Structure and dynamics of the microbial communities underlying the carboxylate platform for biofuel production

E.B. Hollister [1], A.K. Forrest [2], H.H. Wilkinson [3], D.J. Ebbole [3], S.A. Malfatti [4], S.G. Tringe [4], M.T. Holtzapple [2], T.J. Gentry [1]

[1] Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA 77843-2474

[2] Department of Chemical Engineering, Texas A&M University, College Station, TX, USA 77843-3122

[3] Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX USA 77843-2132

[4] DOE Joint Genome Institute, Walnut Creek, CA, USA 94598

July 31, 2010

ACKNOWLEDGMENT

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

Title: Peering into the black box of bioconversion: Characterizing the mixed-microbial communities involved in the carboxylate platform for biomass conversion

Running title: Carboxylate platform microbial communities

Authors: E.B. Hollister [1*], A.K. Forrest [2], H.H. Wilkinson [3], D.J. Ebbole [3], S.A. Malfatti [4], S.G Tringe [4], M.T. Holtzapple [2], T.J. Gentry [1]

[1] Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA 77843-2474

[2] Department of Chemical Engineering, Texas A&M University, College Station, TX, USA 77843-3122

[3] Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX USA 77843-2132

[4] DOE Joint Genome Institute, Walnut Creek, CA, USA 94598

[*] Corresponding author: Mailing address: Department of Soil and Crop Sciences, Texas A&M University, 370 Olsen Blvd, TAMU 2474, College Station, TX, USA 77843-2474, Phone: 1-979-845-5695, Fax: 1-979-845-0456, Email: ehollister@tamu.edu

Abstract

The carboxylate platform employs a mixed microbial community to convert lignocellulosic biomass into chemicals and fuels. Although many of its components are well understood, this study is the first to characterize its microbiology. Mesophilic (40 °C) and thermophilic (55 °C) fermentations employing a sorghum biomass feedstock and a marine sediment inoculum source were profiled using 16S rRNA tag-pyrosequencing at multiple points over the course of a 30-day incubation. The contrasting fermentation temperatures converted similar amounts of biomass, but the mesophilic condition was significantly more productive than the thermophilic condition, and the two differed significantly with respect to propionic and butyric acid production. Pyrotag sequencing revealed the presence of dynamic communities that responded rapidly to temperature and changed substantially over time. Both temperatures were dominated by bacteria associated with the class *Clostridia,* but they shared few taxa in common. The species-rich mesophilic community harbored a variety of *Bacteroidete*s, *Actinobacteria*, and γ-*Proteobacteria*, whereas the thermophilic community was much simpler, composed mainly of *Clostridia* and *Bacilli*. Despite differences in composition and productivity, similar dynamics were observed with respect to the relative abundance of functional classes. Over time, organisms related to known cellulose-degraders decreased in abundance, while organisms related to those capable of utilizing xylose and other degradation byproducts appeared to increase. Improved understanding of the microbiology of the carboxylate platform will help refine its design and operation, facilitate the development of inocula for its large-scale implementation, and contribute to the developing body of knowledge regarding biomass conversion and biofuel production.

Keywords: mixed alcohol bioreactor/cellulosic biofuels/tag-pyrosequencing/microbial communities/carboxylate platform

Introduction

As energy demands continue to increase and the world's fossil fuel reserves steadily decline, the need to develop stable, alternative energy sources continues to grow. Next-generation biofuels (i.e. those derived from cellulosic feedstocks) have the potential to contribute to meeting a portion of those needs (32, 35). It is estimated that the United States has the potential to produce in excess of 1.3 billion tons of biomass feedstock each year for the production of lignocellulosic biofuels (32). Although biomass-based fuels cannot be produced on such a scale as to replace petroleum-based fuels entirely, it is thought that they have the potential to fulfill ~30% of the U.S. annual transportation fuel demands (32).

Multiple strategies have emerged to convert lignocellulosic feedstocks into fuels. The two most common are the sugar platform (i.e., the enzymatic hydrolysis of cellulose into simple sugars followed by aseptic fermentation) and thermochemical conversion (i.e., gasification followed by catalysis). Although each of these has proven useful, their large-scale implementation has been hampered by high costs associated with their construction, processing, and maintenance (17, 29). A lesser known alternative to these strategies is the carboxylate platform, also known as the MixAlco process (22).

The carboxylate platform is a well-optimized, flexible, and cost-effective means to convert lignocellulosic feedstocks -- including crop residues, dedicated energy crops, municipal solid waste, and biosolids -- into chemicals and liquid fuels (2, 18, 22, 39). Contributing to its cost-effectiveness, the carboxylate platform operates under non-sterile conditions and uses a mixed community of naturally occurring anaerobic microorganisms to provide the enzymes necessary for biomass conversion. The mixed-culture approach of this system underlies its flexible nature as well, allowing the platform to convert sugars, as well as proteins, pectin, oils,

and fats. The products from the mixed-culture fermentation are carboxylate salts, which can be transformed through chemical means to a wide variety of products, including alcohols, jet fuel, and gasoline. In addition, the product spectrum from this process is temperature dependent (6, 15) and can be varied in response to market demands.

Over the course of its development, a number of different inoculum sources have been utilized in the carboxylate platform (39). Despite these attempts to identify better-performing microbial communities, the platform's microbial component has remained a "black box" relative to other well optimized and understood aspects of this system. One of the best performing inoculum sources that has been found, to date, is a marine sediment that was collected along the Gulf of Mexico coastline at Galveston Island, TX (39). While this community appears to be well-adapted to the conditions of the carboxylate platform, it is unclear which microorganisms are present in this consortium and, of those, which are involved in the process of biomass conversion in the carboxylate platform.

The objective of this study was to characterize the carboxylate platform's bacterial communities, seeded from the Galveston inoculum, over the course of a 30-day, laboratory-scale incubation, and under both mesophilic (40 °C) and thermophilic (55 °C) conditions. This community-focused approach, involving the production of large scale tag-pyrosequence libraries, was used in combination with reactor performance metrics to characterize the composition of the reactor microbial communities and the relationships that they share with the system's productivity and performance. Increased understanding of the microbial communities involved in the carboxylate platform will help refine the design and operation of the bioreactor system, facilitate the development of inocula for large-scale implementation of this process, and

92 contribute to the developing body of knowledge regarding biomass conversion and biofuel
93 production.
94
95 **Materials and Methods**
96 *Bioreactor substrate, inoculum source, and pre-incubation preparation*
97 A photoperiod-sensitive, high-tonnage sorghum cultivar (*Sorghum bicolor* (L.) Moench)
98 was used as a biomass feedstock and was obtained from William Rooney (Sorghum Breeding
99 and Genetics Program, Texas A&M University). The sorghum biomass was chipped, dried, and
100 treated with hot water and lime (0.1 g $Ca(OH)_2$) and 10 mL distilled H_2O per g dry biomass; 2 h
101 at 100 ºC) to enhance its digestibility (7).
102 The inoculum used in the reactor system was obtained from marine sediment collected
103 from Galveston, TX in May 2008. Sediment was collected from multiple shoreline pits dug to
104 0.5 m depth, adjacent to the high water mark at the time of sampling. At 0.5 m depth, there was
105 a distinct color change from yellow/brown sand to dark gray/black sediment. Upon collection,
106 sediment samples were placed into bottles filled with deoxygenated water, 0.275 g L^{-1} sodium
107 sulfate, and 0.275 g L^{-1} cysteine hydrochloride, as described by Thanakoses *et al.* (39). The
108 bottles were capped and held on ice. Upon return to the laboratory, samples were frozen at -20 ºC
109 until later use. Just prior to inoculation, a sediment sample was thawed, shaken vigorously, and
110 allowed to settle by gravity. Aliquots of the resulting supernatant then used to inoculate the
111 bioreactor vessels.
112
113 *Bioreactor construction and performance monitoring*

Bioreactor vessels were constructed using 1-L polypropylene centrifuge bottles fitted with a stirring and venting apparatus, as described by Fu and Holtzapple (15). Each reactor contained 50 mL mixed microbial inoculum, 36 g lime-treated sorghum, 4 g dried chicken manure (obtained from the Poultry Science Center at Texas A&M University, College Station, TX), 350 mL deoxygenated water, a calcium carbonate buffer ($CaCO_3$, 15 g L^{-1}), and iodoform (CH_3I, 20 g L^{-1}) and was rolled continuously at 2 rpm over the course of the incubation (15). Reactors were flushed with N_2 during the preparation process to create an anaerobic environment, and the iodoform was used to inhibit methane production. Two fermentation temperatures (40 and 55 ºC) were employed, and the experiment was designed so that a set (n = 3) of reactor vessels from each temperature treatment could be sacrificed for microbial community characterization at the early (day 6), middle (day 16), and late (day 30) stages of the incubation.

Carbon dioxide (CO_2) and methane concentrations, as well as pH and total carboxylic acid concentrations were monitored every two days over the course of the incubation. The volume of gas produced was quantified by displacing a $CaCl_2$ solution, and its CO_2 and CH_4 concentrations were measured using an Agilent 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a thermal conductivity detector. pH was measured and monitored using an ORION pH meter (Thermo Electron Corporation, Beverly, MA, USA), and total carboxylic acid concentrations in the fermentation broth were measured with an Agilent 6890 gas chromatograph (Agilent Technologies) equipped with a flame ionization detector and an Agilent 7683 series injector (Agilent Technologies).

The products in the fermentation broth consisted of a mixture of carboxylic acids and carboxylate salts. Samples of the fermentation broth were combined with 1.162 g L^{-1} of a 4-

137 methyl-*n*-valeric acid internal standard and were acidified with 3 M phosphoric acid prior to
138 injection into the gas chromatograph. By acidifying this mixture, each component was converted
139 to its corresponding acid, allowing product concentrations to be reported as carboxylic acids in g
140 L^{-1}. Column head pressure was maintained at 2 atm abs, and samples were heated from 40 °C to
141 200 °C at a rate of 20 °C min^{-1} prior to being held at 200 °C for an additional 2 min. Helium was
142 used as the carrier gas, and the total run time per sample was 11 minutes.
143 As each subset of the fermentation reactions was terminated (i.e., early, middle, and late
144 stages of incubation), samples of both the solid and liquid phases were collected from the reactor
145 vessels for chemical analysis. The reactor bottles were centrifuged, in a Beckman J-6B
146 centrifuge (Beckman Coulter, Inc., Brea, CA, USA) with a swinging bucket rotor at 3297 x g for
147 30 minutes to separate the solids and liquids from one another. An aliquot of the supernatant
148 was collected and subjected to carboxylic acid analysis, as described above, and the solids were
149 analyzed to determine the mass of undigested volatile solids (VS) first by drying at 105°C and
150 then by ashing at 550°C (15). VS content was calculated as the difference between the oven dry
151 weight and the ashed weight of each sample.

152 *DNA extraction and pyrotag sequencing*

153 Reactor materials were collected early in the incubation (day 6), mid-incubation (day 16),
154 and at the end of the incubation (day 30). Solids and liquids from each replicate were combined
155 into one vessel in equal volumes to create a single composite sample for each treatment, at each
156 time point. The composite samples were frozen and stored at -80 ºC until DNA extraction. Just
157 prior to extraction, the reactor samples were thawed and centrifuged at 4000 x g for 10 min.
158 DNA was extracted from the pellet materials using a PowerMax soil DNA extraction kit (Mo
159 Bio Laboratories, Inc., Carlsbad, CA, USA), using a lysozyme-modified version of the

manufacturer's protocol (21). Briefly, 15 g aliquots (wet weight) of reactor material and 15 mL of bead solution were added to each bead beating tube. After 5 min of bead beating, lysozyme was added (final concentration of 1 mg mL^{-1}), and samples were incubated at 37 ºC for 1 h. Following lysozyme treatment, solution "C1" was added and samples were incubated at 65 ºC for 30 min. The manufacturer's protocol was followed from this point onward. Following elution, DNA samples were concentrated and purified using illustra MicroSpin S-400 HR columns (GE Healthcare Bio-Sciences Corp, Piscataway, NJ, USA).

DNA samples were quality checked according to US DOE Joint Genome Institute protocols and were submitted to the Joint Genome Institute for pyrotag sequencing. Hypervariable regions V6-V8 of the 16S rRNA gene were amplified with universal primers 926F and 1392R, including the titanium adapter sequences as well as a 5-base barcode on the reverse primer (12). Reactions were performed in triplicate with 25 cycles of amplification, checked by agarose gel electrophoresis then quantified and pooled in equal amounts for sequencing by standard 454-titanium protocol (12).

Analysis of reactor performance

Reactor performance was characterized through a series of metrics that included conversion, selectivity, yield, and productivity. *Conversion* was calculated as the fraction of VS that was digested over the course of the incubation relative to the quantity of VS contained in the biomass that was initially loaded into the reactor. *Selectivity* was calculated as the proportion of digested material that resulted in carboxylic acid production. *Yield* was determined by calculating the ratio of total carboxylic acids produced relative to the quantity of VS contained in the biomass that was initially loaded into the reactor, and *productivity* was calculated as the

183 overall rate of acid production (g acid L^{-1} d^{-1}). Comparisons of these values, as well as the
184 relative abundances of various acid products, between fermentation temperatures were conducted
185 using paired, two-tailed Student's t-tests, and p-values < 0.05 were considered to represent
186 significant differences.

187 *Microbial community characterization and comparisons*

188 Bioreactor community DNA sequences were quality checked, trimmed to a common
189 length, and clustered (97% similarity) using the computational pipeline described by Kunin et al.
190 (26) and Engelbrektson et al. (12). A representative sequence from each cluster was compared
191 against the Greengenes NAST-aligned database (11) and used to assign putative identities to the
192 entire sequence data set. The quality-checked output was reformatted and input into the
193 Ribosomal Database Project (RDP) pyrosequencing pipeline (9) and the Mothur software
194 package (36) to perform comparisons of the bioreactor bacterial communities across temperature
195 treatments and over time. RPD was used to align sequences, and Mothur was used to calculate
196 distance matrices, assign sequences to operational taxonomic units (OTU, 97% similarity),
197 calculate diversity indices and richness estimates, and determine the degree of overlap shared
198 among the reactor communities. Overlap was calculated using the Yue-Clayton similarity
199 estimator (θ_{YC}), a metric that is scored on a scale of 0 to 1, where 0 represents complete
200 dissimilarity and 1 represents identity (36, 42). When comparing any given set of communities,
201 θ_{YC} considers the distribution of OTUs between the communities, as well as their relative
202 abundance. Additional comparisons of community structure were carried out using nonmetric
203 multidimensional scaling of the OTU data using the Bray-Curtis similarity metric, as
204 implemented in the PAST software program (19).

In an attempt to link community composition to reactor performance, the relative abundance of each OTU was regressed against the reactor performance metrics described above. Given the size of the data set, only those OTUs that accounted for ≥ 1% of the entire community, at any given time point, were considered. Those that had r^2 correlations of 0.5 or better were considered to be "performance related" and were subject to additional analysis. Nearest-neighbor sequences for each of the performance-related OTUs, as well as the five most abundant OTUs present in the middle-of-incubation communities were identified and used to build phylogenetic trees to illustrate the relationships shared among them and aid in their classification. Three nearest neighbor sequences for each reactor OTU were obtained from the Greengenes chimera-checked database (11). Sequences were aligned using the Aligner feature at the Ribosomal Database Project (9), they were bootstrapped with 100 replicates using Phylip's Seqboot function, and then the corresponding distance matrices were calculated using DNAdist with the Jukes-Cantor evolutionary model (13). Phylip's Neighbor function was used to construct neighbor-joining phylogenetic trees, and the Consense function was used to identify a consensus tree (13).

Pyrotag sequence reads were submitted to the NCBI Short Read Archive under accession numbers (in process).

Results

Reactor performance

After 30 days' incubation, the 40 and 55 ºC fermentation temperatures resulted in similar rates of biomass conversion (Table 1); however, the reactors incubated under these temperature regimes differed significantly with respect to their selectivity, yield, and productivity. Each of

these performance metrics was significantly greater at 40 ºC than at 55 ºC, representing increases of 40 to 60% over the 55 ºC values. The two temperature conditions also differed significantly with respect to the abundances of multiple acids within their product spectra (Table 2). Although acetic acid (C_2) dominated and was equally abundant in both reactor systems, the fermentations at 40 ºC produced a significantly greater proportion of propionic (C_3) acid products and a significantly smaller proportion of butyric (C_4) acid products than those at 55 ºC.

Microbial community characterization and comparisons between temperatures and over time

Following the removal of short, ambiguous, and/or low-quality pyrotag reads, the final data set, representing both fermentation temperatures and all time points, consisted of 160,957 high-quality partial 16S rRNA sequences. Sequence library size ranged from 15,344 sequences in the 55 °C-mid incubation sample to 46,359 sequences in the 40 °C-early incubation sample (Table 3). Among all samples 2955 unique OTUs (97% similarity cutoff) were identified. The number of OTUs contained within each sequence library covered an order of magnitude difference, with as few as 140 OTUs occurring in the 55 ºC-mid incubation sample and as many as 1382 OTUs occurring in the 40 ºC-early incubation sample. Likewise, estimates of community diversity (H') and richness (Chao I) varied widely along temperature lines (Table 3).

The reactor communities were dominated by members of the phylum *Firmicutes*, and taxa representing the bacterial classes *Clostridia* and *Bacilli* were common to both fermentation temperatures and all time points. Although the reactor communities were similar to one another at broad taxonomic levels, their OTU composition differed dramatically with respect to incubation temperature and changed substantially over time (Figure 1). Despite the predominance of *Clostridia* and *Bacilli* in the reactors communities (Figure 2), very little overlap was detected between the two fermentation temperatures with respect to species-level

community membership and structure (Table 4). Among the 2955 OTUs identified, only 2% were found at both 40 and 55 ºC. The segregation of taxa by fermentation temperature can also be seen in Figure 3, a neighbor-joining tree of the most abundant OTUs encountered in the mid-incubation communities. Among the OTUs represented, only clusters #1 (55 °C-specific) and #28 (40 °C-specific) share a close, common position on the tree.

Differences among the reactor communities that occurred with respect to time are indicated by Yue-Clayton similarity values in the highlighted portions of Table 4. Samples that were taken closer in time (i.e., early- vs. mid-incubation or mid- vs. late-incubation) were more similar to one another than were those that were taken following longer gaps in time (i.e., early- vs. late-incubation). In the early stages of incubation, members of the class *Clostridia* dominated the reactor communities at both temperatures. A pair of *Clostridiales*-like OTUs accounted for 34% and 19% of the early incubation sequences in the 55 °C and 40 °C communities, respectively (Supplemental table 1). With time, other taxa increased in abundance and displaced these *Clostridiales*-like OTUs as the predominant members of each community. In the 40 °C communities, *Bacteroidetes*, particularly OTUs resembling members of the *Prevotellaceae*, became the most commonly detected taxa, and members of the γ-*Proteobacteria* and *Actinobacteria* increased in abundance as well. In the 55 °C communities, OTUs resembling *Thermoanaerobacterium* spp. dominated the mid-incubation sequence library, and a *Clostridium cellulolyticum*-like OTU dominated the late-incubation sequence library. Concurrent with shifts in dominance among various types of *Clostridia* in the 55 °C community, increases were also detected with respect to the relative abundance of *Bacillus* spp. Additional information regarding the relative abundances of these OTUs and their identities at both temperatures and at each time point along the study can be found in Supplemental Table 1.

Subpopulations correlating with reactor performance

Between the two fermentation temperatures, ten OTUs were identified as sharing strong correlations with reactor productivity (Table 5). While few of these OTUs displayed temperature-exclusive distributions (i.e. Clusters #2, 27, and 28 only occurred at 40°C), most displayed strong temperature affiliations (i.e. the vast majority of sequences affiliated with each OTU were found at one fermentation temperature, rather than being equally distributed across both). Of the ten OTUs identified, half were found to be correlated positively with reactor productivity rates. These included Clusters #1, 2, 5, 14, and 28, each of which was found to be closely related to a *Clostridium* sp (Figure 4). In contrast, the other five OTUs were negatively correlated with reactor productivity rates and drew from a much broader taxonomic range, including *Actinobacteria*, *Proteobacteria*, members of the class *Bacilli*, *Symbiobacteria*, and other *Clostridium*-like organisms. The phylogenetic relationships of these organisms and their closest characterized isolates are illustrated in Figure 4.

Discussion

Although many aspects of the carboxylate platform are well optimized and understood, this study represents the first attempt to characterize the microbial portion of this system. While PCR-based characterizations of community composition, including tag-pyroseqencing, are not without potential biases [e.g., DNA extraction efficiency, the potential for bacteria to carry multiple copies of the 16S gene, primer bias, and sequencing error (26, 27, 40)], they provide valuable insight into communities that contain taxa that are difficult to cultivate and can be useful in studying temporal changes within the same ecosystem (14). Furthermore, recent evidence suggests that thorough controls on data quality and clustering can help to eliminate

297 some of the potential biases associated with pyrotag-induced errors (26). With these caveats in
298 mind, these results suggest that the bacterial communities that underlie biomass conversion in the
299 carboxylate platform are dynamic, respond rapidly to altered temperature conditions, and change
300 substantially over time. The reactors were dominated, in large part, by bacteria associated with
301 the class *Clostridia*. Although the best match to many of these were unclassified *Clostridium*-like
302 organisms originating from a broad range of habitats, this group also included organisms similar
303 to a number of well characterized biomass-degrading and potentially cellulolytic *Clostridium* spp
304 and *Thermoanaerobacterium* spp (Supplemental Table 1). In addition to *Clostridia*, members of
305 the class *Bacilli* were particularly abundant in the 55 °C community, and members of the
306 *Bacteroidetes* were commonly encountered in the 40 °C community.
307 Studies of other biomass conversion systems have isolated and characterized similar taxa
308 in attempts to identify organisms involved in the conversion of biomass and solid waste into
309 acids, alcohols, biogas, and hydrogen (5, 16, 25, 33, 37). *Clostridia* often dominate these
310 systems, though they are frequently found in co-culture with a number of other species. The
311 bacterial class *Clostridia* includes a number of anaerobic, thermophilic species that are
312 commonly associated with the decomposition of lignocellulose and municipal solid waste (5, 8),
313 and many are thought to work most effectively when grown in mixed culture (10, 24).
314 At the laboratory scale, co-cultures of *Clostridium thermocellum* with
315 *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter ethanolicus,* or other
316 species have been used to maximize the hydrolysis and fermentation of lignocellulosic and
317 hemicellulosic feedstocks (4, 10, and references therein). *C. thermocellum* produces cellobiose
318 and cellodextrins from cellulose for its own use, and it participates in the enzymatic
319 saccharification of hemicelluloses to xylose and xylobiose, substrates that can be used by the co-

cultured xylose-degrader (10). Such interactions have been demonstrated in a number of systems, including natural environments, industrial settings, and synthetically-constructed communities (10, 24, 30, 38), and they are considered to represent the model for co-culture cellulose degradation (4).

Sorghum, a complex material containing 29% cellulose, 26% hemicellulose, and 8% lignin on a dry weight basis (34), should be able to provide substrates that support a wide range of metabolisms, including the co-culture model. Based upon the results of this study, this appears to be the case. The communities associated with each of our fermentation temperatures harbored subsets of OTUs that correlated well with reactor productivity and adhere to the cellulose degrader/non-cellulose degrader co-culture model (10, 24, 30). Each community contained two or three dominant *Clostridium*-like OTUs that were positively correlated with reactor productivity and are likely to play an integral role in cellulose degradation within the carboxylate platform. In complement, each community also harbored two or three emergent OTUs that were negatively correlated with reactor productivity and possibly represent the xylose-degrading component of the platform. This group included a *Bacillus*-like bacterium, a *Symbiobacterium*-like bacterium, and a *Clostridium stercorarium*-like bacterium, each of which was detected in the middle- and late-stages of the thermophilic fermentation. In the mesophilic fermentation, this group consisted of taxa that were similar to members of the *Actinobacteria* and *Proteobacteria* (Table 5).

Previous work has demonstrated that the spectrum of acids produced by the carboxylate platform can be varied by altering temperature conditions (6, 15), and the divergent patterns of community succession that we observed between the two temperature treatments may explain this phenomenon. Although both fermentations were seeded with a common inoculum and

343 dominated by *Clostridia* early on, they shared very few taxa in common. With time, these
344 differences increased, as *Bacteroidetes* overtook the 40 °C community and the 55 °C community
345 remained *Clostridia*-rich. At the same time, the acid product spectra of the two communities also
346 began to differentiate themselves from one another, resulting in the 40 °C community producing
347 significantly more propionic acid than the 55 °C community, and the 55 °C community
348 producing significantly more butyric acid than the 40 °C community (Table 2). Members of the
349 *Bacteroides* have been identified as xylose degraders that generate propionate as a fermentation
350 product (33), whereas many *Clostridium* spp are known cellulose degraders that produce
351 butyrate and/or hydrogen gas as major fermentation products (33). Thus, it appears that
352 community composition and succession patterns, particularly with respect to temperature and the
353 type of substrate remaining within the reactor system, may play an important role in determining
354 product outcomes between the two fermentation temperatures.

355 Community divergence may also contribute to the differences we observed between the
356 two fermentation temperatures with respect to productivity. Although both temperatures resulted
357 in the conversion of similar amounts of biomass, the 40 °C community was significantly more
358 productive than the 55 °C community (Table 1). The 40 °C community was also substantially
359 more "species" rich (Table 3) than the 55 °C community. A richer and more diverse community
360 is likely to harbor individuals capable of fulfilling a wider array of niches, particularly with
361 respect to substrate utilization (28), and a more complete utilization of substrate within the
362 reactor could lead to increased productivity. Alternatively, the differences observed with respect
363 to productivity may be rooted in the fact that our definition of productivity was limited to
364 carboxylic acids and did not include other potential conversion products, such as hydrogen gas or
365 methane. Given the similar conversion rates observed between the two fermentation

366 temperatures, the differences that resulted between them with respect to productivity, the use of
367 iodoform to inhibit methanogenesis, and the identities of dominant taxa that emerged within each
368 community, we speculate that a substantial portion of the potential product in the 55 °C
369 fermentation may have been lost as hydrogen gas. Further work will be needed to confirm
370 whether or not this is the case and will be a focus of future investigations of the carboxylate
371 platform.
372
373 **Conclusion**
374 This study provides a first look into the "black box" that underlies the MixAlco
375 carboxylate platform for biomass conversion. Although the 16S rRNA surveys described here
376 have provided useful insight regarding the composition of our reactor communities, the effects
377 that temperature have upon them, and the ways in which they change over time, we have only
378 just begun to shed light on the microbiology of this system. Additional characterizations of the
379 functional genes that are active within this system, and its metagenome at large, are needed to
380 better understand the microbiology of the carboxylate platform. It is anticipated that such
381 information will provide tremendous strides forward in the optimization of this platform and
382 enhance efforts to understand and improve other biomass conversion systems.
383

References cited

1. **Abd-Alla, M. H., S. A. Omar, and A. M. Abdel-Wahab.** 1992. The role of cellulose-decomposing fungi in nitrogenase activity of *Azotobacter chroococcum*. Folia Microbiol. **37:**215-218.
2. **Aiello-Mazzarri, C., F. K. Agbogbo, and M. T. Holtzapple.** 2006. Conversion of municipal solid waste to carboxylic acids using a mixed culture of mesophilic microorganisms. Bioresour. Technol. **97:**47-56.
3. **Akasaka, H., T. Izawa, K. Ueki, and A. Ueki.** 2003. Phylogeny of numerically abundant culturable anaerobic bacteria associated with degradation of rice plant residue in Japanese paddy field soil. FEMS Microbiol. Ecol. **43:**149-161.
4. **Bàez-Vàsquez, M. A., and A. L. Demain.** 2008. Ethanol, biomass, and *Clostridia*, p. 49-54. *In* J. D. Wall, C. S. Harwood, and A. Demain (ed.), Bioenergy. ASM Press, Washington, D.C.
5. **Burrell, P. C., C. O'Sullivan, H. Song, W. P. Clarke, and L. L. Blackall.** 2004. Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. Appl. Environ. Microbiol. **70:**2414-2419.
6. **Chan, W. N., and M. T. Holtzapple.** 2003. Conversion of municipal solid wastes to carboxylic acids by thermophilic fermentation. Appl. Biochem. Biotechnol. **111:**93-112.
7. **Chang, V. S., M. Nagwani, and M. T. Holtzapple.** 1998. Lime pretreatment of crop residues bagasse and wheat straw. Appl. Biochem. Biotechnol. **74:**135-159.

8. **Chynoweth, D. P., and P. Pullammanappallil.** 1996. Anaerobic digestion of municipal solid wastes, p. 74-114. *In* A. C. Palmisano and M. A. Barlaz (ed.), Microbiology of solid wastes. CRC Press, Boca Raton, FL.

9. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje.** 2009. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Res. **37:**D141-145.

10. **Demain, A. L., M. Newcomb, and J. H. D. Wu.** 2005. Cellulase, *Clostridia*, and ethanol. Microbiol. Mol. Biol. Rev. **69:**124-154.

11. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72:**5069-5072.

12. **Engelbrektson, A., V. Kunin, K. C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, and P. Hugenholtz.** 21 January 2010, posting date. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. doi: 10.1038/ismej.2009.153.

13. **Felsenstein, J.** 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by author. Department of Genome Sciences, University of Washington, Seattle, WA.

14. **Fernandez, A., S. Y. Huang, S. Seston, J. Xing, R. Hickey, C. Criddle, and J. Tiedje.** 1999. How stable is stable? Function versus community composition. Appl. Environ. Microbiol. **65:**3697-3704.

15. **Fu, Z., and M. T. Holtzapple.** 27 August 2009, posting date. Fermentation of sugarcane bagasse and chicken manure to calcium carboxylates under thermophillic conditions. Appl. Biochem. Biotechnol. doi: 10.1007/s12010-009-8748-z.

16. **Goberna, M., H. Insam, and I. H. Franke-Whittle.** 2009. Effect of biowaste sludge maturation on the diversity of thermophilic Bacteria and Archaea in an anaerobic reactor. Appl. Environ. Microbiol. **75:**2566-2572.

17. **Granda, C. B., and M. T. Holtzapple.** 2008. Biorefineries for solvents: The MixAlco process, p. 347-360. *In* J. D. Wall, C. S. Harwood, and A. Demain (ed.), Bioenergy. ASM Press, Washington, D.C.

18. **Granda, C. B., M. T. Holtzapple, G. Luce, K. Searcy, and D. L. Mamrosh.** 2009. Carboxylate platform: The MixAlco process part 2: Process economics. Appl. Biochem. Biotechnol. **156:**537-554.

19. **Hammer, O., D. A. T. Harper, and P. D. Ryan.** 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron **4:**9 pp.

20. **Hardman, J. K., and T. C. Stadtman.** 1960. Metabolism of omega-amino acids: 2. Fermentation of delta-aminovaleric acid by *Clostridium aminovalericum* n sp. J. Bacteriol. **79:**549-552.

21. **Hollister, E. B., A. S. Engledow, A. M. Hammett, T. L. Provin, H. H. Wilkinson, and T. J. Gentry.** 4 February 2010, posting date. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J. doi: 10.1038/ismej.2010.3.

22. **Holtzapple, M. T., R. R. Davison, M. K. Ross, S. Aldrett-Lee, M. Nagwani, C. M. Lee, C. Lee, S. Adelson, W. Kaar, D. Gaskin, H. Shirage, N. S. Chang, V. S. Chang,**

and M. E. Loescher. 1999. Biomass conversion to mixed alcohol fuels using the MixAlco process. Appl. Biochem. Biotechnol. **77-9:**609-631.

23. **Juarez, B., M. V. Martinez-Toledo, and J. Gonzalez-Lopez.** 2005. Growth of *Azotobacter chroococcum* in chemically defined media containing p-hydroxybenzoic acid and protocatechuic acid. Chemosphere **59:**1361-1365.

24. **Kato, S., S. Haruta, Z. J. Cui, M. Ishii, and Y. Igarashi.** 2004. Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. FEMS Microbiol. Ecol. **51:**133-142.

25. **Klocke, M., P. Mähnert, K. Mundt, K. Souidi, and B. Linke.** 2007. Microbial community analysis of a biogas-producing completely stirred tank reactor fed continuously with fodder beet silage as mono-substrate. Syst. Appl. Microbiol. **30:**139-151.

26. **Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz.** 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. **12:**118-123.

27. **Lee, Z. M.-P., C. Bussema, III, and T. M. Schmidt.** 2009. rrnDB: Documenting the number of rRNA and tRNA genes in Bacteria and Archaea. Nucleic Acids Res. **37:**D489-493.

28. **Levén, L., A. R. B. Eriksson, and A. Schnürer.** 2007. Effect of process temperature on bacterial and archaeal communities in two methanogenic bioreactors treating organic household waste. FEMS Microbiol. Ecol. **59:**683-693.

29. **Lin, Y., and S. Tanaka.** 2006. Ethanol fermentation from biomass resources: current state and prospects. Appl. Microbiol. Biotechnol. **69:**627-642.

30. **Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius.** 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66:**506-577.

31. **Madden, R. H.** 1983. Isolation and characterization of *Clostridium stercorarium* sp. nov., cellulolytic thermophile. Int. J. Syst. Bacteriol. **33:**837-840.

32. **Perlack, R. D., L. L. Wright, A. F. Turhollow, R. L. Graham, B. J. Stokes, and D. C. Erbach.** 2005. Biomass as feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply, DOE/GO-102005-2135. Oak Ridge National Laboratory, Oak Ridge, TN. (http://feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf).

33. **Ren, N., D. Xing, B. E. Rittmann, L. Zhao, T. Xie, and X. Zhao.** 2007. Microbial community structure of ethanol type fermentation in bio-hydrogen production. Environ. Microbiol. **9:**1112-1125.

34. **Rooney, W. L., J. Blumenthal, B. Bean, and J. E. Mullet.** 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod. Biorefin. **1:**147-157.

35. **Rubin, E. M.** 2008. Genomics of cellulosic biofuels. Nature **454:**841-845.

36. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75:**7537-7541.

495 37. **Shin, H.-S., J.-H. Youn, and S.-H. Kim.** 2004. Hydrogen production from food waste in
496 anaerobic mesophilic and thermophilic acidogenesis. Int. J. Hydrogen Energy **29:**1355-
497 1363.
498 38. **Shiratori, H., H. Ikeno, S. Ayame, N. Kataoka, A. Miya, K. Hosono, T. Beppu, and**
499 **K. Ueda.** 2006. Isolation and characterization of a new *Clostridium* sp. that performs
500 effective cellulosic waste digestion in a thermophilic methanogenic bioreactor. Appl.
501 Environ. Microbiol. **72:**3702-3709.
502 39. **Thanakoses, P., N. A. A. Mostafa, and M. T. Holtzapple.** 2003. Conversion of
503 sugarcane bagasse to carboxylic acids using a mixed culture of mesophilic
504 microorganisms. Appl. Biochem. Biotechnol. **105:**523-546.
505 40. **von Wintzingerode, F., U. B. Göbel, and E. Stackebrandt.** 1997. Determination of
506 microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis.
507 FEMS Microbiol. Rev. **21:**213-229.
508 41. **Yanling, H., D. Youfang, and L. Yanquan.** 1991. Two cellulolytic *Clostridium* species:
509 *Clostridium cellulosi* sp. nov. and *Clostridium cellulofermentans* sp. nov. Int. J. Syst.
510 Bacteriol. **41:**306-309.
511 42. **Yue, J. C., and M. K. Clayton.** 2005. A similarity measure based on species
512 proportions. Commun. Stat. Theory Methods **34:**2123-2131.

Titles and legends to figures

Figure 1. Nonmetric multidimensional scaling of reactor communities, based upon their OTU composition, shows distinct separation of the bioreactor communities with respect to fermentation temperature (40 vs. 55 °C) and time. Early, mid, and late refer to incubation lengths of 6, 16, and 30 days, respectively.

Figure 2. Distribution of bacterial classes (foreground) versus bioreactor acid products (background) over 30 days' incubation at A) 40 °C and B) 55 °C.

Figure 3. Neighbor-joining phylogenetic tree of the five most abundant OTUs found within each bioreactor community following 16 days' incubation at 40 °C (blue) and 55 °C (red) and their three nearest neighbor isolates from the Greengenes chimera-checked database. Bootstrap values (out of a possible 100) are provided at each branch point, and the relative abundance of each cluster (i.e., OTU) within its respective community is provided in parentheses.

Figure 4. Neighbor-joining phylogenetic tree of the 10 OTUs that were found to share strong correlations with reactor productivity rates and their nearest neighbor isolates from the Greengenes chimera-checked database. Bootstrap values (out of a possible 100) are provided at each branch point. The symbol adjacent to each OTU entry indicates its relationship (triangle = positive correlation, square = negative correlation) with reactor productivity, and the color of the text indicated the community with which each OTU was affiliated (blue = 40°C, red = 55°C).

Table 1. Reactor performance metrics following 30 days' incubation. Values represent the mean of three replicates ± SE. Different letters, within a column, indicate a statistically significant difference at $p < 0.05$.

Treatment	Conversion [1]	Selectivity [2]	Yield [3]	Productivity (g acid L^{-1} d^{-1})
40 °C incubation	0.231 ± 0.023 [a]	0.371 ± 0.024 [a]	0.085 ± 0.004 [a]	0.230 ± 0.012 [a]
55 °C incubation	0.259 ± 0.032 [a]	0.232 ± 0.027 [b]	0.059 ± 0.001 [b]	0.159 ± 0.001 [b]

[1] *Conversion* represents the ratio of volatile solids digested to the biomass that was originally loaded into the reactor.

[2] *Selectivity* represents the proportion of digested material that resulted in carboxylic acid production.

[3] *Yield* represents the ratio of total carboxylic acids produced to the biomass that was originally loaded into the reactor.

16 **Table 2.** Distribution of fermentation products under contrasting reaction incubation temperatures following 30 days' incubation.

17 Values represent the mean of three replicates ± SE. Different letters, within a column, indicate a statistically significant difference at p

18 < 0.05 as determined by Student's t-test. ND refers to acid products that were not detected.

19

Treatment	Relative abundance of acid products (%)				
	Acetic (C_2)	Propionic (C_3)	Butyric (C_4)	Valeric (C_5)	Caproic (C_6)
40 °C incubation	63.40 ± 2.62 [a]	14.43 ± 1.14 [a]	16.46 ± 1.82 [a]	2.72 ± 0.44 [a]	2.98 ± 0.88 [a]
55 °C incubation	64.47 ± 1.86 [a]	ND [b]	34.23 ± 1.86 [b]	1.26 ± 0.02 [a]	ND [a]

Table 3. Summary of community diversity attributes based upon OTUs and their relative abundances. OTUs were defined at a 97% similarity level. Early refers to samples taken after 6 days' incubation, mid refers to samples taken after 16 days' incubation, and late refers to samples taken after 30 days' incubation.

Reactor community	Community characteristics			
	Sequence library size	Number of OTUs	Shannon (H')	Chao I richness estimate
Overall	160,957	2955	---	---
40 °C early	46,359	1382	3.61	2878
40 °C mid	41,732	978	3.59	1882
40 °C late	17,033	526	3.46	1162
55 °C early	22,406	387	2.52	847
55 °C mid	15,344	140	2.34	271
55 °C late	18,083	479	3.21	888

Table 4. Overlap of OTUs among the reactor communities, as represented by θ_{YC}, the Yue-Clayton estimator of similarity. θ_{YC} is scored on a scale of 0 to 1, with a score of 0 representing complete dissimilarity and a score of 1 representing identity. Clusters of shading have been placed within the table to highlight the changes within each temperature treatment over time.

	Reactor community					
	40 °C early	40 °C mid	40 °C late	55 °C early	55 °C mid	55 °C late
40 °C early	---					
40 °C mid	0.4236	---				
40 °C late	0.0615	0.5885	---			
55 °C early	0.0001	< 0.0001	0.0412	---		
55 °C mid	0.0001	0.0003	0.0488	0.7555	---	
55 °C late	0.0020	0.0044	0.0346	0.2983	0.4071	---

Table 5. Characteristics of OTUs correlated with productivity rates in the MixAlco system in the A) 40 °C and B) 55 °C reactor communities. The symbol following the r^2 value indicates whether the correlation between the relative abundance of that particular OTU with reactor productivity shared a positive (+) or negative (-) relationship. Accession numbers refer to those belonging to the closest GenBank matches.

OTU ID	Correlation (r^2)	Closest GenBank match	Accession number	Similarity (%)	Potential function	Reference
A)						
2	0.992 (+)	Uncultured *Clostridiales*	GU214156	98	cellulose degradation	
5	0.999 (+)	*Clostridiaceae* str 80Wc	AB078860	98	cellobiose degradation	(3)
28	0.987 (+)	*Clostridium aminovalericum*	NR_029245	96	cellobiose degradation	(20)
13	0.825 (-)	*Actinomycetaceae* bacterium	FJ542912	95	Uncertain	
27	0.720 (-)	*Azotobacter chroococcum*	EF634040	100	xylose degradation; possible degradation of phenolic compounds	(1, 23)
B)						
1	0.996 (+)	Uncultured bacterium	AB221356	99	cellulose degradation	(38)
14	0.559 (+)	*Clostridium cellulosi*	FJ465164	100	cellulose degradation	(41)
16	0.871 (-)	*Clostridium stercorarium*	L09176	96	cellulose and xylose degradation	(31)
70	0.739 (-)	*Symbiobacterium* sp KA13	AB455239	95	uncertain	
174	0.964 (-)	*Bacillus* sp S13	AF403022	99	likely xylose degradation	

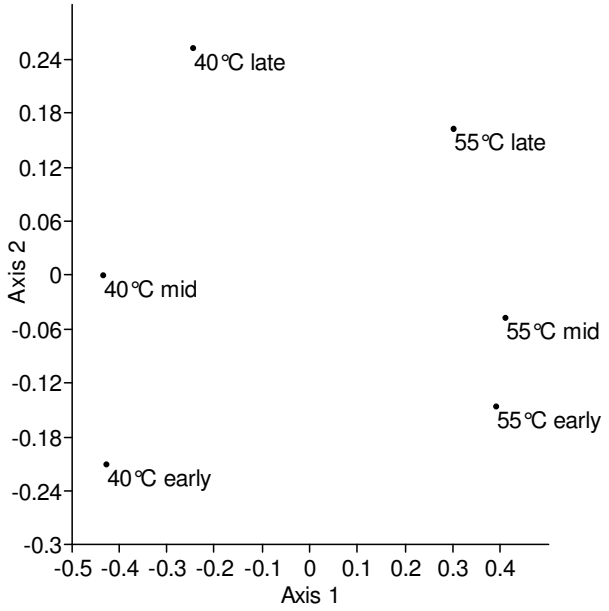

Figure 1. Nonmetric multidimensional scaling of reactor communities, based upon their OTU composition, shows distinct separation of the bioreactor communities with respect to fermentation temperature (40 vs. 55 °C) and time. Early, mid, and late refer to incubation lengths of 6, 16, and 30 days, respectively.

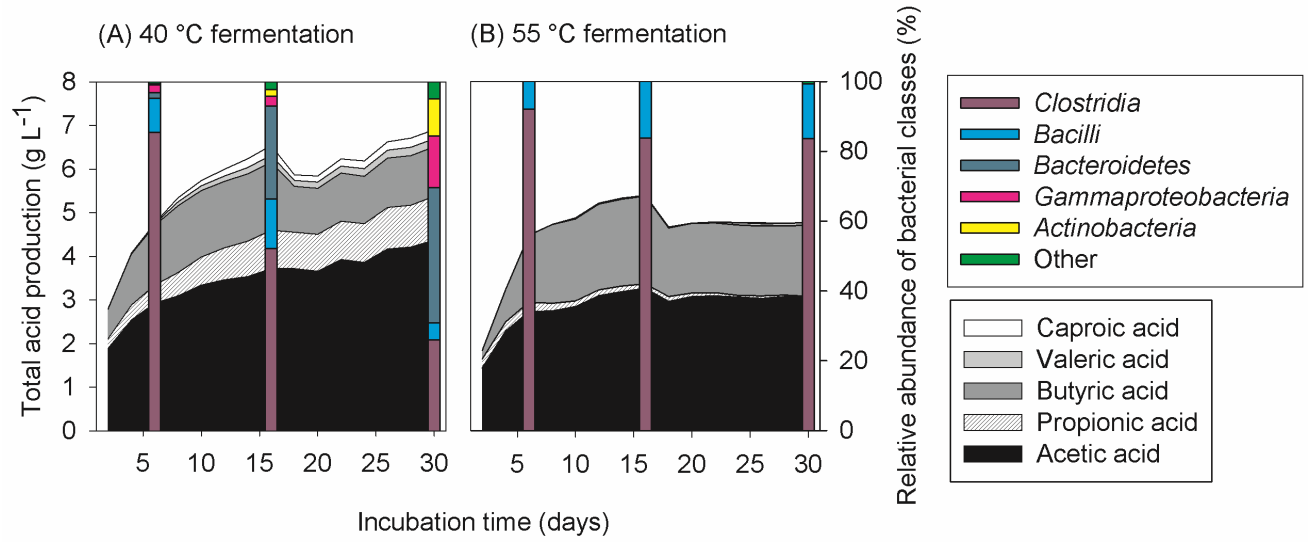

Figure 2. Distribution of bacterial classes (foreground) versus bioreactor acid products (background) over 30 days' incubation at A) 40 °C and B) 55 °C.

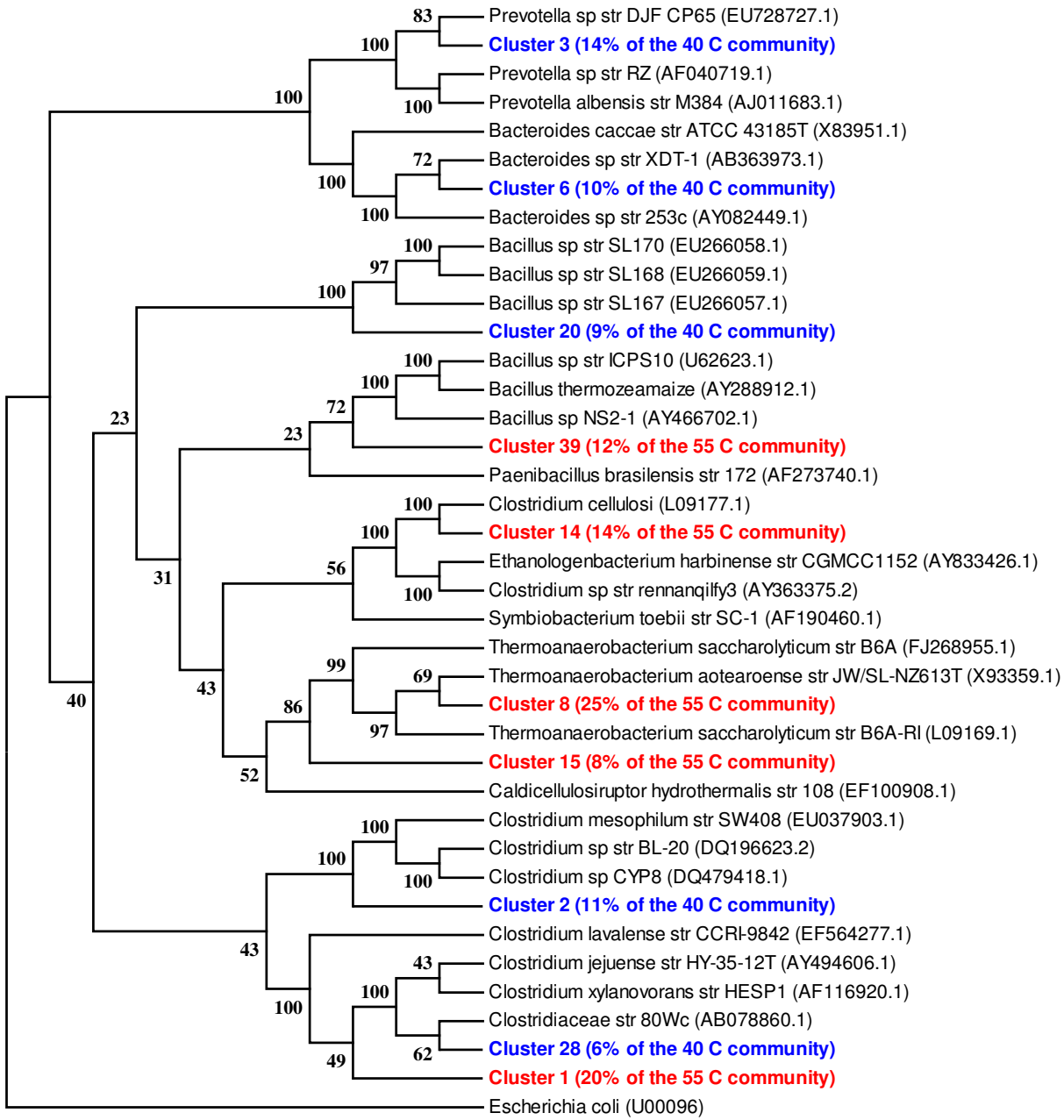

Figure 3. Neighbor-joining phylogenetic tree of the five most abundant OTUs found within each bioreactor community following 16 days' incubation at 40 °C (blue) and 55 °C (red) and their three nearest neighbor isolates from the Greengenes chimera-checked database. Bootstrap values (out of a possible 100) are provided at each branch point, and the relative abundance of each cluster (i.e., OTU) within its respective community is provided in parentheses.

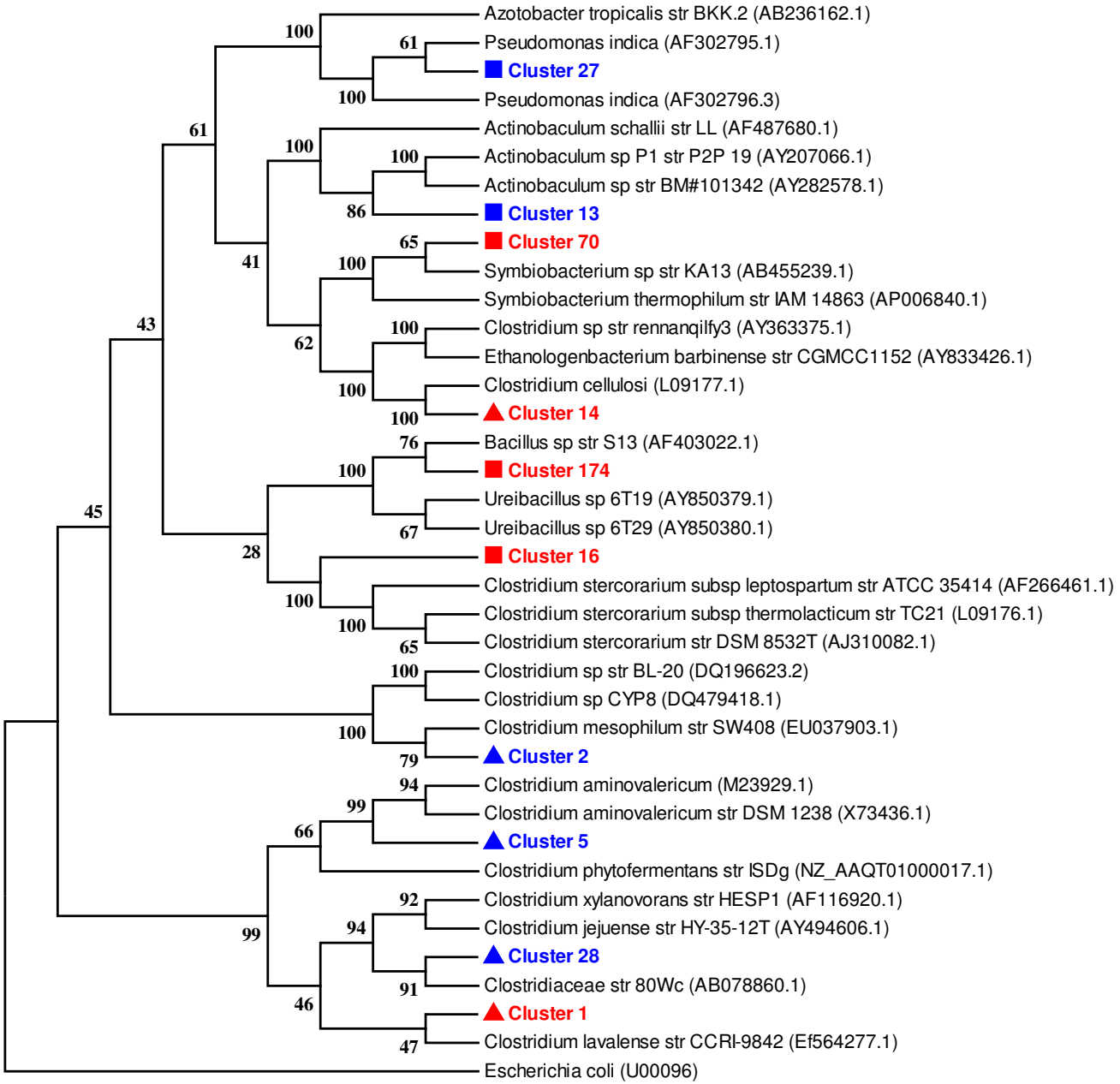

Figure 4. Neighbor-joining phylogenetic tree of the 10 OTUs that were found to share strong correlations with reactor productivity rates and their nearest neighbor isolates from the Greengenes chimera-checked database. Bootstrap values (out of a possible 100) are provided at each branch point. The symbol adjacent to each OTU entry indicates its relationship (triangle = positive correlation, square = negative correlation) with reactor productivity, and the color of the text indicated the community with which each OTU was affiliated (blue = 40°C, red = 55°C).