Disentangling methanogenic wastewater treatment bioreactor communities with probabilistic metabolic models

Andrew Freiburger (afreiburger@u.northwestern.edu)[1,3], Heather Nielsen[2], Daniel Morrow[1,2], Kathleen Beilsmith[3], Chris Henry[3], George Wells[2], Keith Tyo[1]

1. Northwestern University, Department of Chemical Engineering, 2145 Sheridan Road, Evanston, IL 60208
2. Northwestern University, Department of Civil and Environmental Engineering, 2145 Sheridan Road, Evanston, IL 60208
3. Argonne National Laboratory, Data Science and Learning Division, 9700 S. Cass Avenue, Lemont, IL 60439

Background and Motivation

Renewable natural gas (RNG) produced by anaerobic digestion (AD) of anthropogenic waste is a promising, carbon-neutral energy substitute for fossil-derived natural gas. Ex-situ biomethanation reactors isolate methanogenesis from upstream AD processes, utilizing the metabolism of hydrogenotrophic methanogens to **convert waste CO_2 to pipeline-ready RNG** (≥95% CH_4, ≤2% CO_2, ≤5% H_2) according to the reaction:

$$4H_2 + CO_2 \rightarrow CH_4 + H_2O$$

To maintain consistent production of pipeline-ready RNG, the reactor must be operated to select for and maintain an enrichment of hydrogenotrophic methanogens. This study aims to evaluate the relationship between growth media composition and microbial functional pathways, and their combined effect on reactor performance.

Experimental Approach

A bench-scale (0.45 L) membrane biofilm reactor (MBfR) was constructed and operated for 331 days with the aim to produce pipeline-ready RNG. Substrate gases H_2 and CO_2 were delivered via polymeric hollow-fiber membranes, with liquid recirculation and automated influent media, temperature (T = 37±1˚C), and pH control (pH = 7±0.1).

In addition to substrate gases, the reactor received an influent flow of media to supply key trace elements and nutrients. The recipe for the influent media was modified based on reactor performance until the system stabilized. After this, the H_2 and CO_2 loading rates were increased in an effort to determine the maximum production capacity of the system.

Key performance parameters include:

- Gas quality (% CH_4, CO_2, and H_2)
- CH_4 production rate (mL CH_4/min)

Biomass samples were preserved weekly. Samples correlated to significant events – such as changes to the media – were selected for 16S rRNA extraction, PCR, and sequencing.

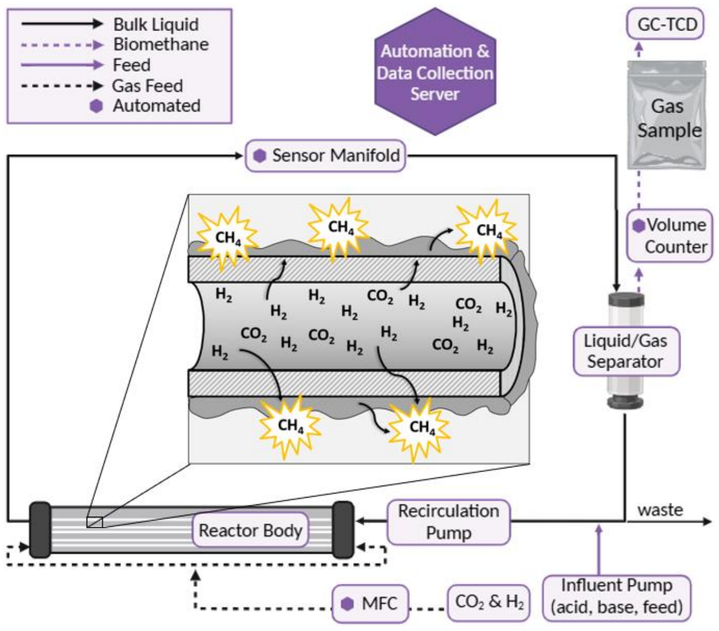


Figure 1. Membrane biofilm reactor (MBfR) configuration. Acronyms: mass flow controller (MFC), gas chromatography thermal conductivity detector (GC-TCD). Substrate gases – H_2 and CO_2 – are delivered in the lumen of hollow-fiber membranes. Trace metal and nutrient media is pumped in, continuously recirculated, and passively wasted via gravity overflow. Image partially drawn using Biorender©.

Three media recipes were tested & iteratively modified during the first 150 days of operation:

- DSMZ 120 – Supports specific methanogenic activity (SMA) tests for *Methanosarcina*. Contains organic carbon sources including acetate, yeast extract, casitone, and methanol
- Speece 1996 – Provides a baseline level of trace elements for methanogenesis in AD with a scaled increase in the loading rates of trace elements based on substrate loading rates
- Wolfe 2011 – Supports SMA tests for hydrogenotrophic methanogens

Observations

Point	Media	Modification
1	DSMZ 120	Removed acetate
2	DSMZ 120	Removed yeast, casitone, and bicarbonate
3	DSMZ 120	Removed methanol
4	Speece	Switched from DSMZ 120 to Speece
5	Wolfe	Switched from Speece to Wolfe
6	Wolfe	Changed organic to inorganic acid in Wolfe
7	Wolfe	Changed back to organic acid

High influent organic carbon:

- Leads to overproduction of CH_4 relative to H_2 loading rates
- Reduces product gas quality (i.e. excess CO_2)
- May support activity of non-preferred metabolic pathways

Stable performance was achieved after switching to Wolfe media.

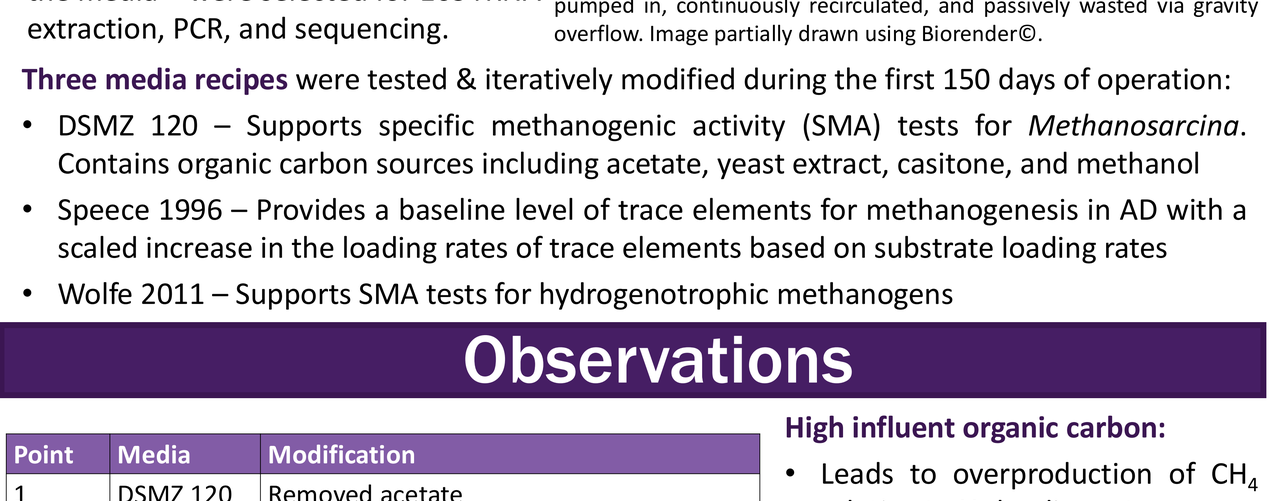
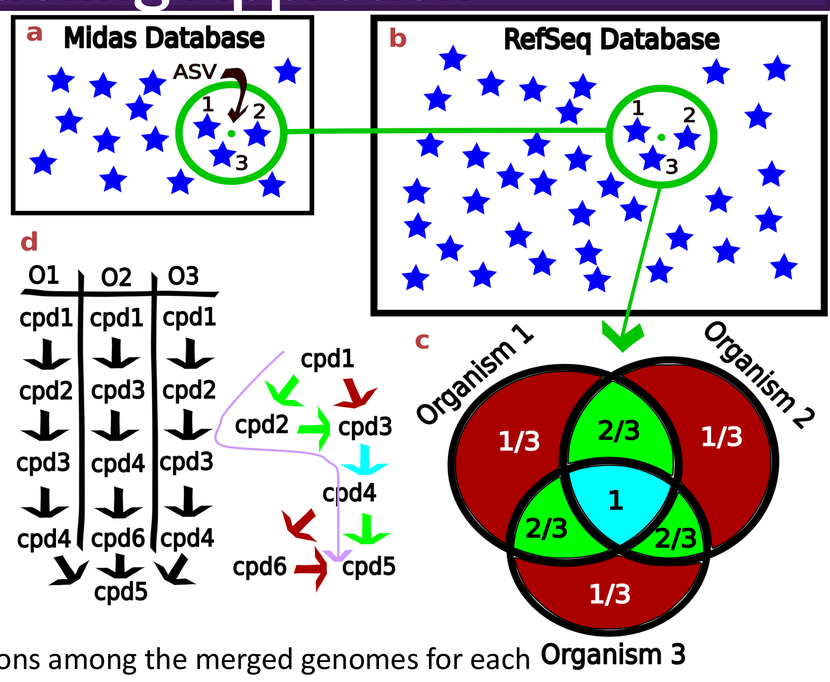


Figure 2. MBfR product gas quality and production rates correlating changes to influent media composition. The projected production rate is the expected CH_4 production rate based on H_2 loading rates. Media changes are described in the table above.

Modeling Approach

The general approach is illustrated in the figure at right. **Part a)** is the initial mapping of the experimental 16S ASVs into the Midas Database, which is a database of taxonomies for the 16S of common digestor sludge organisms. **Part b)** downloads these organisms' genomes via their taxonomy in the RefSeq Database. **Part c)** the genomes for organisms that mapped to each 16S are annotated via RAST and the union of annotations reconstructed into a metabolic model for that 16S.



Part d) the frequency of annotations among the merged genomes for each 16S is added as coefficient weight to the associated reactions in the linear optimization performed with the metabolic models. The pinkish line through the possible pathways in the 16S model utilizes the most conserved reactions in the simulations, attempting to capture biology.

Methods

The metabolic models will be constrained through several notable constraints. The community biomass (bio_{comm}) is a linear combination of the member biomass reactions (bio_1, …), weighted according to their relative abundances (a_1, …) in the community:

$$bio_{comm} = a_1 * bio_1 + a_2 * bio_2 + ... + a_n * bio_n$$

Members are prevented from contributing metabolism without growing themselves, which can be exploited by the optimization but deviates from perceived biological reality:

$$\sum_{rxn}^{R}(rxn_{forward} + rxn_{backward}) = KinCoef * bioRxn_{forward}$$

The limited nutritional environment of these systems is captured in the model through imposing a maximum number of moles of carbon atoms that can be consumed:

$$\sum_{exRxn}^{EX}(totalEle_{exRxn} * (exRxnExpr_{forward} \oplus exRxnExpr_{backward})) = eleLimit$$

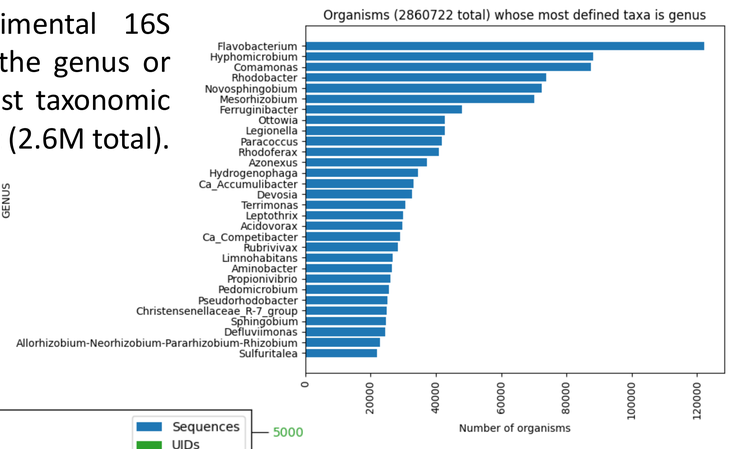
The final step of the community simulation limits the community biomass (bio_{comm}) to a minimum fraction of its maximum growth and minimizes the sum of all intracellular reactions, as a way to find the most parsimonious metabolic pathways, aligned with the assumption that evolution has sculpted cells to be optimally efficient:

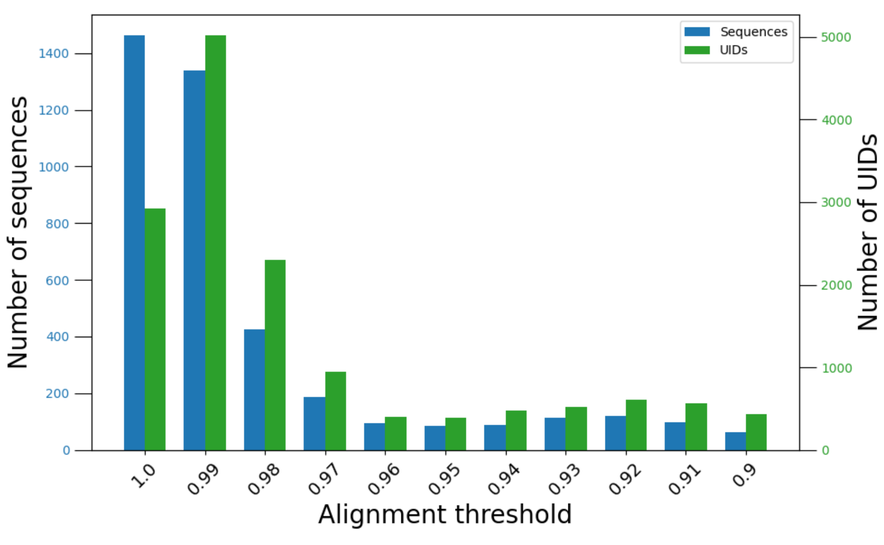
$$min\ \Sigma_r^R(flux_r * p_r)$$

Progress

Part a) mapped 2/3 of the experimental 16S sequences into the Midas database at the genus or species level. The organisms whose best taxonomic resolution is genus are presented at right (2.6M total).

Part b) mapped 4/5 of the Midas-derived taxonomy into the RefSeq to acquire UIDs. This resulted in the distribution in the figure below, where many sequences mapped perfectly or near perfectly, which supports Part c).





The next step is to process these UIDs into RefSeq Accession IDs that can then be used to download the genomes and proceed in reconstructing the metabolic models for simulation. These later steps are already encoded from previous work, so progress should accelerate after the genomes are acquired.

Acknowledgements

This work is possible due to funding through agreement DE-EE0009284 with the US Department of Energy under the direction of Beau Hoffman and Dr. Katie Davis.

Additional thanks to Dr. Yingqian Xiong, Emily Kin, Zhen Jia, Dr. Yingfen Zhao, and Dr. Neil M. Schweitzer for their support and expertise in Northwestern labs.