Supporting information for

**Metabolic Segregation and Characteristic Gene Clusters of the Key Functional Populations in Anaerobic Digestion Consortia**

Yubo Wang1,2, Ruoqun Zhang1, Chunxiao Wang2, Weifu Yan2, Tong Zhang2, Feng Ju1

Author Affiliation:

1. Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang Province, China

2. Environmental Microbiome Engineering and Biotechnology Laboratory, The University of Hong Kong, Pokfulam, Hong Kong

**SUPPORTING INFORMATION**

***Seed sludge used***

AD is not commonly adopted in wastewater treatment plants in mainland China, and no anaerobic digestion sludge is available as seed at the time, so we started with surplus sludge for the enrichment of anaerobic digestion microbiome. The inoculum used in this enrichment experiment was surplus sludge sampled from Chengxi wastewater treatment plant (Hangzhou, China).

***Physiochemical characterization of the biogas contents and the volatile fatty acids***

Contents of H2, CO2 and CH4 were measured by a gas chromatograph (GC) (Agilent 7890B, California, USA) equipped with a thermal conductivity detector, and a 2 m × 2 mm (inside diameter) stainless steel column packed with Porapak N (80–100 mesh). Injector, detector and column temperatures were kept at 57 ◦C, 200 ◦C and 60 ◦C, respectively. Nitrogen gas was used as the carrier gas at a flow rate of 30 mL/min. Standard curves established in this study covered H2 concentrations in the range of 100 ppm-20,000 ppm, CO2 content in the range of 20%-80%, and CH4 content in the range of 20%-80%.

Concentrations of volatile fatty acids (VFAs), i.e., formate, acetate, n-butyrate, iso-butyrate, and propionate were measured by the same GC, with a flame ionization detector and a 10 m × 0.53 mm HP- FFAP fused-silica capillary column [1]. Initial temperature of the column was 70 oC for 3 min, followed with a ramp of 5 oC/min to 80 oC for 2 min, a second ramp of 10 oC/min was followed to achieve a final column temperature of 120 oC for 1 min. Temperatures of the injector and detector were 200 and 250 oC, respectively. Nitrogen was used as the carrier gas at a flow rate of 25 mL/min, and the flow rates of air and hydrogen gas were 400 ml/min and 40 ml/min, respectively. VFAs analysed in this study included acetate, propionate and n-butyrate and iso-butyrate. Before being processed for VFA measurement, 1.5 mL mixed liquor sampled from the AD reactor was filtered through a 0.45 μm membrane filter.

***Quality filtration of the meta-omics sequencing data***

Quality filtration of the metagenome sequencing data is as follows: [Cutadapt](https://github.com/marcelm/cutadapt/) (v1.17) finds and removes the 3’ sequencing adapter from the high-throughput sequencing reads [2]. A sliding window method supported by fastp was applied to drop the low-quality bases of each read, to be specific, the window (window size: 5 bp) slides from 5’ to 3’ of the sequence, and if one window was met with mean quality <20, drop the bases in the window and the right part, and then stop. After the trimming and pruning as introduced above, reads shorter than 50 bp and reads with ambiguous bases (N) would be discarded. Percentage of clean reads that passed this quality control step is summarized in Appendix file1.

***Assembly strategy of the metagenome datasets***

10 sets of single assembly were conducted on the clean DNA paired end reads generated from the 10 metagenome datasets, respectively. 3 sets of co-assembly were conducted respectively on: 1) the 3 metagenome datasets of Consortia-G, corresponding to biomass samples collected from Consortia-G on day-85, day-134 and day-219, respectively; 2) the 3 metagenome datasets of Consortia-P, corresponding to biomass samples collected from Consortia-P on day-85, day-134 and day-219, respectively; and 3) the 3 metagenome datasets of Consortia-B, corresponding to biomass samples collected from Consortia-B on day-85, day-134 and day-219, respectively. Quality of the assembly was evaluated by QUAST[3], which could be found in Appendix file1.

***Thermodynamic evaluation***

Thermodynamic calculation conducted in this study was mainly on the syntrophic oxidation of propionate and butyrate, following the approaches as introduced by Robbert et al [4]. And such thermodynamic evaluation is to evaluate how the H2 dynamics may affect the propionate and butyrate oxidation, assuming H2 is the electron carrier in the syntrophic metabolism.

***Metabolizing role of Eremiobacterota and Thermotoga in the enriched AD consortia***

Another four MAGs identified with high transcription activity include: the three co-occurring *Eremiobacterota* populations ER1, ER2 and ER4, and the one *Thermotoga* population TM1. Take ER1 for example, its relative abundance in Consortia-P was only 0.54%, yet when butyrate was provided as the primary substrate to Consortia-P (datapoint pb5), ER1 accounted for 4.3% of the mRNA gene transcripts; and when fumarate was provided as the primary substrate to Conosrtia-P (datapoint pfu3), ER1 accounted for 2.5% of the mRNA gene transcripts. However, unlike that of FI21,22, FI3,4,7 and DES18-25, these 4 MAGs did not demonstrate consistent pattern of transcription activity towards the stimuli of a certain substrate. As summarized in Figure 4, highly expressed in these four MAGs were genes encoding the protease, amylase, clostripain and the murein-degrading enzymes (genes in grid M10 in Figure 4). Preference for peptides and amino acids as nutrient sources by *Eremiobacterota* has also been suggested in previous pangenome analysis [5]. This suggests that these three co-occurring *Eremiobacterota* populations ER1,2,4 and the one *Mesotoga* population TM1 might be scavengers in the system by utilizing extracellular substances, e.g., protein, starch, and even cell debris. The putative mobility of the *Eremiobacterota* populations ER1,2,4 was indicated by the high transcription activity of genes encoding the assembly of the locomotion organelle flagella (genes in grid M09 in Figure 4). Being an as-yet-uncultured bacterial clade, the first report on *Eremiobacterota* was in cold desert soils of Antarctica [6], and this is the first time that two quite novel (red-value <0.5, annotatable only at the phylum level) *Eremiobacterota* MAGs have been recovered from an anaerobic digestion consortia.

**List of supporting figures and supporting tables**

Figure S1. Stepwise increment of the organic loading rate and the biogas yield during the enrichment of the three consortia over a course of 219 days

Figure S2. Schematic illustration of the stimuli experiment. 1The combined substrate denoted as ‘G: B: P=1:1:1’ in the 4th batch is a combination of glucose: butyrate: propionate as co-substrates in a ratio of 1:1:1 (COD: COD: COD).

Figure S3. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-B, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when crotonate was provided as the primary substrate, d) when glucose: butyrate: propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and e) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.

Figure S4. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-P, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when fumarate was provided as the primary substrate, d) when glucose: butyrate: propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and e) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.

Figure S5. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-G, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when glucose: butyrate: propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and d) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.

Figure S6. Sludge VSS content measured of the three consortia over the course of the 219 days enrichment

Figure S7. a) Relative abundance of the 15 genera in the seed sludge (blue), Consortia-G (grey), Consortia-P (pink) and Consortia-B (magenta). The circular dots represent the relative abundance of these top 15 genera in total, while the cross dots represent relative abundance of the two methanogen archaeal genera. b) A zoomed-in look on the relative abundance of *Methanobacterium* and *Methanosaeta* in the three enriched consortia.

Figure S8. Hydrogen dynamics over the degradation of varying substrates in a) Consortia-G, b) Consortia-P and c) Consortia-B. The varying substrates provided in the serial batch experiments are butyrate, propionate, glucose, combination of glucose, butyrate and propionate in a ratio of 1:1:1 (COD: COD: COD, referred to as G: B: P=1:1:1 in the figure), and crotonate (in Consortia-B) or fumarate (in Consortia-P), respectively.

Figure S9. Gibbs free energy change as a function of H2 concentration of the hydrogenotrophic methanogenesis, anaerobic acetogenic oxidation of butyrate and of propionate. This calculation assumes that H2 is the only electron carrier in between the syntrophic acetogen bacteria and the hydrogenotrophic methanogens; concentration of propionate and butyrate are at 10 µM, and acetate concentration at 1 µM; and the CO2 and CH4 contents in the biogas were ~30% and ~70%, respectively.

Figure S10. Methanogenic activity of the three enriched consortia over the stimuli of varying substrates, i.e., glucose, butyrate, propionate, crotonate, fumarate and combined substrate as glucose: butyrate: propionate = 1:1:1 (COD: COD: COD).

Table S1. Physiochemical parameters detected in the three consortia at the time points when sludge samples were collected for meta-RNA sequencing. Note: 1. For better visualization, S1, S2, S3, ScH2 and Sc were used in Figure 4 to denote the corresponding substrate applied as the stimuli carbon source in each of the three consortia; 2. Sampling time here indicates the degradation time after the addition of each substrate, and the degradation profile of each stimuli-substrate over the course of a batch experiment could be found in Figure S3 -S5.

Table S2. Estimation on the VSS equivalent of the putative fermenting bacteria (FI21,22), the putative SBOBs (FI3,4,7) and the putative SPOBs (DES18-25) identified in the three enriched consortia.

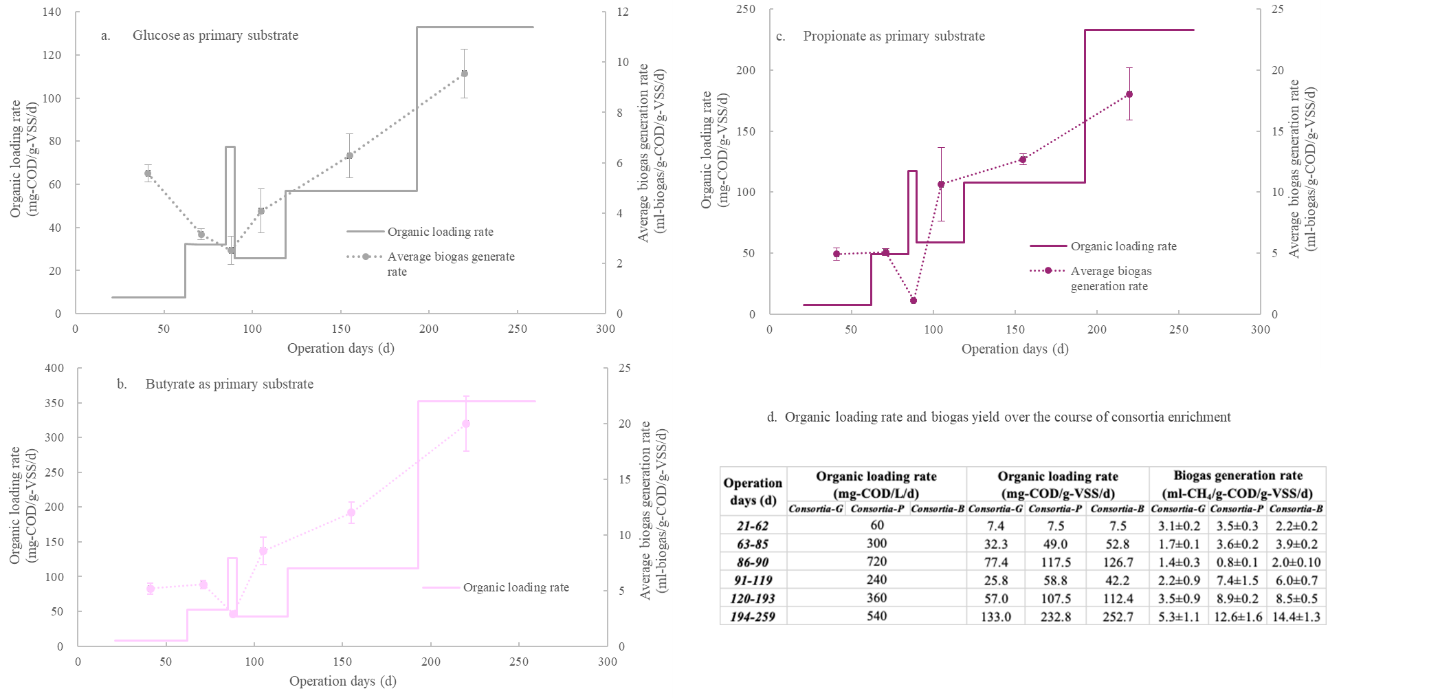


Figure S1. Stepwise increment of the organic loading rate and the biogas yield during the enrichment of the three consortia over a course of 219 days



Figure S2. Schematic illustration of the stimuli experiment. 1The combined substrate denoted as ‘G:B:P=1:1:1’ in the 4th batch is a combination of glucose:butyrate:propionate as co-substrates in a ratio of 1:1:1 (COD:COD:COD).



Figure S3. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-B, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when crotonate was provided as the primary substrate, d) when glucose:butyrate:propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and e) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.



Figure S4. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-P, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when fumarate was provided as the primary substrate, d) when glucose:butyrate:propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and e) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.



Figure S5. Profile of short-chain fatty acids (SCFAs) and H2 dynamics in Consortia-G, a) when butyrate was supplied as the primary substrate, b) when propionate was supplied as the primary substrate, c) when glucose:butyrate:propionate = 1:1:1 (COD:COD:COD) was provided as co-substrates, and d) when glucose was provided as the primary substrate. Arrows in purple indicate the time points when sludge samples were retrieved for metatranscriptome analysis.

Figure S6. Sludge VSS content measured of the three consortia over the course of the 219 days enrichment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Figure S7. a) Relative abundance of the 15 genera in the seed sludge (blue), Consortia-G (grey), Consortia-P (pink) and Consortia-B (magenta). The circular dots represent the relative abundance of these top 15 genera in total, while the cross dots represent relative abundance of the two methanogen archaeal genera. b) A zoomed-in look on the relative abundance of *Methanobacterium* and *Methanosaeta* in the three enriched consortia. |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | |  | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Figure S8. Hydrogen dynamics over the degradation of varying substrates in a) Consortia-G, b) Consortia-P and c) Consortia-B. The varying substrates provided in the serial batch experiments are butyrate (coloured in magenta in the figure), propionate (coloured in pink in the figure), glucose (coloured in green in the figure), combination of glucose, butyrate and propionate in a ratio of 1:1:1 (COD: COD: COD, referred to as G:B:P=1:1:1 in the figure), and crotonate (in Consortia-B) or fumarate (in Consortia-P), respectively. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Figure S9. Gibbs free energy change as a function of H2 concentration of the hydrogenotrophic methanogenesis, anaerobic acetogenic oxidation of butyrate and of propionate. This calculation assumes that H2 is the only electron carrier in between the syntrophic acetogen bacteria and the hydrogenotrophic methanogens; concentration of propionate and butyrate are at 10 µM, and acetate concentration at 1 µM; and the CO2 and CH4 contents in the biogas were 30% and 70%, respectively.



Figure S10. Transcription dynamics and relative dominance of *Methanobacteriaceae* EA7 and *Methanothrix* HA1,2,3 in the hydrogenotrophic- and acetoclastic- methanogenesis metabolism. *cdhA*: Acetyl-CoA decarbonylase subunit alpha, *fdhA*: formate dehydrogenase subunit alpha, *fdhC*: formate transporter gene, and *mcrA:* the methyl coenzyme M reductase subunit alpha. Full names of genes denoted with abbreviations could be found in the Appendix file 1.

Table S1. Physiochemical parameters detected in the three consortia at the time points when sludge samples were collected for meta-RNA sequencing. Note: 1. For better visualization, S1, S2, S3, ScH2 and Sc were used in Figure 4 to denote the corresponding substrate applied as the stimuli carbon source in each of the three consortia, e.g., S1 indicates the use of butyrate as substrate; 2. Sampling time here indicates the digestion time after the addition of each substrate, and the degradation profile of each stimuli-substrate over the course of a batch experiment could be found in Figure S3 -S5.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Consortium** | **Substrate**  **ID1** | **Substrate** | **Sample**  **ID** | **Sampling**  **time2  (h)** | **pH** | **H2 (ppm)** | **CH4 (%)** | **Concentration (mg/l)** | | | | | |
| **n-HBut** | **initial n-HBut** | **iso-But** | **HPr** | **initial HPr** | **HAc** |
| **Consortia-G** |  |  | g0 | na | 7.24 | 0 | 40 | 0 | na | 0 | 0 | na | 0 |
| S1 | Butyrate | gb5 | 23 | 6.98 | 36 | 37 | 467 | 1119 | 90 |  |  |  |
| S2 | Propionate | gp5 | 24 | 7.15 | 28 | 45 | 0 | 0 | 0 | 1004 | 3002 | 0 |
| S3 | Glucose | gg2 | 4 | 7.04 | 4030 | 14 | 0 | 0 | 0 | 178 | 0 | 43 |
|  | Glucose | gg5 | 24 | 7.12 | 56 | 44 | 0 | 0 | 0 | 240 | 0 | 43 |
|  | G:B:P=1:1:1 | gc6 | 24 | 6.87 | 35 | 57 | 106 | 308 | 0 | 518 | 767 | 38 |
| **Consortia-P** |  |  | p0 | na | 7.33 | 0 | 36 | 0 | na | 0 | 0 | na | 0 |
| S1 | Butyrate | pb5 | 24 | 7.21 | 47 | 54 | 491 | 1455 | 113 | 0 | 0 | 0 |
| S2 | Propionate | pp4 | 10 | 6.9 | 25 | 36 | 0 | 0 | 0 | 1547 | 2223 | 0 |
| S3 | Fumarate | pfu3 | 13 | 6.98 | 5682 | 8 | 0 | 0 | 53 | 43 | 0 | 0 |
| ScH2 | G:B:P=1:1:1 | pc4 | 10 | 6.71 | 34936 | 15 | 289 | 326 | 0 | 616 | 757 | 133 |
| Sc | G:B:P=1:1:1 | pc7 | 49 | 7.08 | 51 | 65 | 28 | 326 | 0 | 297 | 759 | 199 |
|  | Glucose | pg5 | 23 | 6.82 | 51506 | 33 | 16 | 0 | 56 | 52 | 0 | 0 |
| **Consortia-B** |  |  | b0 | na | 7.31 | 0 | 45 | 0 | na | 0 | 0 | na | 0 |
| S1 | Butyrate | bb4 | 10 | 7 | 35 | 42 | 646 | 1269 | 136 | 0 | 0 | 0 |
| S2 | Propionate | bp5 | 36 | 7.12 | 20 | 35 | 0 | 0 | 0 | 1315 | 2779 | 0 |
| S3 | Crotonate | bcr7 | 49 | 6.74 | 101 | 40 | 144 | 0 | 62 | 46 | 0 | 407 |
| ScH2 | G:B:P=1:1:1 | bc4 | 10 | 6.78 | 30000 | 10 | 439 | 469 | 0 | 590 | 659 | 101 |
| Sc | G:B:P=1:1:1 | bc7 | 49 | 7.02 | 51 | 67 | 34 | 456 | 0 | 398 | 640 | 91 |
|  | Glucose | bg5 | 23 | 6.91 | 35798 | 27 | 43 | 0 | 58 | 53 | 0 | 0 |

Table S2. Estimation on the VSS equivalent of the putative fermenting bacteria (FI21,22), the putative SBOBs (FI3,4,7) and the putative SPOBs (DES18-25) identified in the three enriched consortia

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Enriched consortia** | **VSS of the  enriched sludge  (mg)** | **16S rRNA based relative abundance** | | | **VSS equivalent (mg)** | | |
| **FI3,4,7** | **DES18-25** | **FI21,22** | **FI3,4,7** | **DES18-25** | **FI21,22** |
| Consortia-B | 711 | 7.34% | 0.67% | 0.01% | 9817 | 896 | 13 |
| Consortia-P | 249 | 1.75% | 7.30% | 0.01% | 4364 | 18204 | 25 |
| Consortia-G | 134 | 0.08% | 1.33% | 1.57% | 569 | 9455 | 11162 |

**References**

1. Fang, H., C. Li, and T. Zhang, *Acidophilic biohydrogen production from rice slurry.* International Journal of Hydrogen Energy, 2006. **31**(6): p. 683-692.

2. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, 2011: p. 10-12.

3. Mikheenko, A., et al., *Versatile genome assembly evaluation with QUAST-LG.* Bioinformatics, 2018. **34**(13): p. i142-i150.

4. Kleerebezem, R. and M.C.M. Van Loosdrecht, *A Generalized Method for Thermodynamic State Analysis of Environmental Systems.* Critical Reviews in Environmental Science and Technology, 2010. **40**(1): p. 1-54.

5. Ji, M., et al., *Candidatus Eremiobacterota, a metabolically and phylogenetically diverse terrestrial phylum with acid-tolerant adaptations.* ISME J, 2021. **15**(9): p. 2692-2707.

6. Ji, M., et al., *Atmospheric trace gases support primary production in Antarctic desert surface soil.* Nature, 2017. **552**(7685): p. 400-403.