A genome-centric metagenomics approach to explain microbial community structure in anaerobic digesters

Neshat, Soheil Asgarineshat; Krithika Arumugam; Uma Shankari Chandra Segaran; Prabu Sekar; Purbojati, Rikky; Nguyen, T. Q. N.; Cokro, Angel Anika; Santillan, Ezequeil; Ng, T. C. A.; Williams, Rohan B. H.; Wuertz, Stefan

2022

Neshat, S. A., Krithika Arumugam, Uma Shankari Chandra Segaran, Prabu Sekar, Purbojati, R., Nguyen, T. Q. N., Cokro, A. A., Santillan, E., Ng, T. C. A., Williams, R. B. H. & Wuertz, S. (2022). A genome-centric metagenomics approach to explain microbial community structure in anaerobic digesters. Singapore International Water Week 2022.

https://hdl.handle.net/10356/156928

A genome-centric metagenomics approach to explain microbial community structure in anaerobic digesters

S. A. Neshat*, Krithika Arumugam*, Uma Shankari d/o Chandra Segaran*, Prabu Sekar*, Rikky Purbojati*, T.Q.N. Nguyen*, A. Anika Cokro*, Ezequeil Santillan T.C.A. Ng*, Rohan B. H. Williams, S. Wuertz *,*****

*Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

**Singapore Centre for Environmental Life Sciences Engineering, National University of Singapore, Singapore, Singapore

***School of Civil and Environmental Engineering, Nanyang Technological University Singapore, 639798, Singapore.

Corresponding authors: Stefan Wuertz, Professor of Civil and Environmental Engineering, swuertz@ntu.edu.sg, School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore and and Rohan B.H. Williams, lsirbhw@nus.edu.sg

Type of presentation: Poster

Theme and sub-topic the paper is to be submitted for:

- Theme 3 (Basic and advanced wastewater treatment processes - Application and integration of novel molecular biology tools)

SUMMARY

A functioning anaerobic digestion (AD) microbiome is integral for sludge management to be successful. Comprehensive ecological insights, and a full accounting of important microbial species, can help to improve and validate operational strategies. We analysed a time-series of metagenome samples obtained from full-scale anaerobic digesters and performed genome-resolved analysis to gain insight into the microbial community structure and potential functions of the AD microbiome. Ninety samples from three full-scale digesters were collected over a period of nine months and their nucleic acids extracted and subjected to shotgun sequencing (Illumina HiSeq2500; average 70M PE reads/sample). The raw reads were quality controlled, trimmed, assembled, binned, and dereplicated to obtain metagenome-assembled genomes (MAGs). Genome quality was assessed using the MIMAG criteria. The taxonomic assignment of the recovered MAGs was conducted using GTDB-Tk, and gene-level functional annotations were obtained using the KEGG, and CAZy databases. From ninety metagenome assemblies, 14,236 MAGs were recovered, of which 37%, 16%, and 1% satisfied medium-quality (Completeness > 50% and contamination <10%), putative high-quality (Completeness > 90% and contamination <5%), and high-quality criteria (Completeness > 90% and contamination <5% and presence of rRNA and tRNA genes), respectively. Taxonomical classification of the MAGs with at least medium quality (n =7666) revealed that 12.9%, 37.4%, and 77.1% of them belong to a novel family, genus, and species, respectively. A co-occurrence network analysis of the community structures in three replicate digesters revealed a highly interconnected network of microorganisms, suggesting the presence of a backbone in a functional AD microbial community. Functional analysis of the recovered MAGs showed the presence of three methanogenesis pathway modules, namely, methanogenesis via CO2, acetate, and trimethylamine. In addition, a specialization was observed in the hydrolytic bacterial community using CAZy annotation. In conclusion, we have obtained a catalogue of 166 MIMAG high-quality MAGs from a time-series metagenome survey of three full-scale anaerobic digesters situated in a tropical wastewater treatment plant, leading to novel ecological insights into the AD microbial community.

KEYWORDS

Anaerobic Digestion, Metagenomics, Genome-centric, microbiome

INTRODUCTION

Anaerobic digestion (AD) is a biological process widely known for its role in solid waste management. In this process, a consortium of microorganisms consisted of bacteria and archaea works together to convert organic materials into biogas (Neshat, Mohammadi, Najafpour, & Lahijani, 2017). Despite wide applications, the microbial community behind this process is not fully understood. Several studies have been conducted using culture independent techniques (such as amplicon sequencing of 16S rRNA genes) to investigate the AD microbial community, yet database limitations made it impossible to identify majority of novel species in AD community (Campanaro, Treu, Kougias, Zhu, & Angelidaki, 2018). In recent years, advent of bioinformatics techniques along with whole genome sequencing enabled us to recover complete genomes of novel species using genome-centric metagenomics. Using this technique, several novel species have been identified in the AD microbiome and novel functional pathways have also been introduced (Campanaro et al., 2016, 2020; Vanwonterghem, Jensen, Rabaey, & Tyson, 2016; Zhu, Campanaro, Treu, Kougias, & Angelidaki, 2019). Yet, more in-depth studies are required to understand the complexity of the AD microbial community dynamics at different settings. Here, using a genome-centric metagenomics approach, we analyzed a time-series of samples from three full-scale mesophilic anaerobic digesters treating wastewater sludge in Singapore to gain ecological insights into the AD microbial community structure, dynamics, and potential functions.

METHODS

Digesters, and sample collection

Three full-scale anaerobic digesters located in a water reclamation plant in Singapore were monitored for a period of 9-months. During this time, samples were collected from each digester on average at a 10-days interval. Collected samples were flash frozen in liquid nitrogen and were kept at -80°C prior to nucleic acids extractions.

DNA extraction and sequencing

DNA extraction was conducted using MPBio FastDNA™ SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA). Extracted DNA was sequenced on an Illumina HiSeq2500 sequencing machine in paired end mode with a read length of 250x250 bp. In total, 6.2 B reads were generated in the DNA sequencing runs. Minimum and maximum sequencing depth of 37 and 153M reads were obtained with a mean sequencing depth of 70M reads/sample.

Genome-centric metagenomics analysis

The analysis of sequencing data was conducted on SCELSE high performance computing cluster (HPC). Fastqc v0.11.3 (Andrews, 2010) was used to quality control all of the samples before and after trimming. Trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014) was used to trim low-quality reads as well as sequencing adapters. Then, SPAdes v3.13 (Bankevich et al., 2012) with kmer sizes of 21,33, 55 and 77 was used to assemble reads into contigs. The assembled contigs were then binned into genome bins by Metabat2 v2.12.1 (Kang, Froula, Egan, & Wang, 2015). These genome bins are known as metagenome assembled genomes (MAGs). The recovered MAGs were quality assessed and dereplicated by dRep v2.3.2 (Olm, Brown, Brooks, & Banfield, 2017). The quality of the recovered MAGs was analysed based on minimum information about metagenome assembled genome (MIMAG) criteria (Bowers et al., 2017). Only the MAGs satisfying at least the MIMAG medium-quality criteria were selected for further analysis. Taxonomical classification of the selected MAGs was performed by GTDB-Tk v1.1.0 (Chaumeil, Mussig, Hugenholtz, & Parks, 2020). High quality (HQ) MAGs were gene annotated by Prokka v1.13 (Seemann, 2014) followed by blasting against KEGG database release 93.0 (Kanehisa & Goto, 2000) by kofamscan v1.2.0 (Aramaki et al., 2020) and KEGG decoder v1.1 (Graham, Heidelberg, & Tully, 2018). The functional annotation was also conducted by EggNOG-mapper using the CAZy database. Relative abundances were calculated using CoverM.

RESULTS AND DISCUSSION

Using genome-centric metagenomics pipeline, we recovered a sum of 14,236 metagenome assembled genomes (MAGs). More than half of the recovered MAGs were of medium quality or better according to MIMAG criteria. The taxonomical classification of these MAGs revealed that 12.9%, 37.4%, and 77.1% of them belong to a novel family, genus, and species, respectively. Calculation of relative abundances shows that the community at phylum level is dominated by members of *Actinobacteriota*

phylum followed by *Proteobacteria* and *Chloroflexota* phyla providing a diverse range of hydrolytic functions. The archaeal members of the community were not as diverse as the bacterial members. *Halobacterota* and *Euryarchaeota* were the only archaeal phyla present in these digesters. The average relative abundances in three studied digesters at phylum level as well as the relative abundance of the recovered MAGs are depicted in Figure 1. The functional analysis of the recovered MAGs showed that despite low diversity in archaeal MAGs, three complete methanogenesis pathways, namely, methanogenesis via CO_2, Acetate, and trimethylamine were present in the community.

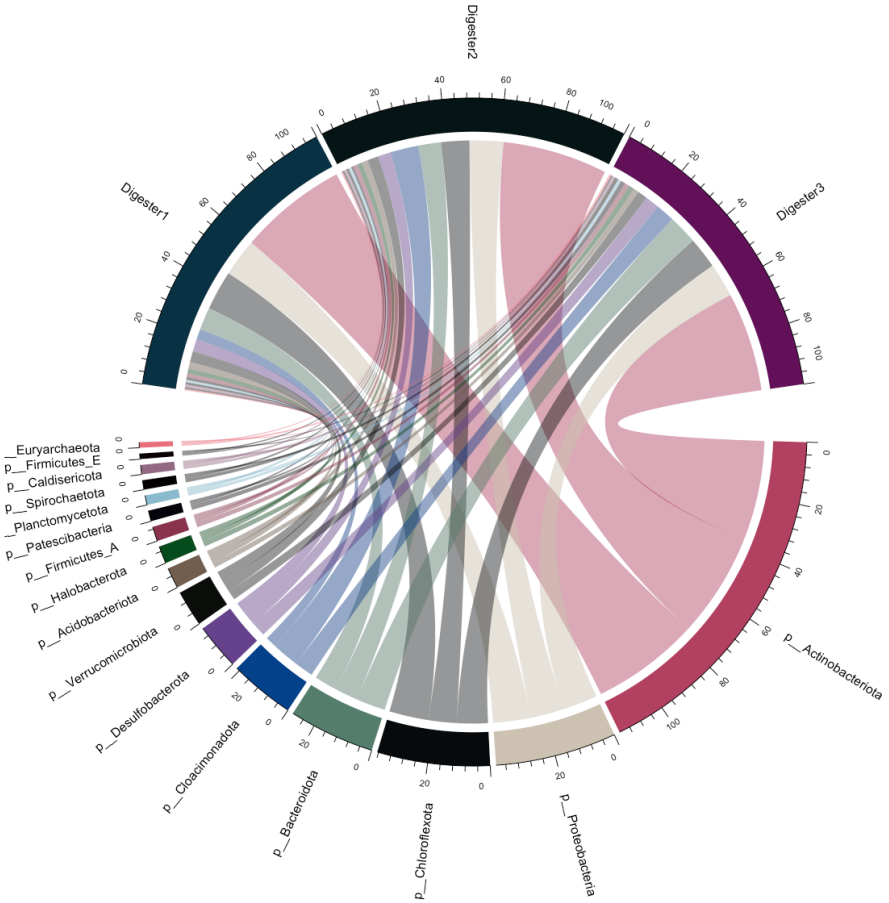

Figure 1 The structure of the microbial community in the studied digesters (top side) and the abundance of the recovered MAGs from the analysed metagenomics samples (bottom side) at phylum level.

CONCLUSIONS

Exploring a time-series of samples from three replicate anaerobic digesters in Singapore using genome-centric metagenomics, we obtained a catalogue of 166 MIMAG high-quality MAGs and analyzed their potential functions. Analyzing potential functions of these MAGs revealed a specialization of bacterial species in the hydrolysis stage of AD. In addition, a wide range of pathways were found to be at work in full-scale anaerobic digesters producing methane from sewage sludge. The novel ecological insights obtained in this work enhance our understanding of AD.

REFERENCES

1. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
2. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, *36*(7), 2251–2252.
3. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Prjibelski, A. D. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell

sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
4. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*.
5. Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., … Eloe-Fadrosh, E. A. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), 725–731.
6. Campanaro, S., Treu, L., Kougias, P. G., De Francisci, D., Valle, G., & Angelidaki, I. (2016). Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. *Biotechnology for Biofuels*, *9*(1), 26.
7. Campanaro, S., Treu, L., Kougias, P. G., Zhu, X., & Angelidaki, I. (2018). Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Scientific Reports*, *8*(1), 1926.
8. Campanaro, S., Treu, L., Rodriguez-R, L. M., Kovalovszki, A., Ziels, R. M., Maus, I., … Luo, G. (2020). New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnology for Biofuels*, *13*(1), 1–18.
9. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database*. Oxford University Press.
10. Graham, E. D., Heidelberg, J. F., & Tully, B. J. (2018). Potential for primary productivity in a globally-distributed bacterial phototroph. *The ISME Journal*, *12*(7), 1861–1866.
11. Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.
12. Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*.
13. Neshat, S. A., Mohammadi, M., Najafpour, G. D., & Lahijani, P. (2017). Anaerobic co-digestion of animal manures and lignocellulosic residues as a potent approach for sustainable biogas production. *Renewable and Sustainable Energy Reviews*, *79*.
14. Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864–2868.
15. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069.
16. Vanwonterghem, I., Jensen, P. D., Rabaey, K., & Tyson, G. W. (2016). Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. *Environmental Microbiology*, *18*(9), 3144–3158.
17. Zhu, X., Campanaro, S., Treu, L., Kougias, P. G., & Angelidaki, I. (2019). Novel ecological insights and functional roles during anaerobic digestion of saccharides unveiled by genome-centric metagenomics. *Water Research*, *151*, 271–279.