate. Four triplicate sets of batch bottles were prepared (no source of electrons for methanogens). In short, in an anaerobic hood, we added 0.75 mL basal medium (described in Agler et al. [2]), ~8 mmol g^{-1} VS acetic acid, ~4 mmol g^{-1} VS *n*-butyric acid, and 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer to a total volume of 7.25 mL, added 0.25-mL inoculum and 0.1-mL sodium sulfide, corrected to pH 5.5 at 30^{o}C with NaOH (total volume ~10 mL), capped the bottles with butyl rubber stoppers and crimp caps, and flushed each bottle with nitrogen for 10 min (set **A**). To set **B**, we also added ~10 mmol g^{-1} VS carbon dioxide and hydrogen; to set **C**, we added ~6 mmol g^{-1} VS ethanol; and to set **D**, we added ~ 10 mmol g^{-1} VS carbon dioxide and hydrogen and ~6 mmol g^{-1} VS ethanol. The inoculum was collected from the well-mixed bioreactor before the substrate was changed from dilute-acid pretreated corn fiber to yeast fermentation beer on day 30. We periodically measured the headspace pressure in the batch bottles and analyzed headspace gas for methane, carbon dioxide, and hydrogen. We measured the liquid substrates and the products ethanol,

acetic acid, and *n*-butyric acid at the time of inoculation and at the end of 93 h. At the end of the run, we measured the volatile solids (VS) concentration in each bottle to normalize the measurements for the amount of biomass in each batch bottle.

Chemical Analysis. We performed all chemical analyses on a regular schedule. At the end of each 48-h feeding cycle we measured the biogas production and recorded the temperature and pressure to standardize the measurements. Biogas composition was measured weekly. For hydrogen composition, we used a Gow-Mac Series 580 GC (Gow-Mac Instrument Co, Bethlehem, PA) with a 5' x 1⁄4" stainless column packed with 60/80 Carboxen 1000 packing material (Supelco, Sigma-Aldrich, St. Louis, MO). The temperature of the column, injector, and detector were 100^{o}C, 110^{o}C, and 105^{o}C, respectively, and the current to the TCD detector was 70 mA. Carbon dioxide and methane were measured with an SRI 8610C GC with a 1m x 1⁄4" Rt-XLSulfur column (Restek, Corp, Bellefonte, PA). The temperature of the column, injector, and detector were 40^{o}C, 25^{o}C, and 101^{o}C, respectively, and the current was 167 mA. We determined the composition of the effluent and the stripping solution by measuring the individual carboxylic acids and ethanol concentration after every feeding cycle. Individual carboxylic acids were measured with an HP 5890 Series II GC (Hewlett-Packard, Palo Alto, CA) equipped with an autosampler with a 15m x 0.53mm Nukol column. Ethanol was measured with the same GC setup and a Supelco 6' 1⁄4" x 2mm glass column packed with 10% CW-20M (treated with 0.01% H3PO4) on 80/100 Chromasorb WAW support.

Genomic DNA (gDNA) extraction. We collected biomass by first mixing the bioreactors well for 5 min and sampling ~50 mL. We rapidly transferred three aliquots of sample to 2-mL vials and centrifuged at 10,000 rpm for 10 min, disposed of the supernatant, then froze them immediately at -80^{o}C until further analysis. We extracted genomic DNA (gDNA) from ~200 mg of biomass using the MoBio PowerSoil 96-well gDNA isolation kit (MoBio Labs, Inc, Carlsbad, CA), according to the MoBio protocol, except that cell lysis was performed by beadbeating.

16S rRNA Gene Sequencing and Data Analysis. To amplify 16S rRNA genes, PCR was carried out in triplicate for each sample and for water blanks. The PCR mastermix included 2.5 U Agilent Easy-A High Fidelity PCR Cloning Enzyme, 5 µl of 10X Easy-A reaction buffer

(Agilent Technologies, Inc., Santa Clara, CA), and 1μL each of 10 uM forward and reverse primers and 1 μL of 10mM dNTP. The forward primer combined the 454 primer 'B' and the universal bacterial primer 8F: 5'-GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG-3'. The reverse primer was a concatenation of the 454 primer 'A', followed by a barcode, unique for each sample, followed by the universal bacterial primer 338R: 5'-GCCTCCCTCGCGCCATCAGXXXXXXXXXXXXCATGCTGCCTCCCGTAGGAGT-3'. On each 96-well PCR plate we included negatives composed of randomly selected reverse primers and no template. Triplicates were pooled with the Mag-Bind EZ Pure magnetic purification kit (Omega Bio-Tek, Norcross, GA), and were eluted into 40 ul TE buffer according to the manufacturer's instructions. Pooled triplicates were run on a 1% agarose gel to verify the product. All negatives had no visible band and were not analyzed further. The concentration of dsDNA in each pooled triplicate was measured via fluourometric analysis with the PicoGreen dsDNA quantitation kit (Invitrogen Corp, Carlsbad, CA). The samples were pooled in equimolar amounts into a single sample with a final concentration 8.88 ng μl^{-1} dsDNA. Sequencing was performed on the Roche 454 pyrosequencing platform using Titanium chemistry and beginning sequencing at 454 adaptor A (Engencore, Columbia, SC).

We used the QIIME 1.4.0 pipeline [3] for sequence denoising, quality filtering, processing, and data analysis. Our efforts resulted in on average 5,875 high-quality 16S rRNA gene sequences per sample from which we picked and assigned taxonomy to 839 operational taxonomic units (OTUs; 97% ID). We determined OTUs whose relative abundance (i.e. the fraction of all sequence reads assigned to a specific OTU) was correlated to the *n*-caproic acid production rate with a correlation coefficient (r) of at least 0.8 and significance of correlation (p-value) < 0.05 (Table S1). We plotted the relative abundance vs. time for five of these OTUs that had a relative abundance by day 90 of at least 0.05 (Figure 2A). To determine how diverse the microbiome OTUs were, we calculated the phylogenetic diversity and the Gini coefficient and plotted both vs. time (Figure 2B). Phylogenetic diversity is a measure of the OTU diversity where higher numbers represent a more diverse sample. It uses the whole-sample phylogenetic tree to take into consideration the phylogenetic relatedness of OTUs. The Gini coefficient is a measure of microbiome evenness, where for a Gini coefficient of 0 (perfectly even) sequences are perfectly distributed between all OTUs, and 1 (perfectly uneven) would indicate that all

sequences are concentrated in one OTU. The measures were calculated with error bars by randomly subsampling 500 sequences from each sample, calculating both measures, repeating this process 100 times, and calculating standard deviation from the 100 replicate measurements. To measure changes in the overall sample microbiome structure, we calculated between-sample unweighted UniFrac distances, which takes into account not only changes in OTU presence/absence, but also in OTU phylogenetic relatedness. To visualize changes to the microbiome structure, we plotted the samples on the first two principal coordinates of the UniFrac distances, directly showing 49.8% of the total microbiome variation (Figure 2C).

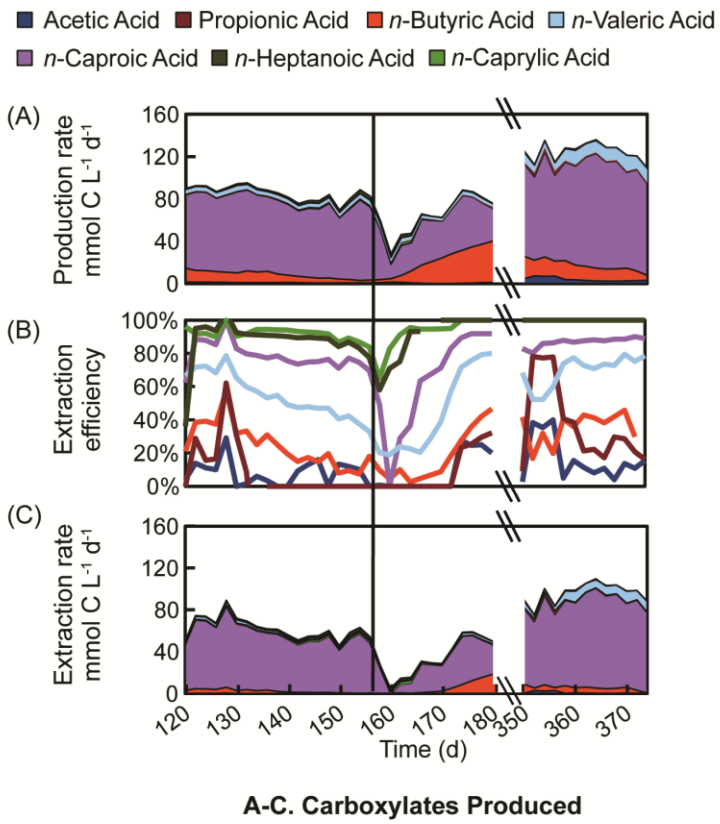
Shotgun metagenome sequencing and data analysis. We quantified gDNA extracted from 10 bioreactor samples via fluourometric analysis with the PicoGreen dsDNA quantitation kit (Invitrogen Corp, Carlsbad, CA). The samples were then sequenced on an Illumina HiSeq 2000 system on two lanes (Columbia NextGen Genome Center, Columbia University, New York, NY). Quality filtering was performed on the reads using a trimming threshold of two consecutive low-quality bases, no unknown bases, and a final minimum length of 75 bp, and removal of identical sequences, using the QIIME 1.4.0 pipeline [3]. We uploaded the reads to MG-RAST [4] for further analysis. After all quality control, each sample contained an average of 24,286,497 reads with an average length of 98 bp (~2.4 billion bp per sample). Using MG-RAST we created a functional (gene) annotation table by applying a maximum e-value cutoff of $1e^{-2}$ and a minimum percent identity of 50% to annotations based on the KEGG Orthology (KO) database. We used the metagenomic functional annotations as a way to supplement our findings in the 16S rRNA gene sequencing analysis of the communities. Specifically, we looked at the taxonomic breakdown of genes involved in conversion of yeast-fermentation beer to *n*-caproic acid.

We used two methods to determine the genes for which taxonomic identification can provide information about the role of bacteria in the bioreactor. First, we looked at genes significantly correlated with rates of *n*-caproic acid production (Figure S4). To do so, we converted the functional annotation table from MG-RAST into a table compatible with QIIME 1.4.0. We used QIIME to determine the genes whose relative abundance across the 10 samples was correlated with *n*-caproic acid production, limiting the analysis to genes that appeared in at least 5 of 10 samples. This produced 10 genes correlated at $R^2 > 0.9$ and with at least 1000 reads

per sample by day 126 (Table S3). Second, we looked at genes catalyzing the major steps leading to conversion of yeast-fermentation beer components to *n*-caproic acid (Figure S3). Specifically, we considered genes involved in: 1. Complex molecule hydrolysis (starch, xylan/xylose, and cellulose/cellobiose), 2. Glycolysis (we assume this was the central primary carbon metabolism pathway), 3. Ethanol oxidation, and 4. Chain elongation. Some of the enzymes could theoretically be used for other processes; for example, chain elongation genes could be involved in β-oxidation, but we labeled Figure S3 based on the most likely role of the gene in *n*-caproic acid production. We did not include any genes with less than 1000 assigned reads (Table S2). To simplify Figure S3, we also combined several steps for the processes: xylan hydrolysis, cellulose and cellobiose hydrolysis, ethanol oxidation, and glycolysis (Table S2). To determine the taxonomy distribution of genes, we used MG-RAST's "best hit" implementation to annotate read taxonomy at the level of genus by applying a maximum e-value cutoff of $1e^{-4}$ and a minimum percent identity of 50% to annotations based on the M5NR database. We used relatively strict cutoffs, because the best-hit method of gene annotation may produce faulty annotations, especially when two potential annotations have a close quality hit in the implemented BLAST search. Since our goal here is only to indicate potential roles of taxonomic groups, the annotations at these cutoff values should be sufficient.

References:

1. E. Alkaya, S. Kaptan, L. Ozkan, S. Uludag-Demirer and G. k. N. Demirer, *Chemosphere*, 2009, **77**, 1137-1142.
2. M. T. Agler, Z. Aydinkaya, T. A. Cummings, A. R. Beers and L. T. Angenent, *Bioresour. Technol.*, 2010, **101**, 5842-5851.
3. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight, *Nat. Methods*, 2010, **7**, 335-336.
4. F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening and R. Edwards, *BMC Bioinformatics*, 2008, **9**, 386.



A-C. Carboxylates Produced

Fig. S1. Performance of the *n*-caproic acid producing bioreactor on days 120-180 and 350-374, including the effects of an extraction module failure on day 157, and the maximum *n*-caproic acid production rate on day 364 after extraction rates were increased on day 300 due to increasing the extraction membrane surface area 3.5 times. Days 120-130 correspond to the same days in Figure 1: A. Production rate of C2-C8 carboxylic acids on days 120-180 and 350-374; B. Extraction efficiency of C2-C8 carboxylic acids as the percentage of produced acid that was extracted in-line; and C. Extraction rate of C2-C8 carboxylic acids is the rate of in-line recovery of each acid. The vertical line in A-C represents the failure of an extraction module on day 157. Performance data between days 180 and 350 was omitted to only focus on one membrane failure and the sustained production of *n*-caproic acid after installation of a new and larger extraction system on day 300. In the interim, problems arose from cracking membrane housing, and these problems disappeared after switching to stainless steel housing.

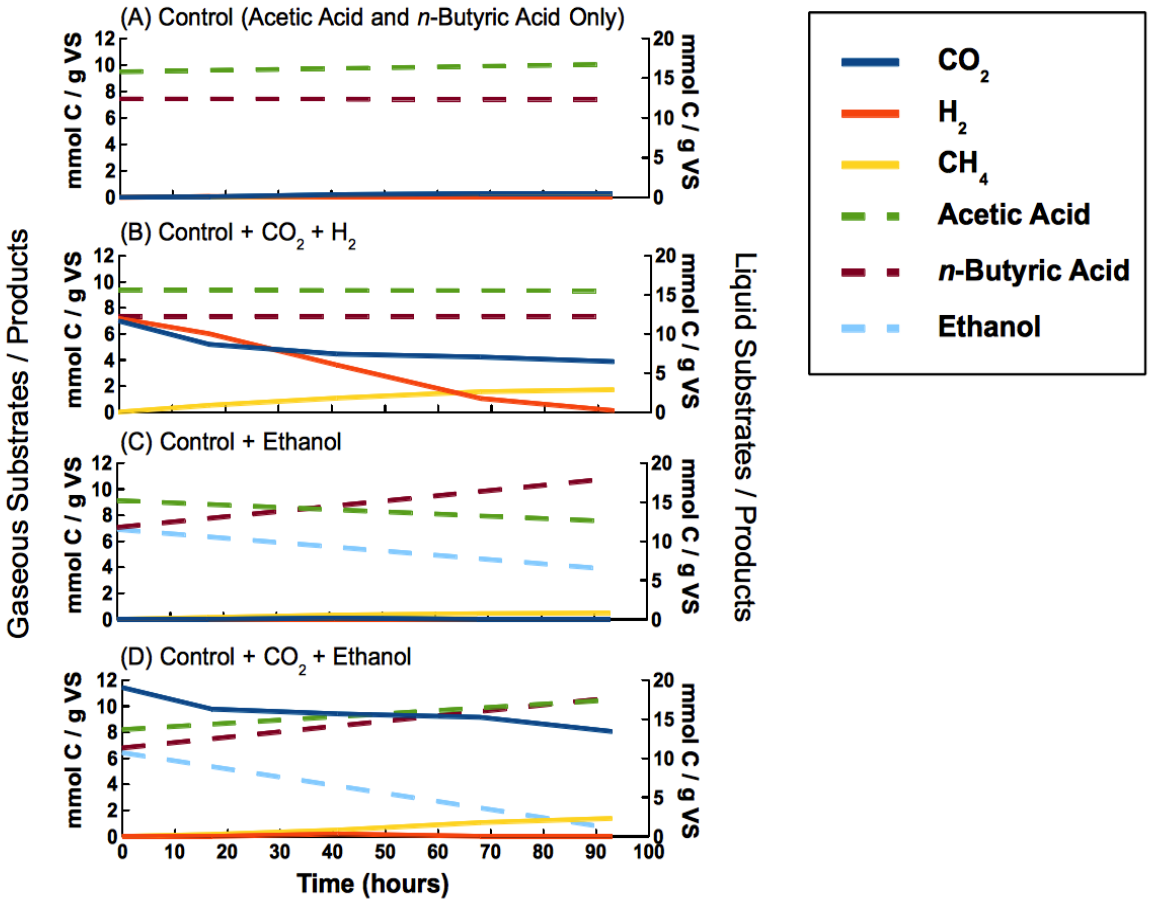


Fig. S2. A 93-h methanogenic activity test at pH 5.5 and 30°C demonstrates that microbiomes did not produce methane from acetic acid but produced it from CO_2 with H_2 or ethanol as sources of reducing equivalents: A. Control set demonstrates that methane is not produced by microbiomes in the presence of acetic acid; B. The microbiome produces methane from CO_2 and H_2 but acetic acid is not consumed; C. By consuming ethanol, the microbiome elongates some of the acetic acid to *n*-butyric acid (*n*-caproic acid was not detected); and D. When CO_2 and ethanol are added, microbiomes consume most of the ethanol because acetic acid is elongated to *n*-butyric acid and CO_2 is reduced to CH_4. Despite acetic acid elongation, acetic acid concentrations increase because ethanol is oxidized to acetic acid either by methanogens or by other microbes, which transfer H_2 to methanogens.

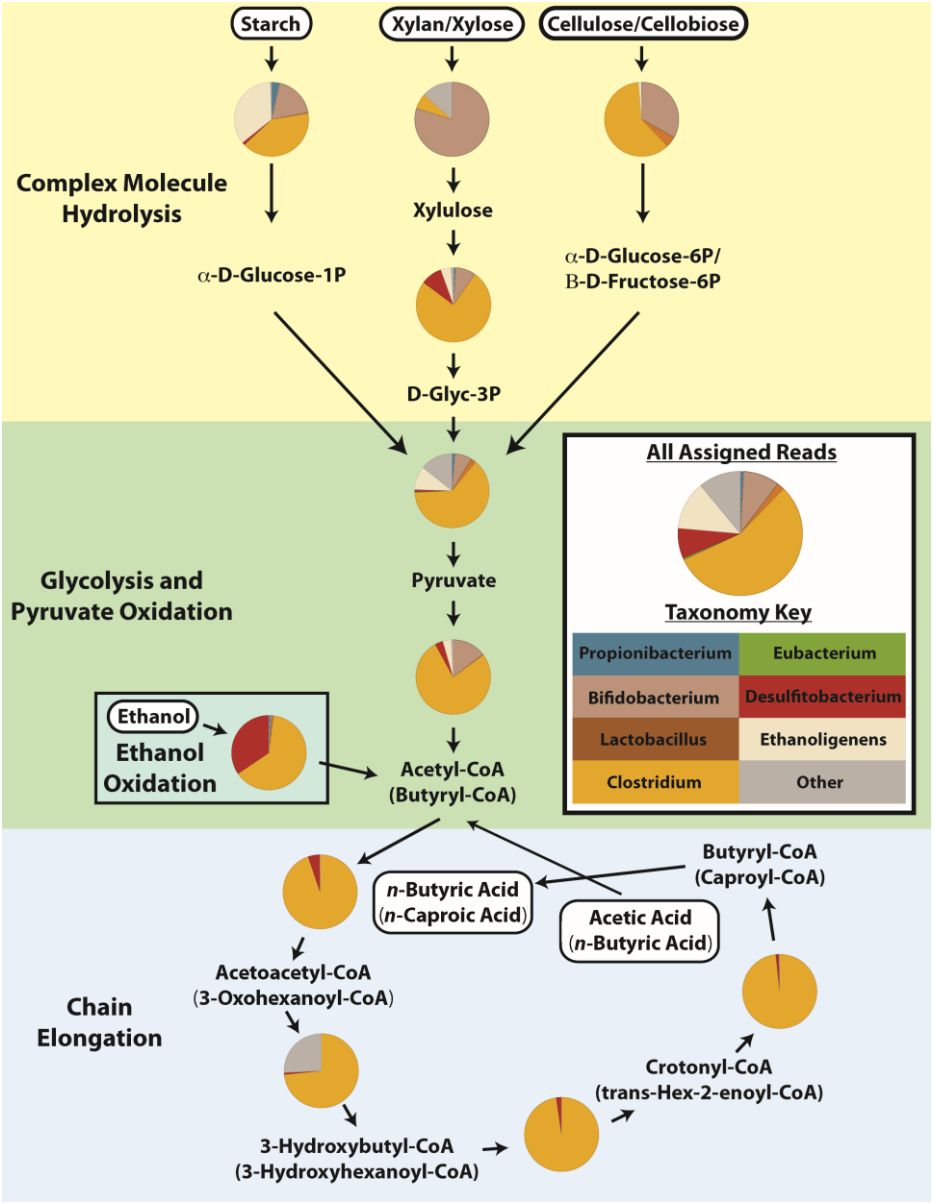


Fig. S3. Results of the shotgun metagenomic sequencing analysis. Taxonomic breakdown of gene pools involved in specific steps of carbon metabolism, indicate bacterial taxa that could be involved in specific steps of carbon metabolism. The figure is only a subset of all potential metabolic genes, but includes those that we expect to be important in conversion of yeast-fermentation beer to *n*-caproic acid. All data is from a sample taken on day 126 when conversion occurred at high rates. Extracellular molecules are highlighted in white. For simplicity, some enzymatic steps have been combined and genes with only a small number of assigned metagenomic reads (<1000) have been left out. For comparison, the taxonomic breakdown of all taxonomically-assigned reads is shown. The genera shown are the same as those shown in Figure S4A.

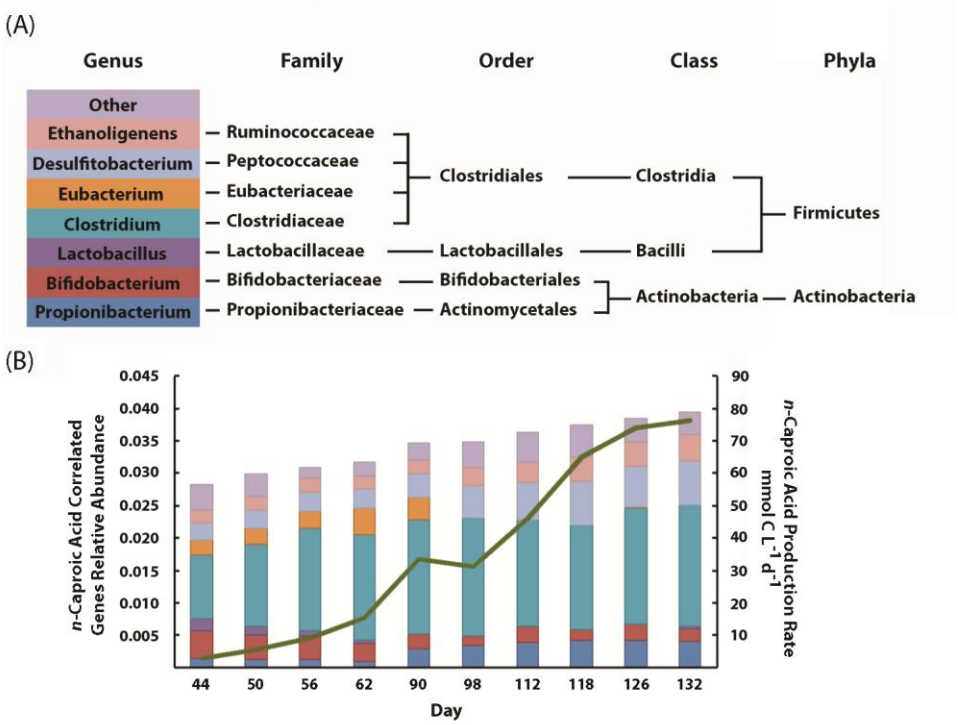
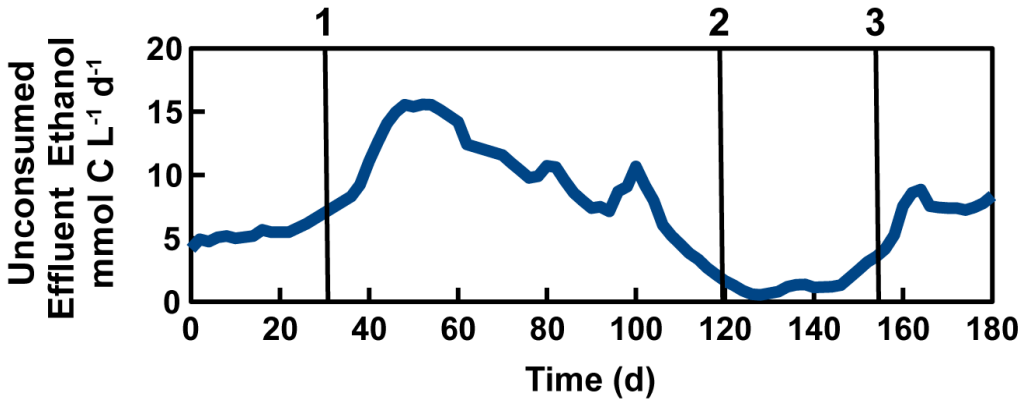


Fig. S4. Results of our shotgun metagenomic sequencing analysis. Genera important to *n*-caproic acid formation determined by the taxonomic breakdown of the seven genes most correlated to the production rate of *n*-caproic acid ($R^2 > 0.9$): A. Taxonomy key and phylogenetic relatedness of the bacterial genera; and B. Total relative abundance and breakdown of taxonomy for genes associated with production rate of *n*-caproic acid. The production rate of *n*-caproic acid (green line) is shown for reference.



300
301 **Fig. S5.** Ethanol concentration in effluent of the *n*-caproic acid producing bioreactor. Ethanol in
302 the effluent occurred when it was not oxidized during chain elongation. The black lines
303 represent: 1. The switch to yeast fermentation beer substrate; 2. The HRT decrease from 15 days
304 to 12 days, and 3. The failure of the extraction modules.
305

Table S1: Correlation coefficient (with significance) for OTUs whose relative abundance was positively correlated to the *n*-caproic acid production rate (with 16S rRNA gene sequencing analysis). [1]OTUs with r > 0.80 and p-value < 0.05 and whose relative abundance reached at least 0.05 by day 90 (Figure 2). [2]OTUs without taxonomic information could only be identified as belonging to the domain bacteria.

OTU	Significance (p-value)	Correlation Coefficient (r)	Phyla	Class	Order	Family	Genera	Species
[1]466	0.002	0.960	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiales_Family_XI_Incertae_Sedis	g__Sporanaerobacter	s__Sporanaerobacter_acetigenes
[1]733	0.008	0.926	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
746	0.008	0.926	[2]					
260	0.010	0.917	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
521	0.011	0.912	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__Clostridium_kluyveri
[1]2	0.015	0.899	[2]					
[1]97	0.015	0.897	[2]					
[1]225	0.016	0.895	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__Clostridium_kluyveri
649	0.027	0.864	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
792	0.033	0.849	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
674	0.041	0.830	p__Firmicutes					
551	0.052	0.808	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Oscillospira	otu_2126
123	0.070	0.776	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
509	0.086	0.749	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__Clostridium_kluyveri
13	0.097	0.734	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
695	0.098	0.732	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	
289	0.099	0.731	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	
251	0.099	0.731	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	
26	0.106	0.721	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
406	0.112	0.712	p__Bacteroidetes	c__Bacteroidia	o__Bacteroidales	f__Porphyromonadaceae		
634	0.134	0.684	[2]					
463	0.135	0.683	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
38	0.140	0.677	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
578	0.152	0.662	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	
415	0.247	0.561	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
287	0.253	0.555	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
680	0.254	0.554	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
259	0.287	0.523	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
441	0.305	0.507	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Oscillospira	otu_2126
284	0.332	0.483	[2]					
673	0.370	0.450	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
374	0.378	0.444	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
180	0.380	0.442	p__Firmicutes					
586	0.380	0.442	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	
286	0.419	0.410	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	
33	0.422	0.408	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
144	0.433	0.400	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
497	0.434	0.398	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
637	0.449	0.386	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
620	0.528	0.326	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
490	0.531	0.324	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__Clostridium_acetobutylicum
278	0.576	0.291	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
402	0.582	0.286	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__Clostridium_kluyveri
689	0.625	0.255	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	
719	0.646	0.241	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
282	0.649	0.239	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae	g__Clostridium	otu_2116
176	0.678	0.218	[2]					
369	0.685	0.214	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
550	0.695	0.206	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
457	0.723	0.187	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
1	0.761	0.161	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		
742	0.771	0.154	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Ruminococcaceae		

OTU	Significance (p-value)	Correlation Coefficient (r)	Phyla	Class	Order	Family	Genera	Species
440	0.802	0.133	**p__Firmicutes**	**c__Clostridia**	**o__Clostridiales**	**f__Lachnospiraceae**		
428	0.860	0.094	**p__Bacteroidetes**	**c__Bacteroidia**	**o__Bacteroidales**			**otu_973**
119	0.919	0.054	**p__Firmicutes**	**c__Clostridia**	**o__Clostridiales**	**f__Ruminococcaceae**		
439	0.967	0.022	**p__Firmicutes**	**c__Clostridia**	**o__Clostridiales**	**f__Ruminococcaceae**		**otu_2109**
803	0.975	0.017	**p__Firmicutes**	**c__Clostridia**	**o__Clostridiales**	**f__Ruminococcaceae**	**g__Clostridium**	**otu_2116**

Table S2: Genes included in the analysis for Figure S3 (with shotgun metagenomic sequencing analysis). Only those genes with >1000 assigned reads are included in the figure. [1]The transketolase involved in conversion of xylulose to a glycolysis intermediate is separated from other xylan degradation genes because the taxonomy distribution of the gene was very different.

Pathway Involvement	EC Number	Name	>1000 Reads
Starch to Glycolysis	2.4.1.1	Phosphorylase	Yes
Xylan/Xylose to Xylulose	3.2.1.37	Xylan 1,4-β-Xylosidase	Yes
	1.1.1.21	Aldehyde Reductase	No
	1.1.1.9	D-Xylulose Reductase	Yes
	5.3.1.5	Xylose Isomerase	Yes
	2.7.1.17	Xylulokinase	No
[1]Xylulose to Glycolysis	2.2.1.1	Transketolase	Yes
Cellulose to Glycolysis	3.2.1.4	Cellulase	No
	3.2.1.91	Cellulose 1,4-β-cellobiosidase	No
	3.2.1.21	β -glucosidase	Yes
	2.7.1.1	Hexokinase	No
	2.7.1.2	Glucokinase	No
	2.7.1.63	Polyphosphate-glucose phosphotransferase	No
	5.1.3.15	Glucose-6-phosphate 1-epimerase	No
	5.1.3.3	Aldose 1-epimerase	No
Glycolysis to Pyruvate	5.4.2.2	Phosphoglucomutase	Yes
	5.3.1.9	Glucose 6-phosphate isomerase	Yes
	2.7.1.11	6-Phosphofructokinase	Yes
	2.7.1.146	ADP-specific phosphofructokinase	No
	4.1.2.13	Fructose-bisphosphate aldolase	Yes
	1.2.1.12	Glyceraldehyde 3-phosphate dehydrogenase	Yes
	1.2.1.59	Glyceraldehyde 3-phosphate dehydrogenase (NADP)	No
	5.3.1.1	Triose-phosphate isomerase	Yes
	2.7.2.3	Phosphoglycerate kinase	Yes
	5.4.2.4	Bisphosphoglycerate mutase	No
	3.1.3.13	Bisphosphoglycerate phosphatase	No
	5.4.2.1	Phosphoglycerate mutase	Yes
	4.2.1.11	Phosphopyruvate hydratase	Yes
	2.7.1.40	Pyruvate kinase	No
Pyruvate to Acetyl-CoA	2.3.1.54	Formate *C*-acetyltransferase (Pyruvate-formate lyase)	Yes

Pathway Involvement	EC Number	Name	>1000 Reads
Ethanol to Acetyl-CoA	1.1.1.1	Alcohol dehydrogenase	Yes
	1.1.1.2	Alcohol dehydrogenase (NADP)	No
	1.2.1.3	Aldehyde dehydrogenase (NAD)	No
	1.2.1.5	Aldehyde dehydrogenase (NADP)	No
Chain Elongation	2.3.1.9	Acetyl-CoA *C*-acetyltransferase	Yes
	1.1.1.36	Acetoacetyl-CoA reductase	No
	1.1.1.35	3-Hydroxylacyl-CoA dehydrogenase	No
	1.1.1.157	3-Hydroxybutyryl-CoA dehydrogenase	Yes
	4.2.1.55	3-Hydroxybutyryl-CoA dehydratase	Yes
	4.2.1.17	Enoyl-CoA hydratase	No
	1.3.8.1	Butyryl-CoA dehydrogenase	Yes
	2.8.3.6	3-Oxoadipate CoA transferase	No

Table S3: Coefficient of correlation to *n*-caproic acid production rate for genes included in the analysis for Figure S4 (with shotgun metagenomic sequencing analysis). Only those genes with >1000 assigned reads are included in the figure.

EC Number	Name	Correlation Coefficient [r | r^2]	>1000 Reads
2.4.1.7	**Sucrose phosphorylase**	**0.98 | 0.95**	**Yes**
2.4.2.2	**Pyrimidine-nucleoside phosphorylase**	**0.97 | 0.94**	**Yes**
1.17.4.2	**Ribonucleoside-triphosphate reductase**	**0.96 | 0.94**	**Yes**
5.4.2.7	**Phosphopentomutase**	**0.96 | 0.92**	**Yes**
4.2.3.3	**Methylglyoxal synthase**	**0.96 | 0.92**	**Yes**
5.4.2.8	**Phosphomannomutase**	**0.95 | 0.91**	**Yes**
5.4.3.3	**β-lysine 5,6 aminomutase**	**0.95 | 0.91**	**No**
2.2.1.1	**Transketolase**	**0.95 | 0.90**	**Yes**