

Lawrence Berkeley National Laboratory

LBL Publications

Title

Genome-resolved correlation mapping links microbial community structure to metabolic interactions driving methane production from wastewater

Permalink

<https://escholarship.org/uc/item/1js3t5j5>

Journal

Nature Communications, 14(1)

ISSN

2041-1723

Authors

Kieft, Brandon
Finke, Niko
McLaughlin, Ryan J
[et al.](#)

Publication Date

2023

DOI

10.1038/s41467-023-40907-5

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Genome-resolved correlation mapping links microbial community structure to metabolic interactions driving methane production from wastewater

Received: 12 May 2023

Accepted: 15 August 2023

Published online: 04 September 2023

 Check for updates

Brandon Kieft¹, Niko Finke¹, Ryan J. McLaughlin^{1,2}, Aditi N. Nallan^{1,2}, Martin Krzywinski³, Sean A. Crowe^{1,4} & Steven J. Hallam^{1,2,5,6,7,8} ✉

Anaerobic digestion of municipal mixed sludge produces methane that can be converted into renewable natural gas. To improve economics of this microbial mediated process, metabolic interactions catalyzing biomass conversion to energy need to be identified. Here, we present a two-year time series associating microbial metabolism and physicochemistry in a full-scale wastewater treatment plant. By creating a co-occurrence network with thousands of time-resolved microbial populations from over 100 samples spanning four operating configurations, known and novel microbial consortia with potential to drive methane production were identified. Interactions between these populations were further resolved in relation to specific process configurations by mapping metagenome assembled genomes and cognate gene expression data onto the network. Prominent interactions included transcriptionally active *Methanolinea* methanogens and syntrophic benzoate oxidizing *Syntrophorhabdus*, as well as a Methanoregulaceae population and putative syntrophic acetate oxidizing bacteria affiliated with Bacteroidetes (Tenuifilaceae) expressing the glycine cleavage bypass of the Wood–Ljungdahl pathway.

Renewable natural gas (RNG), primarily composed of biogenic methane (CH₄) and carbon dioxide is an important non-fossil energy resource useful in the transition to a low carbon future¹. Despite widespread adoption of technologies such as anaerobic digestion (AD) to produce biogenic CH₄, industrial-scale AD converting organic waste (e.g., municipal black and gray water) to RNG tends to experience operational challenges, including (i) variable RNG yields, (ii) lower production efficiencies than theoretical values, and (iii) substantial amounts of solid residues that can be costly to dispose². These

challenges can confound the economics of RNG production and arise in part from a prevailing “black box” paradigm that does not fully consider the microbial communities, also known as microbiomes, driving AD conversion processes³. Discovering design principles that shape the network properties of AD microbiomes offers a new paradigm for optimizing RNG production and waste resource recovery.

Over the past decade, high-throughput sequencing approaches have been used to describe microbial community structure, function and dynamics associated with AD at different operating scales^{4–9},

¹Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. ²Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. ³Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada.

⁴Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. ⁵Genome Science and Technology Program, University of British Columbia, 2329 West Mall, Vancouver, BC V6T 1Z4, Canada. ⁶Bradshaw Research Institute for Minerals and Mining (BRIMM), University of British Columbia, Vancouver, BC V6T1Z4, Canada. ⁷Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.

⁸ECOSCOPE Training Program, University of British Columbia, Vancouver, BC V6T 1Z3, Canada. ✉e-mail: shallam@mail.ubc.ca

including recent efforts to characterize the global microbiome of wastewater ADs¹⁰. Importantly, these studies indicate that the AD milieu supports regionally distinct microbial communities containing heterogeneous constituents engaged in conserved metabolic interactions driving organic waste conversion to CH₄¹¹. In addition to microbial dynamics, considerable evidence also indicates that industrial-scale AD performance is strongly influenced by a variety of process parameters including substrate chemistry, loading rates, retention time, temperature, and metal micronutrient (e.g., Fe, Ni, Mo, Mn) concentration^{12–15}. Taken together, an emerging scientific consensus^{16–18} suggests that a new paradigm for understanding and improving AD systems should involve constructing microbial interaction networks in relation to physical and chemical parameter information under time-resolved conditions at relevant operating scales. This is particularly needed for industrial-scale operations which are difficult to access and study over unit time^{4,8,18–25}.

Here, we present a two-year time series study of the Lulu Island municipal wastewater treatment plant (WWTP) operated by Metro Vancouver in Richmond, British Columbia, Canada. Our study is intended to facilitate further understanding of microbial community structure, function, and dynamics in relation to industrial-scale RNG production. The Lulu Island WWTP uses standard practices and produces >5000 m³ of RNG per day from mixed sludge as a starting material. Replicated mixed sludge samples were collected biweekly from Lulu ADs and archived for DNA and RNA extraction. Small subunit ribosomal RNA (SSU or 16S rRNA) gene amplicons were generated from 116 replicated samples, including two periods of standard AD operation, a period of AD operation with an additional allochthonous waste stream, and a period that used serial (instead of parallel) AD operation. The resulting amplicon sequence variant (ASV) data was used to identify a core set of microorganisms across the time series, construct a co-occurrence network to generate statistically informed hypotheses related to potential syntrophic interactions, and, in combination with process parameter information, identify indicator microorganisms associated with different process configurations and relevant WWTP conditions such as final nitrate levels, volatile solids destroyed, total RNG produced, and RNG methane content. A subset of time series samples representing each process configuration was also selected for metagenomic whole genome shotgun and metatranscriptomic sequencing to produce metagenome-assembled genomes (MAGs) and estimate expression levels based on transcript read mapping. MAGs were associated with cognate ASV nodes in the co-occurrence network. Ultimately, these multi-omic datasets and statistical approaches offered potential mechanistic explanations (through encoded and expressed functions) for metabolic interactions between co-occurring microorganisms in the AD, providing insight into the metabolic network driving RNG production in the Lulu Island WWTP.

Results and discussion

Lulu Island waste resource recovery ecosystem

The Lulu Island WWTP operated by Metro Vancouver in Richmond, British Columbia, Canada (Longitude: –123.14498° or 123° 8' 42" W, Latitude: 49.11491° or 49° 6' 54" N) provides primary and secondary treatment of >30 billion liters of mixed-sourced wastewater from ~200,000 residents each year. Primary treatment includes a series of tanks where wastewater undergoes screening, aeration, mechanical separation, settling, and clarification. This effluent then enters a secondary treatment stream, where it is pumped through a trickling filter, solids-contact tank, secondary clarifier, and disinfection tank before being released into receiving water. A portion of sludge from the primary and secondary treatment process streams is mixed and thickened, then split equally into two mesophilic (38 °C) anaerobic digesters (ADs) manifesting a 30-day retention time. The ADs are typically operated in parallel with identical mixed waste inputs. Methane and other gases from the ADs are scrubbed to renewable

natural gas (RNG) which is either used to generate heat for Lulu Island WWTP operations or sold to FortisBC, the local distributor of natural gas.

Metro Vancouver provided physicochemical parameters and triplicate waste secondary sludge (WSS) samples from solid contact tanks and overflow digestate samples from anaerobic digesters (AD1 and AD2) on a biweekly basis between October 13, 2016 and December 12, 2018. During this time interval, ADs experienced four different process configurations (Standard Operation I, standard operation with chemically enhanced primary treatment sludge (CEPT Operation), Standard Operation II, and Serial Operation; Fig. 1A) which differed both in upstream primary and secondary wastewater treatment steps and in the flow of material into the ADs. Each configuration represented experimental perturbations of municipal operations to determine how different process parameters contribute to RNG production,

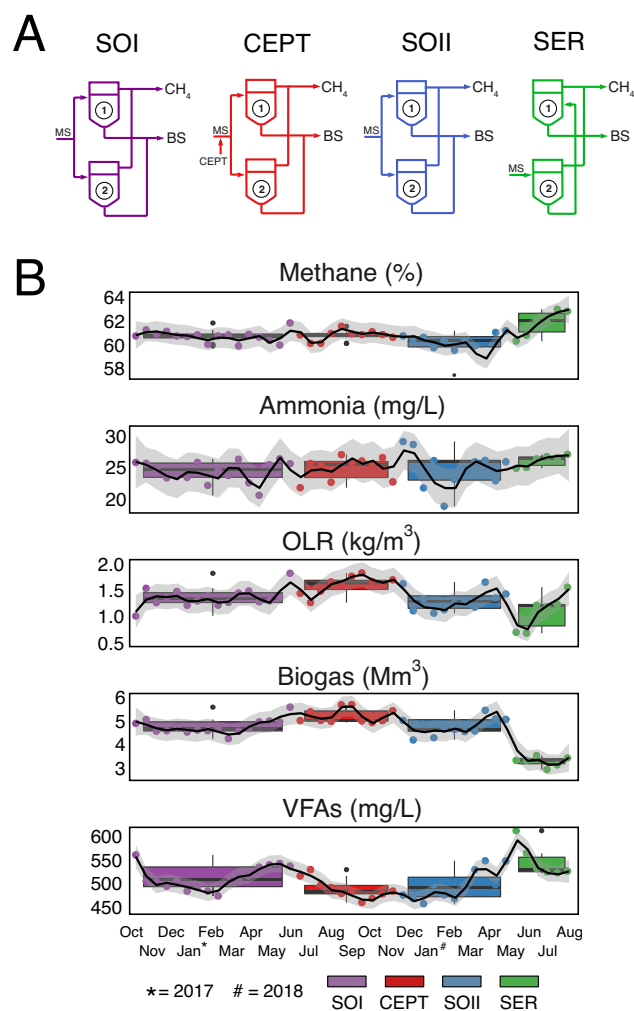


Fig. 1 | Summaries of the operating conditions and physicochemical conditions of the digester across the time series. **A** Wire diagram of the four process configurations across the two-year time series. SOI Standard Operation I, CEPT Chemically Enhanced Primary Treatment Operation, SOII Standard Operation II, SER Serial Operation, MS mixed sludge waste from primary and secondary treatment), CH₄ (biogas), BS biosolids, Labels “1” and “2” indicate the two anaerobic digesters (all samples for this study were taken from AD1). **B** Temporal physicochemical sparklines of AD performance. Methane % volumetric percent of biogas as methane, HRT hydraulic retention time, OLR organic loading rate, Biogas (total volumetric biogas generated), VFAs volatile fatty acids. Box plots of configuration groups represent 75th, 50th (median), and 25th percentiles, with whiskers representing 90th percentiles and outlier position as filled points above whiskers ($n = 16$ for SOI, 10 for CEPT, 11 for SO II, and 6 for SER).

volatile solids conversion, influent, and effluent sludge characteristics (e.g., nutrient and metal composition, chemical oxygen demand) in relation to microbial community structure, function, and dynamics. This study is focused on data collected and analyzed from one of the ADs (AD1), which was sampled continuously throughout the two-year time series.

Physicochemical parameter information

Physicochemical parameter information was collected throughout the time series to evaluate both mixed sludge characteristics entering the AD as well as organic waste conversion processes e.g., solids destruction, VFA concentrations, RNG production (Supplementary Data 1). In pairwise comparisons, physicochemical profiles from the two Standard Operations (I and II) were indistinguishable (Supplementary Data 4). However, samples from the intervening CEPT Operation differed significantly from the two Standard configuration modes based on a sharp increase in organic loading (kg/m^3), which ultimately led to a temporary increase in total biogas production approaching 5739 m^3 (Fig. 1B). The final Serial Operation was the most distinct configuration of the four because of a fundamentally different flow regime, initiated by a rapid shift to in-line (serial) rather than equivalent flow of mixed sludge split between the two ADs (Fig. 1A). This caused the inflow rate (ML/day) into AD1 to increase sharply at the beginning of Serial Operation, along with a concomitant decrease in hydraulic retention time (days). During this interval, digestate characteristics such as volatile fatty acid concentration (mg/L), final ammonia in the effluent (mg/L), and percent CH_4 in RNG all increased. While the CH_4 to carbon dioxide (CO_2) ratio of RNG increased during the Serial configuration (indicating higher quality), the total volume of RNG decreased, suggesting a possible tradeoff in RNG quantity vs. quality which was then explored further from a microbiological perspective.

Microbial community structure

Given the observed impact of process configuration on Lulu Island AD RNG production, we explored the sample archive to identify relationships between physicochemical parameters, such as VFA concentrations and RNG production, and microbial community structure. Genomic DNA extracted from 43 replicated AD1 samples spanning the time series was used to generate amplicons targeting the V4 region of the bacterial and archaeal (prokaryotic) 16S rRNA gene. Resulting data sets had an average of 23,833 quality-filtered 250-bp paired-end reads per sample resolving 928 unique ASVs (Supplementary Data 5). Consistent with physicochemical parameter information, microbial community structure also differed across samples in relation to process configuration (Supplementary Data 6). Interestingly, while the conditions in Standard Operation II returned to the same state as in Standard Operation I (after the intervening CEPT Operation), the microbial community did not return to its previous structure, indicating formation of a new stable state. The majority of unique ASVs (92%) had >80% sequence identity to cognate sequences in the global Microbes in Wastewater Treatment Systems and Anaerobic Digesters (MiDAS) 16S rRNA gene database¹⁰, indicating that a phylogenetically and globally conserved set of microbial lineages are adapted to driving hydrolysis, fermentation and methanogenesis in the AD milieu (Fig. 2).

Despite high recall of ASVs to the MiDAS database, the structure of the Lulu Island AD community differed in several important ways from previously described datasets from mesophilic ADs (Fig. S1)^{23, 26–28}. Although Bacteroidetes, Firmicutes, and Proteobacteria were common and conserved community members, the most abundant taxonomic group in Lulu Island samples was Cloacimonetes (candidate phylum WWE1), with a cumulative relative abundance exceeding 25% in most samples across the time-series (Figs. S1, S2). This phylum is typically found in low abundance in mesophilic ADs treating wastewater, although it has been observed to dominate bioenergy facilities processing crop residues, and some mesophilic WWTPs (typically with

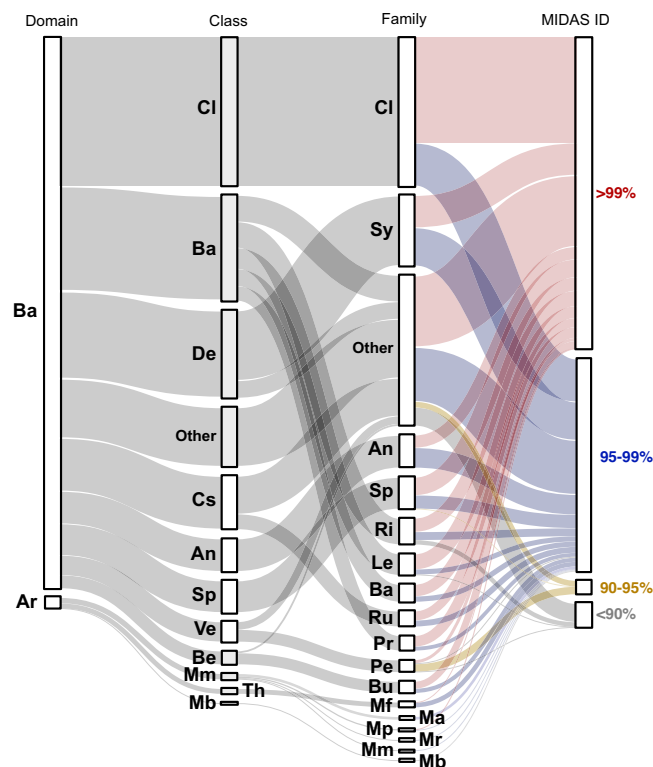


Fig. 2 | A Sankey diagram of microbial community structure shown as relative abundances of family-level taxa grouped into their respective domains and classes. The width of ribbons represents the cumulative relative abundances of all ASVs within each taxonomic lineage. The final column assigns ASVs to categories of percent identity of their V4 16S sequences to the MiDAS database. The two-letter codes on the plot represent taxonomic names. *Domain:* Ba Bacteria, Ar Archaea. *Class:* Cl Cloacimonadia, Ba Bacteroidia, De Deltaproteobacteria, Ot Other, Cs Clostridia, An Anaerolineae, Sp Spirochaetia, Ve Verrucomicrobiae, Be Betaproteobacteria, Mm Methanomicrobia, Th Thermococci, Mb Methanobacteria, *Family:* Cl Cloacimonadaceae, Sy Syntrophaceae, Ot Other, An Anaerolineaceae, Sp Spirochaetaceae, Ri Rikenellaceae, Le Lentimicrobiaceae, Ba Bacteroidetes (vadinHA17), Ru Ruminococcaceae, Pr Prolixibacteraceae, Pe Pedosphaeraceae, Bu Burkholderiaceae, Mf Methanofastidiosaceae, Ma Methanosetaeaceae, Mp Methanospirillaceae, Mr Methanoregulaceae, Mm Methanomicrobiaceae, Mb Methanobacteriaceae.

long retention times or high organic loading rates), where it likely plays a role in amino acid fermentation and syntrophic propionate oxidation^{8,29}. In addition to Cloacimonetes, ASVs associated with the candidate phylum Marinimicrobia were also relatively abundant, reaching up to 5.4% in some samples. Although prevalent and active in marine ecosystems under low oxygen conditions^{30–32}, Marinimicrobia are often observed in ADs where they may play a role in hydrogen production and nitrogen mineralization^{33,34}. Additional candidate groups, including Atribacteria, Kiritimatiellaeota, and Hydrogenedentes, were also identified in Lulu Island samples at relative abundances approaching 1%, suggesting that these candidate groups are important AD community members whose metabolic roles are conserved across time^{8,35}.

Amplicon sequence variants affiliated with Methanogenic archaea had low relative abundances across the time series and included the families Methanofastidiosaceae (0.9%), Methanosetaeaceae (0.5%), Methanospirillaceae (0.3%), Methanoregulaceae (0.2%), Methanomicrobiaceae (0.1%), and Methanobacteriaceae (0.1%). Methanofastidiosaceae, the most abundant methanogenic lineage in Lulu Island AD samples, is a candidate group that lacks several canonical methanogenesis pathway components, likely using methylated thiol compounds for energy and CH_4 production³³. Methanosetaeaceae, primarily

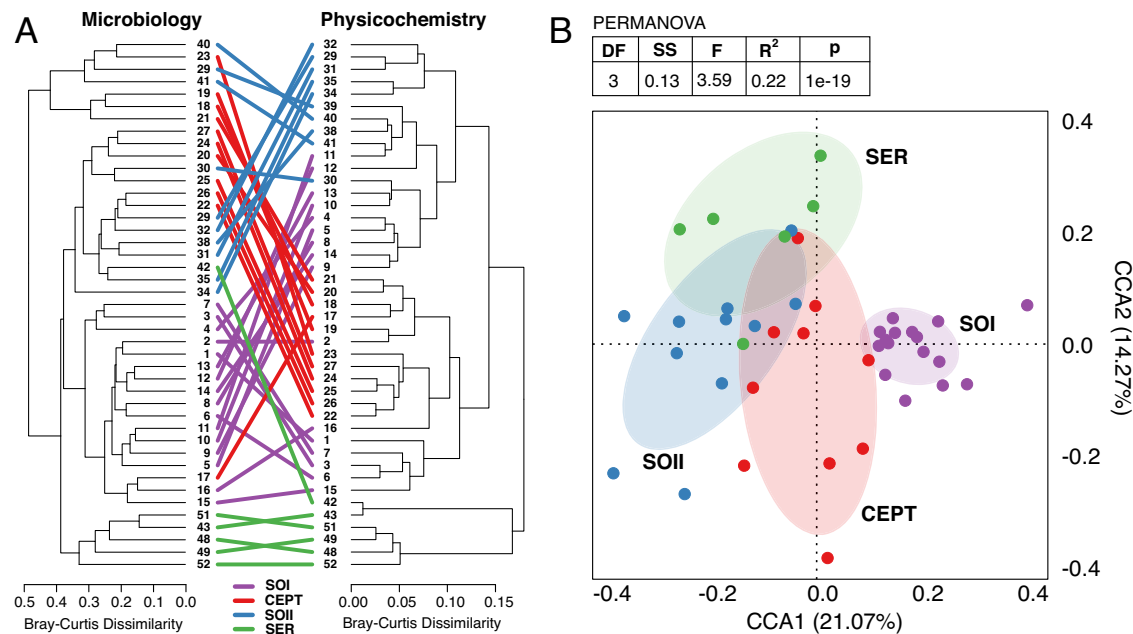


Fig. 3 | Relationships between microbial community structure and physicochemical conditions. **A** The structures of the AD microbial community (ASV data) and the AD physicochemical data are expressed as Bray-Curtis dissimilarity plotted in a dendrogram. Identical samples in each data-type dendrogram are connected with a line and colored corresponding to their process configuration. **B** The combined table of ASV and physicochemical data underwent principal coordinate

analysis and samples were plotted in the first two coordinate dimensions. The variance explained by each dimension is shown as a percent in the axis text. Samples are colored by their process configuration and ellipses sized by the 95% confidence of the centroid of each configuration group are laid in the background. The two-sided PERMANOVA table above the plot shows results of a test between the four configuration groups.

represented by *Methanotherx* populations in Lulu Island samples, convert acetate to CH₄ without the need for syntrophic interactions, while the four remaining lineages are obligate hydrogenotrophic methanogens^{36–38} that depend on syntrophic interactions with bacteria for electron equivalents driving CO₂ reduction to CH₄³⁹.

Relationships between AD microbial communities and physicochemical parameters

In ecosystems with well-controlled resource inputs such as ADs with constrained feedstock, organic loading rate, retention time, temperature, etc., environmental conditions and microbial community structure can fluctuate together in stable patterns^{4,40,41}. In the Lulu Island AD milieu, intervals with similar process configurations typically selected for similar microbial communities based on hierarchical cluster analysis (Fig. 3A), while sampling timepoints with distinct physicochemical parameters selected for distinct communities. Notably, at transition points between process configurations, when physicochemical parameters rapidly changed, microbial community composition rapidly responded in a time period shorter than the two-week resolution of our time series sampling, wherein microbial communities at the initial timepoint of a new operating condition still resembled that of the previous configuration before a new stable state was reached by the next sampling point (Fig. 3A).

By applying a dimension-reduction approach (i.e., canonical correspondence analysis) to a sample-by-sample distance matrix of the combined physicochemical parameter and ASV abundance information, process configuration was identified as a statistically significant variable in shaping the AD microbes and physicochemistry based on permutational analysis of variance tests (Fig. 3B). Using these combined datasets, it was observed that samples taken directly after transition to a new process configuration were more similar to the previous configuration. Given that no significant decline in RNG production was associated with these transitions (Supplementary Data 1), it appeared that detectable shifts in microbial community structure did not necessarily degrade AD performance with respect to overall

RNG yield. From these combined data, it was further determined from model fitting that the set of physicochemical parameters that most influenced microbial community structure was organic loading rate (kg/m³) and the concentrations of volatile acids, ammonia, and nitrate (mg/L) in the AD digestate, accounting for 27.6% of the temporal variation in microbial structure observed across the time series (Supplementary Data 7). These key parameters are typically identified as the main drivers of microbial structure and activity in mesophilic WWTPs⁴².

To account for population-level relationships between microbes and process configurations, ASV temporal distribution patterns were correlated with 28 measured physicochemical parameters across the time series. These results identified microbial populations and physicochemical parameters that were positively correlated (Fig. S4). One set of positively correlated variables included total biogas production (m³), organic loading rate (kg/m³), and several populations of Clostridia, Sphingobacteriales, Chloroflexi, and Spirochaetes. A second set of positive correlations, loosely associated with the first set, included AD hydraulic retention time (days), percent biogas composed of CO₂, concentration of effluent nitrate (mg/L), and several populations of Bacteroidales, Proteobacteria, and acetoclastic methanogens affiliated with *Methanotherx*. These groupings support previous observations that certain heterotrophic microorganisms, typically those engaged in hydrolysis coupled to fermentation, as well as acetoclastic methanogenic archaea, are most competitive when organic loading rate and retention time are high^{43, 44}. Although total gas production was positively correlated with this set of conditions and taxa, the percentage of CO₂ was also high relative to other timepoints, indicating that a larger share of organic carbon input was metabolized to CO₂ rather than to CH₄. This correlation of higher CO₂ percent with dominance of acetoclastic methanogens is not unexpected given the stoichiometric imbalance in CO₂:CH₄ between different methanogenic pathways^{45,46}. A third set of positively correlated variables included percent CH₄, temperature (°C), volatile fatty acids (mg/L), and several populations of Clostridiales, Bacteroidetes, and hydrogenotrophic methanogens

affiliated with *Methanoculleus* (Fig. S4). A fourth set, loosely correlated with the third, included BOD and COD, ammonia (mg/L) and suspended solids (mg/L), as well as hydrogenotrophic methanogens affiliated with *Methanomicrobiales*. Previous studies indicate that increases in both ammonia and fatty acid concentrations tend to select for hydrogenotrophic methanogenesis under mesophilic conditions^{12,47–51}. Shifts from acetoclastic to hydrogenotrophic methanogenesis have also previously been observed in WWTP ADs^{39,52,53}, reducing the methane potential of effluent^{54,55}, and resulting in higher methane content of RNG⁵⁶.

Taken together, these results indicate significant coupling between physicochemical parameters and both population- and community-level microbial structure across process configurations. Methanogenic archaea and some putative syntrophic bacteria that cooperate to reduce CO₂ to CH₄ were strongly selected during the Serial Operation, while direct acetate-reducing methanogens and other fermentative and non-syntrophic bacteria were selected in other process configurations. Physicochemical parameter measurements generally corroborated these patterns, including higher ammonia associated with increased hydrogenotroph/syntroph abundances, higher VFA concentrations associated with increased abundances of fermentative bacterial lineages, and higher RNG CO₂ content associated with increased acetoclastic methanogen abundances. Subsequent investigation of the time series focused on identifying microbial indicators engaged in metabolic interactions with potential to influence RNG purity and yield.

Identification of microbial indicators

During the time series, >25 complete volumetric turnover events occurred based on an average AD retention time of 30 days. Despite this recurring bottleneck, a robust core microbiome could be identified in the Lulu Island AD containing >30% of identified ASVs (339) in at least 80% of samples. This core collectively accounted for 64.4% of total 16S rRNA gene sequence reads. Given that input wastewater to the Lulu Island WWTP varied in origin and upstream processing across the time series, evidence of a robust core microbiome supports the hypothesis that selection factors such as environmental filtering and ecological interactions were likely more important than the initial community composition of waste inputs and the founder effect^{11,57,58}. Based on previous work, selection pressures that help maintain a core AD microbiome include strong physical constraints such as retention time and organic loading⁵⁹, chemical constraints such as low trace metal concentrations and highly anoxic, reducing conditions⁶⁰, and biological constraints such as metabolic interactions between co-occurring microorganisms⁶¹. Research to identify and characterize these modes of selection in the AD milieu are becoming increasingly important to the biotechnology sector, typically with the assumption that operational controls can be identified and leveraged to select for communities that improve RNG production^{6,20,28,62,63}.

Although a robust core microbiome persisted throughout the time series, the relative abundance of many taxa was significantly impacted by shifts in physicochemical parameters and process configuration. Through indicator species analysis, with configuration as the conditional variable and all ASVs tested for significant over-representation based on temporal abundance patterns, 138 indicator ASVs were identified, including both common and conditionally rare taxa (Fig. S3). Indicator ASVs were usually distributed such that each configuration was represented by a unique set of ASVs, even within a single family-level taxon, suggesting there may be subtle underlying diversification patterns supporting functional redundancy in the AD milieu. For example, while there were many indicators affiliated with the *Syntrophaceae* family of *Proteobacteria*, there were unique sets of *Syntrophaceae* indicator ASVs for Standard I, CEPT, and Standard II, and Serial Operations. Other taxa with indicator ASVs were more specific in their representation of a given process configuration.

For example, 10/13 ASVs from the *Rikenellaceae* family of *Bacteroidetes* were affiliated with the Serial Operation. Another notable pattern was the partitioning of archaeal indicator ASVs between configurations based on methanogenic phenotype. For example, the sole archaeal indicator for Standard Operation I was a population of *Methanosaetaceae*, which perform acetoclastic methanogenesis, while the two archaeal indicators for Serial Operation were affiliated with *Methanoregulaceae* and *Methanobacteriaceae*, which only perform hydrogenotrophic methanogenesis. Taken together, indicator analysis at the ASV level provided further evidence that the activity of co-occurring populations, both as core and configuration-dependent consortia, helps shape the active community structure driving RNG production.

Time-resolved correlation network analysis

Significant correlations were identified between AD process parameters, and both microbial community structure and the abundance of specific microbial populations relevant to RNG production over time. Based on the distributed nature of biomass conversion to methane between different microbial populations in WWTP ADs, we hypothesized that different process configurations would not only select for different populations but also that metabolic interactions among and between populations would vary under selection. To test this hypothesis, a co-occurrence network based on normalized ASV abundances across the two-year time series was constructed from a sparse inverse covariance matrix. Metadata information about nodes in the network, such as indicator ASV status, functional information from a paired MAG, and connectivity to other nodes, were then mapped onto the network to identify configuration-dependent subnetworks containing ensembles with potential to drive RNG production.

The ASV co-occurrence network was composed of 390 ASVs which had significant covariance across the time series (Fig. 4). The average clustering coefficient of the network (i.e., a measure of grouping or density among nodes) was 0.16, placing it well-within the range of microbial food webs or functional networks, and higher in connectivity than randomly produced associations between populations (Supplementary Data 2)⁶⁴. ASVs which were indicators for a given configuration tended to be connected in the network (a result of covariance between the indicator analysis and the co-occurrence model) and formed strongly correlated subnetworks within the larger parent network. Similarly, closely related ASVs (e.g., those with >97% 16S V4 rRNA homology) tended to coalesce into subnetworks, suggesting that populations with low phylogenetic distance tended to share similar temporal distribution patterns and conserved functional roles due to similar selection pressures (see nodes and edges in Supplementary Data 8 and Supplementary Data 9).

Genome-resolved correlation network mapping

Although taxonomically labeled ASVs can provide a way to predict trait-based information that is useful for inferring metabolic interactions, many AD microorganisms have poor taxonomic classifications and additional genome-resolved analysis is needed to assign potential functional roles^{8, 9,20,27,43}. Instead of using reference databases to infer functions from ASV nodes in the network, MAGs were generated from a set of 17 representative samples used for metagenomic whole genome shotgun sequencing across the time series with the goal of linking MAGs to ASVs and then ascribing functions to network nodes. The resulting metagenome datasets had an average assembled length of 657.4 Mbp, from which 40 high-quality MAGs (HQ: >90% complete and <5% contamination, with at least one full ribosomal RNA operon) and 475 medium-quality MAGs (MQ: >50% complete and <10% contamination, with at least one gene copy of the three ribosomal RNA subunits) were binned. Metagenome read mapping indicated that HQ and MQ bins represented on average 12.03% of the total quality base pairs sequenced per sample (Supplementary Data 2). The resulting

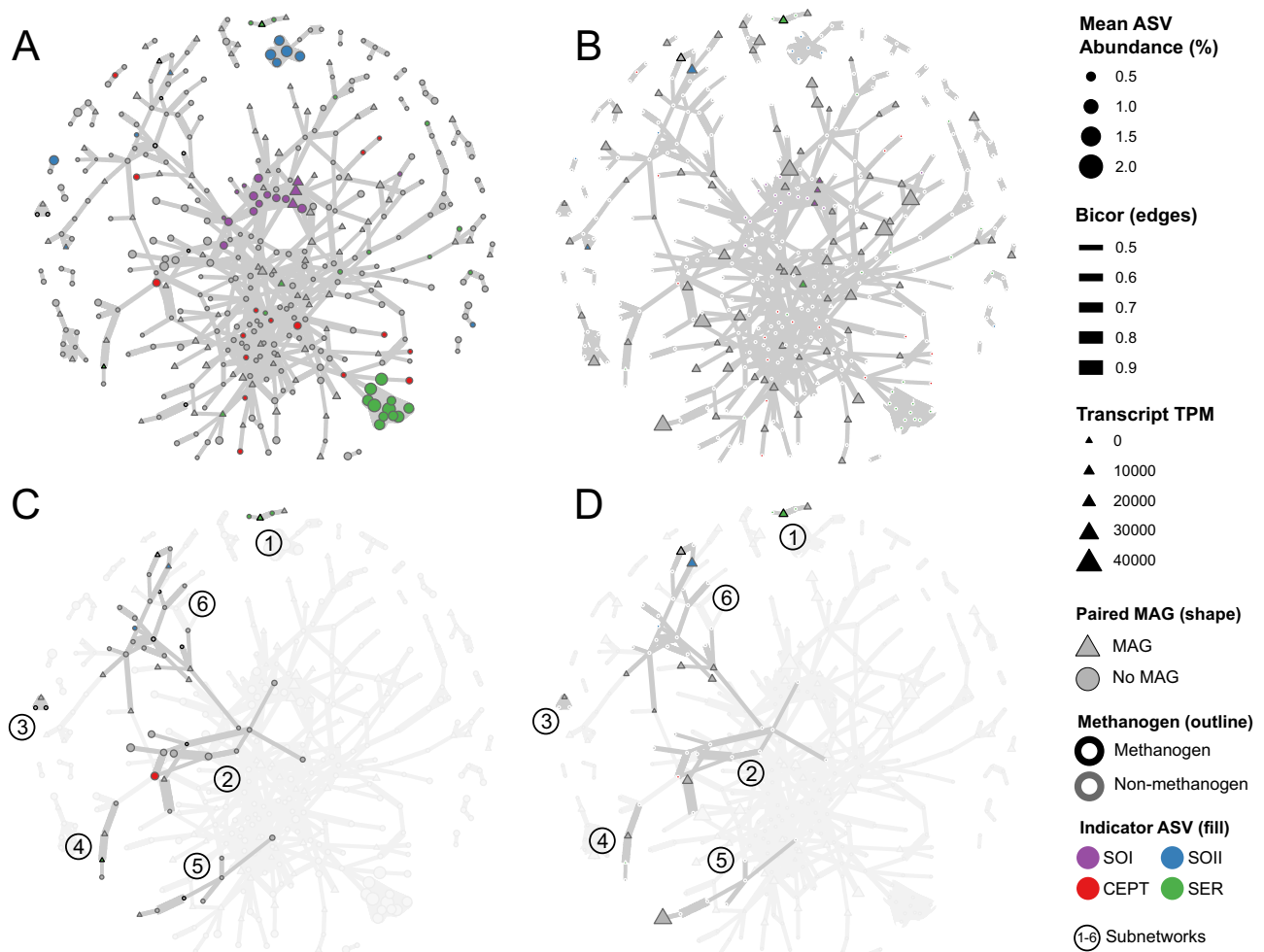


Fig. 4 | A weighted co-occurrence network of the microbial community depicting multiple-test-corrected biweight midcorrelation values (edges) of ASVs (nodes) across the two-year time series. Panels (A) and (C) depict the network wherein node sizes represent ASV mean relative abundance across all samples, while Panels (B) and (D) depict node sizes which represent transcript

expression levels (TPM) of the representative node MAG. Nodes are shaded with color if they are indicator ASVs for a given process configuration, outlined in black if they are methanogen ASVs, and shaped as triangles if they have a representative MAG. Subnetworks are numbered and highlighted in Panels (C) and (D).

MAGs were mapped onto cognate ASVs based on 16S rRNA gene sequence homology, enabling genome-resolved analysis of nodes within the network. A total of 233 ASVs (25.1% of total ASVs) could be paired with a unique MAG at a 16S rRNA gene sequence homology cutoff of >99.5% over the full length of the V4 region (-1 allowed mismatch). Of the 233 ASV-MAG pairs identified by homology, a total of 186 shared the same lowest-common ancestor taxonomy between the ASV V4 region and the MAG genomic sequence and were used for genome-resolved correlation network mapping (see methods).

Thus, the ASV time-series data, enabled by high-resolution sampling, were used to build a network of co-occurring ASVs that characterized temporally coherent taxa which responded similarly to changes in WWTP process configuration. Then, metagenomic sequence information was used to overlay a functional architecture onto the network that used representative MAGs to describe potential metabolic interactions between co-occurring populations at the level of genes, reactions and pathways. The 186 ASV-MAG pairs were distributed throughout the parent network, suggesting good coverage of taxonomic lineages and process configurations (Fig. 4).

In addition to linking ASVs to MAGs via 16S rRNA gene identity, metatranscriptomes were generated from the 17 samples with metagenomes (Supplementary Data 3), allowing assignment of gene expression information (as transcripts per million, TPM) to MAGs in the network (Supplementary Data 10). A notable observation from this

analysis was that methanogens displayed a strong decoupling between the metagenome- and metatranscriptomes-based measured relative abundances regardless of methanogenic lineage⁶⁵⁻⁶⁷, most of which had low TPM in metagenomes but very high TPM in metatranscriptomes (Supplementary Data 10). However, when only considering bacterial MAGs and not methanogen MAGs, the correlation between metagenome TPM abundance and metatranscriptome TPM abundance was strong ($\rho = 0.53$; $t = 13.81$; $df = 472$; p -value < 0.001), indicating that bacterial relative abundance observed in metagenome libraries was predictive of enzymatic activity (metatranscriptomes) in the AD milieu.

Subnetworks of methanogenic consortia

Taken together, the genome-resolved correlation network contained information about (1) temporal patterns of population abundance from the ASV time series, (2) indicator taxa for the various AD process configuration, (3) a subset of network nodes (47.7%) associated with MAGs, and (4) gene-level expression of each MAG. Six subnetworks of co-abundant methane-producing populations were identified within the parent network (Subnetworks 1–6), which were constructed by identifying hydrogenotrophic methanogen ASVs and all nodes connected by primary or secondary co-occurrence edges (one or two degrees of separation). These subnetworks represented ensembles with potential to drive RNG production across the time series, with

supporting evidence for biogas-producing syntrophic metabolisms gleaned from metabolic pathway reconstruction and gene expression data associated with MAGs mapped onto cognate ASVs within subnetworks.

Subnetwork 1 included four populations: three Rikenellaceae ASVs annotated as MiDAS species 240 (M.S.240) and M.S.3232, two of which were Serial Operation indicators (one of the Rikenellaceae indicator ASV had a corresponding MAG that was annotated to Tenuifilaceae in GTDB), and one Methanoregulaceae ASV (M.S.4938), which was also a Serial Operation indicator with an associated MAG (Supplementary Data 8). This small Subnetwork was isolated from the rest of the parent network and, given that 75% of nodes were indicator taxa, the organisms in this subnetwork were significantly more abundant during the Serial Operation when RNG purity was highest. The two MAGs in Subnetwork 1 were both high-quality, with completeness and contamination metrics of 93.3% and 1.4% (Tenuifilaceae; 3300028576_27) and 99.0% and 0% (Methanoregulaceae; 3300036947_39). Due to the strong co-occurrences during Serial Operation and the quality of the MAGs in Subnetwork 1, the encoded and expressed functions of the two genomes were examined for evidence of potential metabolic interactions. As expected, the Methanoregulaceae MAG encoded and expressed all genes necessary for hydrogenotrophic methanogenesis (e.g., *mcrABCDG*, genes Ga0377204_000018.861 – 865 and *mtrABCDEFH*, genes Ga0377204_000018.866 – 873) and lacked acetate kinase (*ack*) and phosphoacetyl transferase (*pta*) needed for acetoclastic methanogenesis (Figure S5; Supplementary Data 11). The Tenuifilaceae MAG that mapped to the subnetwork ASV encoded and expressed a partial Wood Ljungdahl (WL) pathway (*fdh*, *fhs*, *fol*, *met*) but lacked the CODH/ACS complex considered necessary for canonical acetate-oxidizing syntrophy by reverse WL⁶⁸ (Figure S5; Supplementary Data 11). The MAG did, however, both encode and express genes necessary for oxidizing acetate to 5,10-methylenetetrahydrofolate via the glycine cleavage system (Ga0255340_1000079.146, Ga0255340_1000129.100, and Ga0255340_1003402.5), indicating a potential route connecting acetate to its partial reverse WL methyl branch^{33,56}. While this glycine cleavage system syntrophic mode has been proposed in several taxa there is currently no definitive evidence of its activity^{33,56,62,69–71}, and many organisms use this system in amino acid biosynthesis reactions and other diverse functions independent from the reverse WL pathway⁷². The Tenuifilaceae MAG also encoded and expressed organoheterotrophic functions common to other Bacteroidetes, including a sus-like biopolymer degradation and transport system and several peptidases (Supplementary Data 11), indicating that these three ASVs could also be co-occurring with Methanoregulaceae through non-syntrophic modes of interaction (e.g., by catalyzing the rate-limiting depolymerization steps upstream of methanogenesis or through co-selection because of similar preferences for the conditions in the AD under Serial Operation).

Subnetworks 2, 3, and 5, centered around *Methanobacterium lacus*, *Methanospirillum* M.S.2576, and *Methanospirillum* M.S.2576, respectively, none of which had associated high-quality MAGs. Subnetworks 2 and 5 each had an ASV from a syntrophic bacterial lineage (*Smithella* and Syntrophaceae) directly connected to the methanogen node (Supplementary Data 9), while Subnetwork 3 was an isolated group containing three nodes (two methanogen and one Paludibacteraceae ASVs). Subnetwork 2 was the only subnetwork containing hydrogenotrophic methanogens that also included a Cloacimonadaceae ASV despite the latter lineage exhibiting the highest total relative abundance in the time series analysis and representing 10.5% of nodes in the parent network. This suggests that Cloacimonadaceae are likely not closely associated with syntroph-dependent methanogens in the Lulu Island AD milieu.

Subnetwork 4 contained a Clostridia ASV (Christensenellaceae R-7 M.S.240) with an associated MAG (3300028677_44), a Bacteroidales

UCG-001 ASV (M.S.1138), a Synergistaceae ASV (M.S.2022), and a *Methanobacterium lacus* ASV with an associated MAG (3300028677_53) that was also an indicator for Serial Operation (Supplementary Data 8). The two MAGs in Subnetwork 2 had completeness and contamination of 78.3% and 0.4% (Christensenellaceae) and 94.4% and 0.8% (Methanobacteriaceae), respectively (Supplementary Data 10), and the Christensenellaceae was directly connected with the syntrophy-dependent *Methanobacterium* methanogen. This Firmicutes family includes populations of peptide/amino acid fermenters and potentially H₂-producers^{73,74}, and though it has no isolated representatives, the Christensenellaceae R-7 lineage has been observed to comprise up to 6% of the AD community in some mesophilic digesters¹⁰. According to functional data from the Christensenellaceae MAG in Subnetwork 4, this population encoded and expressed several peptide and amino acid metabolism functions (Supplementary Data 11), including three major operons (two *livFGHKM* and one *azlCD*) for branched-chain amino acid transport, over 15 metabolic amino/exo/endo/oligopeptidase enzymes, and the hallmark fermentation enzyme pyruvate:ferredoxin oxidoreductase (Ga0255346_1000279.56). Between the Bacteroidales (as a depolymerizer⁷⁵), Christensenellaceae (as a fermenter), Synergistaceae (as an SAOB^{56,76}), and *Methanobacterium* (as a hydrogenotrophic methanogen) ASVs, Subnetwork 4 retained the metabolic capacity to carry out hydrolysis, fermentation, acetate oxidation, and methanogenesis of complex organic material to RNG. The observation that this *Methanobacterium* ASV was also a statistical indicator for the Serial Operation further suggests that this ensemble was most abundant in time series analysis during the period of high local RNG purity.

Subnetwork 6 was the largest co-occurring set of ASVs and included 25 nodes with 7 total ASV-MAG pairs (Supplementary Data 8). This Subnetwork was formed by four smaller but contiguous subnetworks. With four hydrogenotrophic methanogens in the 25-node subnetwork, Subnetwork 6 accounted for 40% of hydrogenotrophic methanogen ASVs in the parent network. Among the 25 nodes were also two indicator ASVs for Standard Operation II. Subnetwork 6 included several syntrophic lineages that had primary or secondary edges connected to the methanogen nodes, including two *Syntrophomonas* (both M.S.3971), one *Syntrophorhabdus* (M.S.998), and one Synergistaceae (*Thermovirga* M.S.988)^{56,77}. Based on 10,000 permutations of 25 random nodes from the parent network, the probability of observing four syntroph nodes and four hydrogenotrophic methanogen nodes by chance was 2.8%, showing that Subnetwork 6 was rich in syntrophic interactions driving RNG production.

Nodes with representative MAGs in Subnetwork 6 were two Clostridia ASVs (D8A-2 lineage and *Ruminococcus*) and a single ASV from Desulfobacterota (*Syntrophorhabdus* M.S.998), Deltaproteobacteria (*Phaselicystis* M.S.2086a), Paludibacteraceae (M.S.2677), Patescribacteria (Candidatus Falkowbacteria M.S.5033), and *Methanolinea* (M.S.4938). The directly co-occurring *Methanolinea* and *Syntrophorhabdus* MAGs were 99.0% complete with 0% contamination (3300036947_39) and 72.9% complete with 8.6% contamination (3300028576_34), respectively, and each had a full rRNA operon. The 16S rRNA gene sequence of the *Syntrophorhabdus* MAG (across the whole 1482 bp gene) was 93.48% similar to the best cultured representative *Syntrophorhabdus aromaticivorans* str. UI. Similar to this type strain, the *Syntrophorhabdus* MAG encoded and expressed benzoate-CoA ligase for benzoate degradation to benzoyl-CoA (Ga0255340_1019139.2) and two benzoyl-CoA reductase operons for converting benzoyl-CoA to the dienoyl-CoA intermediate (Ga0255340_1011848.9 and Ga0255340_1011848.11, plus Ga0255340_1031544.4 and Ga0255340_1031544.6), including two operons of the heterodisulfide reductase involved in this endergonic step; the benzoyl-CoA reductase enzymes were highly expressed (TPM = 17.53; 30th most abundant ORF in the genome), indicating that this pathway is a critical metabolic step for *Syntrophorhabdus*

(Supplementary Data 11). To complete the oxidation of benzoate to H₂, CO₂, and fatty acids, the MAG encoded and expressed syntenic genes for dienoyl-CoA hydration (Ga0255340_1005473.7), breaking and hydrolyzing the intermediate ring (Ga0255340_1005473.3 and Ga0255340_1005473.5), and finally for beta-oxidation (Ga0255340_1013965.3-7).

Taken together, these observations provide an overlaid picture of population-level co-occurrence, metabolic interactions facilitating that co-occurrence, and gene expression data to support those metabolic interactions. This study also provides a model for conducting time series amplicon analysis coupled with genome-resolved correlation mapping to identify known and novel metabolic interactions underlying organic waste conversion to RNG with potential to define new design principles that improve RNG quality and yield at scale. These same design principles can in turn be extended to the production of other value-added compounds, hastening the transition from waste treatment to sustainable waste resource recovery.

Methods

Sample collection and processing

We conducted a time series sampling campaign of AD microbial community structure, function, and dynamics at the Lulu Island WWTP in order to identify taxa associated with enhanced RNG production. Metro Vancouver provided physicochemical process parameters and triplicate samples of overflow digestate from anaerobic digesters (AD1 and AD2) at the Lulu Island WWTP on a biweekly basis for two years resulting in a sample collection archive consisting of 28 different physicochemical parameter measurements as well as microbial DNA for amplicon and metagenomic whole genome shotgun sequencing, and RNA for metatranscriptomic sequencing across 43 sampling dates (Supplementary Data 1). After transport to the lab, mixed sludge samples were centrifuged for 15 min at 14,000 × *g*, followed by chemical flocculation of the remaining cells in the supernatant with FeCl₃ and NaOH⁷⁸. After a final spin the supernatant was discarded, and the pellet stored at −80 °C prior to nucleic acid extraction. Mixed sludge DNA was extracted from frozen pellets using a DNeasy PowerSoil kit from Qiagen according to the manufacturer's specifications. DNA quantity and purity (260/280 values) were measured using a NanoDrop (Thermo Fisher). Mixed sludge RNA was extracted from frozen pellets using RNeasy PowerSoil kit from Qiagen. RNA quantity and quality were measured using a NanoDrop.

Small subunit ribosomal RNA gene amplicon sequencing

Microbial community structure and dynamics of mixed sludge samples was determined using PCR amplicon libraries targeting the V4 region of the bacterial and archaeal (prokaryotic) small subunit ribosomal RNA (SSU or 16S rRNA) gene using the 515F (GTGY-CAGCMGCCGCGGTAA) and 806R (GGACTACNVGGGTWTCTAAT) primers according to Earth Microbiome Project guidelines⁷⁹. A barcode was added to each amplicon library for multiplex sequencing on the Illumina MiSeq platform. The QIIME2 function demux was used to demultiplex the resulting FASTQ-formatted sequence files into separate samples, which were uploaded to the NCBI SRA database under PRJNA902729 as SRR22733982 – SRR22734080. The QIIME2 DADA2 pipeline was used to denoise amplicon libraries, filter chimeras, and create a feature table of amplicon sequence variants (ASVs)^{80,81}. ASVs were filtered to remove those with less than one observation on average across all samples and assigned taxonomy based on the QIIME2 feature-classifier-sklearn function with confidence set to 0.75 using a local install of the Silva v132 database⁸². ASV sequences were also assigned taxonomy using the same method to the GTDB bacterial and archaeal 16S rRNA gene database (v202) to compare taxonomy to the paired MAGs (see later methods). 16S V4 rRNA gene amplicon sequences for all samples were uploaded to the NCBI GenBank

database under BioProject PRJNA902729 as BioSamples SAMN32228586 - SAMN32228684.

Amplicon sequence analysis and metadata integration

Physicochemical process parameters and metadata were integrated into the QIIME2 feature taxonomy and count table using the add-metadata and convert functions of biom to create a tab separated table of ASV designators, their taxonomic affiliation, their amplicon read count across samples, and metadata associated with each sample⁸¹. A batch correction was employed to remove any effects of sequencing date (sequence libraries were generated on three separate plates) as a factor explaining biological variation using the R packages DESeq2 and limma^{83,84}. Some physicochemical parameters were not measured daily; therefore, an interpolation technique was used to estimate concurrent biological and physicochemical data. Briefly, any biological sampling dates missing from the physicochemical data table were added as an empty row, then interpolated using the smooth-fitting splinefun() function of the base R 'stats' package; parameters were checked manually by plotting to ensure reliable data interpolation between empirical physicochemical values. The 16S rRNA amplicon data (ASV phylogenetic tree, 16S rRNA gene sequences, ASV taxonomy, sample metadata) were imported into R and merged into a master phyloseq object using the function merge_phyloseq() of the 'phyloseq' R package⁸⁵.

Statistical analyses and data visualization of ASV data

All analysis and plotting code for the ASV analysis in this work can be found in the Supplementary Software. Taxonomic community composition of ASVs (as an average across all samples) was plotted using the sankeyNetwork() function of the 'networkD3' R package. In the final category of the Sankey diagram, the percent identity of ASV V4 sequences to the MiDAS database was determined using a local install of the BLASTn software (v2.5.0) with default parameters^{10,86}.

Hierarchical clustering of both ASV and physicochemical data was performed using the hclust() function of the base 'stats' R package, with distance matrices calculated using the Bray-Curtis dissimilarity metric on the ASV count table and physicochemical metadata table exported from the master phyloseq object. The hierarchical clustering dendrograms from the two datasets were corresponded using the tanglegram() function of the 'dendextend' R package⁸⁷. Canonical correspondence analysis (CCA) was performed using a Hellinger transformation of the combined physicochemical and ASV data with the decostand() and cca() functions of the 'vegan' R package⁸⁸. The results were plotted using the 'ggplot2' R package⁸⁹. Permutational analysis of variance (PERMANOVA) of samples across the four AD Configurations in the CCA was calculated with the adonis() function of the 'vegan' R package.

The community co-occurrence network was calculated by directly passing the master phyloseq object to the spiec.easi() function from the 'SpiecEasi' R package, with the glasso method, a minimum lambda ratio of 0.1, and the bstars selection criteria⁹⁰. Nodes in the network represent ASVs observed in at least 25% of samples across the time series and are sized by mean relative abundance. Vertices in the network represent positive co-occurrence values, with vertex width being proportional to the co-occurrence value calculated from the spiec.easi() function.

Microbial taxonomic compositional data across individual samples and lineages was exported from the master phyloseq object and plotted in barplot and scatterplot form using functions within the 'ggplot2' R package. Indicator species analysis using the multipatt() function of the 'indicpecies' R package was used to derive statistically significant associations between ASVs and each configuration⁹¹. Each indicator ASV was grouped by taxon and results were plotted using 'ggplot2'.

A maximum rank correlation test was performed to determine the set of physicochemical variables which best explained ASV community structure using the *bioenv()* function of the ‘vegan’ R package with default parameters. Correlations between temporal physicochemical parameter values and abundances of individual ASVs were calculated with the *cor()* function of the base ‘stats’ R package and plotted as a heatmap using ‘ggplot2’.

Metagenomic shotgun sequencing

A subset of 17 time series samples (Supplementary Data 1) were sequenced at the Joint Genome Institute on either the Illumina HiSeq-2000 ITB or Illumina NovaSeq platforms (2x151bp reads) to generate metagenome-assembled genomes (MAGs). Read processing was performed in accordance with the JGI standard operating procedure 1064 using the *jgi_meta_run.py* (version 2.0.1) processing pipeline^{92,93}. The resulting paired-end reads were then assembled using SPAdes assembler (version 3.11) using a range of Kmers with the parameters “spades.py -m 2000 -tmp-dir tmp.5774024.0 -o spades3 --only-assembler -k 33,55,77,99,127 --meta -t 32 --reads1.fasta --reads2.fasta”⁹⁴. The entire filtered read set was mapped to the final assembly and coverage information was generated using *bbmap* (version 37.78) using default parameters except *ambiguous=random* (<https://jgi.doe.gov/data-and-tools/bbtools/>). Raw sequence data in FASTQ format are available on NCBI SRA and analysis and assemblies are available in the JGI GOLD database. Project and accession numbers are listed in Supplementary Data 3.

Genome-resolved metagenomic analysis and network mapping

Binning of assembled contigs into MAGs was performed using MetaBAT v2.12.1⁹⁵ implemented through the Genomes OnLine Database (GOLD) based on the established JGI workflow, and completion and contamination was assessed using CheckM v1.0.12⁹⁶. Taxonomic assignment was performed using GTDB (release 86) with GTDB-tk v0.2.2⁹⁷. MAG quality and taxonomic assignment were augmented with RNA detection including noncoding RNA (ncRNA), tRNAs, and rRNA (5S, 16S, 23S) genes using tRNAscan-SE v2.0.6 in “bacterial” and “archaeal” search modes and *cmsearch* from the INFERNAL v1.1.3 package against the Rfam 13.0 database using the trusted cutoffs parameter (*-cut_tc*)^{98,99}. Reported hit overlapping by ≥ 1 bp and belonging to the same Rfam class were identified and the lower scoring of the two was removed.

High- and medium-quality MAGs were identified on the basis of current community standards (Supplementary Data 2)¹⁰⁰. MAGs with >90% completeness, <5% contamination, and at least one of each rRNA subunit gene were retained as “high-quality” MAG populations. MAGs with >50% completeness, <10% contamination, and at least one rRNA subunit gene were retained as “medium quality” MAG populations. High- and medium-quality MAGs were uploaded as draft genomes to NCBI MAGs and were functionally annotated using the JGI’s IMG workflow, which implements NCBI RPSBLAST to assign COG IDs to COG database v2014, EBI’s *pfam_scan* tool (which uses HMMER v3.0) to assign PFAM IDs to Pfam database v30, LAST to assign KEGG KO Terms from IMG genes to the KEGG database v77.1, and LAST to assign EC numbers from IMG genes (using KO Terms) with a homology-based approach with a local installation of BLASTn v2.5.0⁹³.

DNA sequences of ASVs from the time series (~273 bp) were then searched as queries against a custom reference database of SSU rRNA gene sequences encoded in high- or medium-quality MAGs binned from the 17 metagenomes. Hits with >99.5% nucleotide identity over the full length of the query region were accepted (i.e., ASV V4 rRNA sequences with 0 or 1 mismatch along the 273 bp fragment to a MAG V4 rRNA region). For each of the 233 unique ASV-MAG pairs identified using this method, the GTDB taxonomy of each was compared and only pairs with a matching lowest-common-ancestor taxonomy were accepted (i.e., if an ASV was annotated to the order level and its MAG to

the family level, the order level taxonomy must agree between the ASV and MAG). This resulted in 186 query-reference hits with coherent taxonomies, which were considered as ASV-MAG pairs and allowed for MAGs to be mapped onto ASV nodes in the co-occurrence network. Based on functional annotations from the JGI IMG workflow, the presence of methanogenic and potential syntrophic metabolic pathways in a select set of MAGs (i.e., those forming strong subnetworks in the co-occurrence analysis) were inferred from key encoded enzymes and electron carriers. Marker genes were selected to cover the necessary and sufficient steps of methanogenesis and syntrophy by known pathways^{56,101–103}.

Metatranscriptomic sequencing and read mapping

Total RNA was extracted (from the same physical samples as for DNA) using the RNeasy Power Soil kit from Qiagen according to manufacturer specifications. RNA quality was checked using a bioanalyzer (Agilent) to measure concentration and RNA integrity number (RIN) a measure of the quality and integrity of the RNA. Samples were also measured on the NanoDrop (Thermo Fisher) for concentration and the 260/280 value to assess RNA purity. A total of 14 RNA extracts with 260/280 >1.90 and RIN ≥ 5.0 were sent to GeneWiz for high-throughput sequencing of the community mRNA using “Standard RNA-seq” package. Details of this procedure are proprietary; more information can be found here (<https://www.genewiz.com/Public/Services/Next-Generation-Sequencing/RNA-Seq/>). Briefly, total RNA underwent rRNA depletion to concentrate mRNA transcripts, which were library prepped and sequenced Illumina HiSeq-2000 ITB platform (2 × 150bp reads). Raw reads were adapter trimmed, quality trimmed, and quality filtered using TrimmomaticSE (v0.35) with the parameters *-phred33, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36*. Quality-filtered RNA-seq read data in FASTQ format were uploaded to the NCBI under BioProject [PRJNA902729](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA902729) as BioSamples SAMN34106045 - SAMN34106058.

To infer and quantify gene expression of protein-coding genes on contigs assembled from each metagenome, quality-filtered transcripts from each sample were mapped to their respective metagenome library contigs using *bowtie2* v2.2.5 with parameters *-end-to-end* and *-very-sensitive*¹⁰⁴. The resulting BAM mapping files were further processed to identify and remove PCR duplicates using *samtools* and the *MarkDuplicates* function from Picard Toolkit (<http://broadinstitute.github.io/picard/>). The mRNA copy number of genes from each metagenome was then calculated using a protocol designed by the Environmental Genomics Group SciLifeLab at KTH Stockholm (<https://metagenomics-workshop.readthedocs.io/en/latest/annotation/quantification.html>). Briefly, HTSeq was used to count the number of reads mapped to features (e.g., genes, tRNA, etc.) on reference contigs using the GFF3 feature file generated during metagenome analysis. The number of metatranscript reads mapped to each feature were then normalized using the TPM (Transcripts per million) method. The resulting TPM tables for each of the 14 metagenome-metatranscriptome pairs (containing rows of gene features with TPM expression values) were parsed to calculate the percentage of metatranscript reads mapped to MAG bins and to associate TPM to functional pathways of interest in MAGs from the co-occurrence subnetworks (Supplementary Data 3).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data generated for this study can be found deposited in public repositories. For 16S rRNA amplicon sequences, reference the NCBI GenBank database under BioProject [PRJNA902729](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA902729) as BioSamples SAMN32228586 - SAMN32228684. For whole-genome shotgun

metagenome sequences, reference the NCBI SRA database under BioProjects [PRJNA502381](#), [PRJNA502380](#), [PRJNA518419](#), [PRJNA502378](#), [PRJNA502379](#), [PRJNA502376](#), [PRJNA502377](#), [PRJNA502375](#), [PRJNA620832](#), [PRJNA620833](#), [PRJNA620758](#), [PRJNA620835](#), [PRJNA620756](#), [PRJNA620838](#), [PRJNA620839](#), [PRJNA620843](#), and [PRJNA620754](#) (See Supplementary Data 3 for details and associations to JGI GOLD analysis accessions). For shotgun metatranscriptomic sequences, reference the NCBI SRA database under BioProject [PRJNA902729](#) as BioSamples SAMN34106045 - SAMN34106058.

Code availability

While all open source and freely available tools were used for this analysis, a markdown of the code used to conduct statistical analysis and create data visualizations in R v4.1.2 is referenced in the Supplementary Software.

References

- Adekunle, K. F. & Okolie, J. A. A review of biochemical process of anaerobic digestion. *Adv. Biosci. Biotechnol.* **06**, 205 (2015).
- Jimenez, J. et al. Instrumentation and control of anaerobic digestion processes: a review and some research challenges. *Rev. Environ. Sci. Biotechnol.* **14**, 615–648 (2015).
- Shaw, G. T.-W., Liu, A.-C., Weng, C.-Y., Chou, C.-Y. & Wang, D. Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. *PLoS ONE* **12**, e0181395 (2017).
- Werner, J. J. et al. Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl Acad. Sci. USA* **108**, 4158–4163 (2011).
- Zhang, T., Shao, M.-F. & Ye, L. 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J.* **6**, 1137–1147 (2012).
- Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J. & Tyson, G. W. Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.* **27**, 55–64 (2014).
- Cabezas, A. et al. How to use molecular biology tools for the study of the anaerobic digestion process? *Rev. Environ. Sci. Biotechnol.* **14**, 555–593 (2015).
- Calusinska, M. et al. A year of monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but different core microbiomes in bio-waste and wastewater anaerobic digestion systems. *Biotechnol. Biofuels* **11**, 196 (2018).
- Campanaro, S. et al. The anaerobic digestion microbiome: a collection of 1600 metagenome-assembled genomes shows high species diversity related to methane production. Preprint at *bioRxiv* <https://doi.org/10.1101/680553> (2019).
- Nierychlo, M. et al. MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge. *Water Res.* **182**, 115955 (2020).
- Tao, Y. et al. Biogas productivity of anaerobic digestion process is governed by a core bacterial microbiota. *Chem. Eng. J.* **380**, 122425 (2020).
- Hori, T., Haruta, S., Ueno, Y., Ishii, M. & Igarashi, Y. Dynamic transition of a methanogenic population in response to the concentration of volatile fatty acids in a thermophilic anaerobic digester. *Appl. Environ. Microbiol.* **72**, 1623–1630 (2006).
- Comino, E., Rosso, M. & Riggio, V. Investigation of increasing organic loading rate in the co-digestion of energy crops and cow manure mix. *Bioresour. Technol.* **101**, 3013–3019 (2010).
- Goux, X. et al. Microbial community dynamics in replicate anaerobic digesters exposed sequentially to increasing organic loading rate, acidosis, and process recovery. *Biotechnol. Biofuels* **8**, 122 (2015).
- Wintsche, B., Jehmlich, N., Popp, D., Harms, H. & Kleinstueber, S. Metabolic adaptation of methanogens in anaerobic digesters upon trace element limitation. *Front. Microbiol.* **9**, 405 (2018).
- De Vrieze, J. & Verstraete, W. Perspectives for microbial community composition in anaerobic digestion: from abundance and activity to connectivity: Microbial connectivity in anaerobic digestion. *Environ. Microbiol.* **18**, 2797–2809 (2016).
- Upadhyay, A. et al. A review of basic bioinformatic techniques for microbial community analysis in an anaerobic digester. *Fermentation* **9**, 62 (2023).
- Shaw, G. T.-W., Weng, C.-Y., Chen, C.-Y., Weng, F. C.-H. & Wang, D. A systematic approach re-analyzing the effects of temperature disturbance on the microbial community of mesophilic anaerobic digestion. *Sci. Rep.* **9**, 6560 (2019).
- Mei, R. et al. Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Res.* **124**, 77–84 (2017).
- Zhang, Q. et al. High variations of methanogenic microorganisms drive full-scale anaerobic digestion process. *Environ. Int.* **126**, 543–551 (2019).
- Campanaro, S., Treu, L., Kougias, P. G., Luo, G. & Angelidaki, I. Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. *Water Res.* **140**, 123–134 (2018).
- Tonanzi, B. et al. Microbial community successional changes in a full-scale mesophilic anaerobic digester from the start-up to the steady-state conditions. *Microorganisms* **9**, 2581 (2021).
- Kirkegaard, R. H. et al. The impact of immigration on microbial community composition in full-scale anaerobic digesters. *Sci. Rep.* **7**, 9343 (2017).
- Rivière, D. et al. Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J.* **3**, 700–714 (2009).
- Wu, L. et al. Long-term successional dynamics of microbial association networks in anaerobic digestion processes. *Water Res.* **104**, 1–10 (2016).
- Guo, J. et al. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. *Microb. Cell Factories* **14**, 33 (2015).
- Świątczak, P., Cydzik-Kwiatkowska, A. & Rusanowska, P. Microbiota of anaerobic digesters in a full-scale wastewater treatment plant. *Arch. Environ. Prot.* **43**, 53–60 (2017).
- Nguyen, A. Q. et al. Effects of operational disturbance and subsequent recovery process on microbial community during a pilot-scale anaerobic co-digestion. *Int. Biodeterior. Biodegrad.* **138**, 70–77 (2019).
- Pelletier, E. et al. “candidatus cloacamonas acidaminovorans”: genome sequence reconstruction provides a first glimpse of a new bacterial division. *J. Bacteriol.* **190**, 2572–2579 (2008).
- Wright, J. J., Konwar, K. M. & Hallam, S. J. Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **10**, 381–394 (2012).
- Wright, J. J. et al. Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *ISME J.* **8**, 455–468 (2014).
- Hawley, A. K. et al. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nat. Commun.* **8**, 1507 (2017).
- Nobu, M. K. et al. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* **9**, 1710–1722 (2015).
- Vendruscolo, E. C. G. et al. Microbial communities network analysis of anaerobic reactors fed with bovine and swine slurry. *Sci. Total Environ.* **742**, 140314 (2020).

35. Nobu, M. K. et al. Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* **10**, 273–286 (2016).
36. de Bok, F. A. M., Luijten, M. L. G. C. & Stams, A. J. M. Biochemical evidence for formate transfer in syntrophic propionate-oxidizing cocultures of *Syntrophobacter fumaroxidans* and *Methanospirillum hungatei*. *Appl. Environ. Microbiol.* **68**, 4247–4252 (2002).
37. Imachi, H. et al. *Methanolinea tarda* gen. nov., sp. nov., a methane-producing archaeon isolated from a methanogenic digester sludge. *Int. J. Syst. Evol. Microbiol.* **58**, 294–301 (2008).
38. Manzoor, S., Schnürer, A., Bongcam-Rudloff, E. & Müller, B. Complete genome sequence of *Methanoculleus bourgensis* strain MAB1, the syntrophic partner of mesophilic acetate-oxidising bacteria (SAOB). *Stand. Genom. Sci.* **11**, 80 (2016).
39. Demirel, B. & Scherer, P. The roles of acetotrophic and hydrogenotrophic methanogens during anaerobic conversion of biomass to methane: a review. *Rev. Environ. Sci. Biotechnol.* **7**, 173–190 (2008).
40. Falk, M. W., Song, K.-G., Matiassek, M. G. & Wuertz, S. Microbial community dynamics in replicate membrane bioreactors – Natural reproducible fluctuations. *Water Res.* **43**, 842–852 (2009).
41. Peng, X. "Nick", Gilmore, S. P. & O'Malley, M. A. Microbial communities for bioprocessing: lessons learned from nature. *Curr. Opin. Chem. Eng.* **14**, 103–109 (2016).
42. Jiang, C. et al. Characterizing the growing microorganisms at species level in 46 anaerobic digesters at Danish wastewater treatment plants: a six-year survey on microbial community structure and key drivers. *Water Res.* **193**, 116871 (2021).
43. Carballa, M., Regueiro, L. & Lema, J. M. Microbial management of anaerobic digestion: exploiting the microbiome-functionality nexus. *Curr. Opin. Biotechnol.* **33**, 103–111 (2015).
44. Li, Y. et al. Enhancement of anaerobic methanogenesis at a short hydraulic retention time via bioelectrochemical enrichment of hydrogenotrophic methanogens. *Bioresour. Technol.* **218**, 505–511 (2016).
45. Polag, D. et al. Online monitoring of stable carbon isotopes of methane in anaerobic digestion as a new tool for early warning of process instability. *Bioresour. Technol.* **197**, 161–170 (2015).
46. Ziels, R. M. et al. Microbial rRNA gene expression and co-occurrence profiles associate with biokinetics and elemental composition in full-scale anaerobic digesters. *Microb. Biotechnol.* **11**, 694–709 (2018).
47. Schnürer, A. & Nordberg, A. Ammonia, a selective agent for methane production by syntrophic acetate oxidation at mesophilic temperature. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* **57**, 735–740 (2008).
48. Wang, H., Fotidis, I. A. & Angelidaki, I. Ammonia effect on hydrogenotrophic methanogens and syntrophic acetate-oxidizing bacteria. *FEMS Microbiol. Ecol.* **91**, fiv130 (2015).
49. Westerholm, M., Moestedt, J. & Schnürer, A. Biogas production through syntrophic acetate oxidation and deliberate operating strategies for improved digester performance. *Appl. Energy* **179**, 124–135 (2016).
50. Yang, Y. et al. Enhanced anaerobic digestion of ammonia-rich swine manure by zero-valent iron: With special focus on the enhancement effect on hydrogenotrophic methanogenesis activity. *Bioresour. Technol.* **270**, 172–179 (2018).
51. Tian, H., Yan, M., Treu, L., Angelidaki, I. & Fotidis, I. A. Hydrogenotrophic methanogens are the key for a successful bioaugmentation to alleviate ammonia inhibition in thermophilic anaerobic digesters. *Bioresour. Technol.* **293**, 122070 (2019).
52. Petersen, S. P. & Ahring, B. K. Acetate oxidation in a thermophilic anaerobic sewage-sludge digester: the importance of non-aceticlastic methanogenesis from acetate. *FEMS Microbiol. Lett.* **86**, 149–158 (1991).
53. Dykstra, S., Jansen, L. & Gallert, C. Syntrophic acetate oxidation replaces acetoclastic methanogenesis during thermophilic digestion of biowaste. *Microbiome* **8**, 105 (2020).
54. Angelidaki, I., Boe, K. & Ellegaard, L. Effect of operating conditions and reactor configuration on efficiency of full-scale biogas plants. *Water Sci. Technol.* **52**, 189–194 (2005).
55. Markphan, W., Mamimin, C., Suksong, W., Prasertsan, P. & O-Thong, S. Comparative assessment of single-stage and two-stage anaerobic digestion for biogas production from high moisture municipal solid waste. *PeerJ* **8**, e9693 (2020).
56. Zhu, X. et al. Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion ecosystem. *Microbiome* **8**, 22 (2020).
57. Treu, L., Kougias, P. G., Campanaro, S., Bassani, I. & Angelidaki, I. Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. *Bioresour. Technol.* **216**, 260–266 (2016).
58. Peces, M., Astals, S., Jensen, P. D. & Clarke, W. P. Deterministic mechanisms define the long-term anaerobic digestion microbiome and its functionality regardless of the initial microbial community. *Water Res.* **141**, 366–376 (2018).
59. Appels, L., Baeyens, J., Degreve, J. & Dewil, R. Principles and potential of the anaerobic digestion of waste-activated sludge. *Prog. Energy Combust. Sci.* **34**, 755–781 (2008).
60. Shamurad, B. et al. Low-temperature pretreatment of organic feedstocks with selected mineral wastes sustains anaerobic digestion stability through trace metal release. *Environ. Sci. Technol.* **54**, 9095–9105 (2020).
61. Amani, T., Nosrati, M., Mousavi, S. M. & Kermanshahi, R. K. Study of syntrophic anaerobic digestion of volatile fatty acids using enriched cultures at mesophilic conditions. *Int. J. Environ. Sci. Technol.* **8**, 83–96 (2011).
62. Lee, S.-H. et al. Monitoring bacterial community structure and variability in time scale in full-scale anaerobic digesters. *J. Environ. Monit.* **14**, 1893 (2012).
63. Solli, L., Håvelsrud, O. E., Horn, S. J. & Rike, A. G. A metagenomic study of the microbial communities in four parallel biogas reactors. *Biotechnol. Biofuels* **7**, 146 (2014).
64. Steele, J. A. et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* **5**, 1414–1425 (2011).
65. Yu, K. & Zhang, T. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PLoS ONE* **7**, e38183 (2012).
66. Bremges, A. et al. Deeply sequenced metagenome and metatranscriptome of a biogas-producing microbial community from an agricultural production-scale biogas plant. *GigaScience* **4**, 33 (2015).
67. Kakuk, B. et al. Early response of methanogenic archaea to H₂ as evaluated by metagenomics and metatranscriptomics. *Microb. Cell Factories* **20**, 127 (2021).
68. Müller, B., Sun, L. & Schnürer, A. First insights into the syntrophic acetate-oxidizing bacteria—a genetic study. *MicrobiologyOpen* **2**, 35–53 (2013).
69. Zeng, Y. et al. Chasing the metabolism of novel syntrophic acetate-oxidizing bacteria in thermophilic methanogenic chemostats. Preprint at <https://doi.org/10.1101/2021.07.06.451242> (2021).
70. Li, C. et al. Syntrophic acetate-oxidizing microbial consortia enriched from full-scale mesophilic food waste anaerobic digesters showing high biodiversity and functional redundancy. *mSystems* **7**, e00339–22 (2022).

71. Zhu, X., Campanaro, S., Treu, L., Kougias, P. G. & Angelidaki, I. Novel ecological insights and functional roles during anaerobic digestion of saccharides unveiled by genome-centric metagenomics. *Water Res.* **151**, 271–279 (2019).
72. Cronan, J. E. Assembly of lipoic acid on its cognate enzymes: an extraordinary and essential biosynthetic pathway. *Microbiol. Mol. Biol. Rev.* **80**, 429–450 (2016).
73. Morotomi, M., Nagai, F. & Watanabe, Y. Description of *Christensenella minuta* gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **62**, 144–149 (2012).
74. Rago, L. et al. A study of microbial communities on terracotta separator and on biocathode of air breathing microbial fuel cells. *Bioelectrochemistry* **120**, 18–26 (2018).
75. Mei, R., Nobu, M. K., Narihiro, T. & Liu, W.-T. Metagenomic and metatranscriptomic analyses revealed uncultured bacteroidales populations as the dominant proteolytic amino acid degraders in anaerobic digesters. *Front. Microbiol.* **11**, 593006 (2020).
76. Yi, Y. et al. Identification of novel butyrate- and acetate-oxidizing bacteria in butyrate-fed mesophilic anaerobic chemostats by DNA-based stable isotope probing. *Microb. Ecol.* **79**, 285–298 (2020).
77. Su, X.-L. et al. *Acetobacteroides hydrogenigenes* gen. nov., sp. nov., an anaerobic hydrogen-producing bacterium in the family Rikenellaceae isolated from a reed swamp. *Int. J. Syst. Evol. Microbiol.* **64**, 2986–2991 (2014).
78. Kenward, P. A. et al. Recovering cellular biomass from fluids using chemical flocculation. *Environ. Microbiol. Rep.* **10**, 686–694 (2018).
79. Walters, W. et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *Msystems* **1**, e00009–e00015 (2016).
80. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
81. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
82. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
84. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
85. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217 (2013).
86. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **50**, D20–D26 (2021).
87. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
88. Oksanen, J. et al. *vegan: Community Ecology Package.* (2017).
89. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* (Springer-Verlag, 2009).
90. Kurtz, Z. D. et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
91. Cáceres, M. D. & Legendre, P. Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**, 3566–3574 (2009).
92. Grigoriev, I. V. et al. The genome portal of the department of energy joint genome institute. *Nucleic Acids Res.* **40**, D26–D32 (2012).
93. Chen, I.-M. A. et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
94. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
95. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
96. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* <https://doi.org/10.1101/gr.186072.114> (2015).
97. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
98. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
99. Chan, P. P. & Lowe, T. M. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol. Clifton NJ* **1962**, 1–14 (2019).
100. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
101. Hattori, S. Syntrophic acetate-oxidizing microbes in methanogenic environments. *Microbes Environ.* **23**, 118–127 (2008).
102. Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the Methanogenic Archaea. *Ann. N. Y. Acad. Sci.* **1125**, 171–189 (2008).
103. Costa, K. C. & Leigh, J. A. Metabolic versatility in methanogens. *Curr. Opin. Biotechnol.* **29**, 70–75 (2014).
104. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

Acknowledgements

This work was performed under the auspices of the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation (CFI), and the US Department of Energy (DOE) Joint Genome Institute (JGI) and the Facilities Integrating Collaborations for User Science (FICUS) JGI-EMSL (Environmental Molecular Science Laboratory) project (50967) supported by the Office of Science of US DOE Contract DE-AC02-05CH11231. Essential automation support came from the Biofactorial high-throughput biology facility in the Life Sciences Institute at the University of British Columbia. We would like to thank Connor Morgan-Lang, Kateryna Ievdokymenko, Julia Anstett, and Siddarth Raghuvanshi for insightful conversations regarding data analysis, Michael Yoon for helping with sample processing, and Tom Pfeifer in the Biofactorial automation facility at UBC for help in establishing workflows and conducting extractions on the Access workstation. We thank Metro Vancouver staff in the Wastewater Treatment Operations and Collaborative Innovation groups for initiating and supporting this work to better understand microbial interactions in its AD systems.

Author contributions

B.K., N.F., R.M., S.C., and S.J.H. designed experiments and sampling campaign. N.F. and S.C. were responsible for physicochemical measurements. B.K., R.M., A.N.N., and S.J.H. performed the microbiological data analyses. B.K. wrote analysis and figure generation code and performed integrative analysis for DNA and RNA data with input from S.J.H. M.K. contributed to figure design and interpretation. B.K. wrote the manuscript with input from N.F., R.M., A.N.N., S.C., M.K., and S.J.H.

Competing interests

S.J.H. is a co-founder of Koonkie Inc., a bioinformatics consulting company that designs and provides scalable algorithmic and data analytics solutions in the cloud. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40907-5>.

Correspondence and requests for materials should be addressed to Steven J. Hallam.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023