

36. OTDK Biology Section

**GENOME-CENTRIC METAGENOMICS AND
METATRANSCRIPTOMICS ANALYSIS OF A HUNGARIAN STATE-
OF-ART BIOGAS PLANT MICROBIOME**

Cheung Teur Teur Sally

MSc Biology

2022

University of Szeged

Supervisor:

Prof. Dr. Kornél Kovács

Dr. Roland Wirth

Table of content

Abbreviations:.....	3
1. Introduction.....	3
2. Literature Review.....	5
2.1 Biogas AD metabolic pathways and involved microbes.....	5
2.2 Chemical fermentation parameters and core microbiome	9
2.3 Multi-omics.....	11
2.4 DNA versus transcript abundance.....	13
3. Working hypothesis and goals	15
4. Materials and Methods.....	16
4.1 Sludge collection.....	16
4.2 Chemical fermentation parameters and experiments	16
4.3 Sequencing of metagenomics and metatranscriptomics	17
4.4 Metagenomics binning and assembly	17
4.5 Metatranscriptomics mapping.....	18
4.6 Metagenomics and metatranscriptomics visualization.....	18
5. Results and Discussions.....	18
5.1 Substrates and Chemical Fermentation Parameters	18
5.2 Genome-centric metagenome data.....	20
5.3 Metatranscriptomics: functional analysis and taxonomy roles	26
6. Conclusions.....	30
Acknowledgments	31
Bibliography:	32

Abbreviations:

AD – anaerobic digestion	GH – glycoside hydrolase
BMP – biological methane potential	GT – glycosyl transferase
C/N – carbon to nitrogen ratio	MAG – metagenome assembled genome
CAZymes – carbohydrate-active enzymes	Mcr – methyl-coenzyme M reductase
CBM – carbohydrate-binding motifs	MDM – microbial dark matter
CE – carbohydrate esterase	RPKM – reads per kilobase per million
ESBP – extracellular solute-binding proteins	SAO – syntrophic acetate oxidation
SZBP – Szeged Biogas Plant	TS – total solid
TAN – total ammonia nitrogen	VOA – volatile organic acid
TIC – total inorganic carbon	VS – volatile solid
TPM – transcript per million	WLP – Wood-Ljungdahl pathway

1. Introduction

Biogas can be utilized as a renewable energy carrier to generate electricity, transportation fuel or heat. The raw biogas is mainly composed of 50-75% methane, 25-50% carbon dioxide (CO₂) and 2-8% nitrogen (Li et al, 2019). The main component is methane (CH₄), an odorless, colorless, organic and highly flammable gas that is formed in nature within short time periods in swamps, rice fields and in geological time scale as “natural gas”. In 1776, Alessandro Volta discovered that the bubbly gas from the marshes is a product of anaerobic decomposition of the organic materials at the bottom of the swamplands (Cloy & Smith, 2022). Methane gained industrial interest and economic value when the excavation of “natural gas” and exploitation as a fossil energy carrier developed rapidly in the 21st century. The disadvantage of methane is that it is a greenhouse gas, thus it contributes to global warming and climate change. In this respect CH₄ is more dangerous than CO₂, which is the main product upon burning methane or other fossil fuels, because CH₄ is 25 times more potent in trapping the heat inside Earth’s atmosphere. The strong potency of methane to trap heat is by absorbing efficiently infrared radiation that would otherwise be reflected into interstellar space. When the light energy from the Sun reaches the Earth, part of the energy is absorbed by the Earth’s crust. A portion of the incoming light lose energy upon the impact, the rest is innately redirected, which has lower energy content, hence the reflected light is enriched in the infrared components. Infrared light is absorbed by heteroatom containing compounds, such as CH₄, CO₂ and many others. The

energy is thus trapped in the atmosphere and causes the alarming increase in global temperature. In fact, most methane emissions come from the natural gas, landfills, agriculture, coal mines and oil operations (U.S. Energy Information Administration, 2022). One of the methods for waste valorization is to produce biogas by converting organic materials collectively called biomass. A major difference between the excavated, fossil “natural gas” CH_4 and CH_4 made from biomass decomposition is that the latter is formed from renewable biomass, hence it does not generate net greenhouse gas emission and therefore does not contribute to the dangerous climate change. The CH_4 molecule from fermentation of biomass in engineered systems called anaerobic degradation (AD) is the same as the fossil one, but one is harmless the other is harmful for our environment (Awasthi et al, 2022). The leftover biomass (digestate) is nutrient-rich and has various functions like fertilizers for crop yields. Meanwhile, this agriculture yield also generates biowaste that is recycled (biomass) to feed the biogas reactors with substrates. The wide range of substrates not only comes from agricultural waste but also is produced from domestic animal and household wastes. The biogas production cycle completes carbon neutrality and contributes to sustainable development in the expanding global demand for clean energy and a circular bioeconomy.

The most abundant type of biogas substrate is agricultural waste and by-products. The major global agricultural wastes are 353.34 Mt (megatonne) wheat straw, 180.73 Mt sugarcane bagasse, 128.02 Mt corn stover and 731.3 Mt rice straw (Sarkar et al, 2012). Europe produces 950 Mt of biomass annually (Saleem, 2022). Lignocellulosic materials are the complex biomass making up the cell walls in plants and the largest source for generation biogas. The first-generation biomass comes from food crops. This type of crop utilization caused the controversial debate of food versus fuel. The problem is that diverting crops to biofuel production could detriment food security and increase land conversion to habitats destructing soil quality (Naik et al, 2010). Thus, second-generation biomass was introduced as a solution for non-edible, energy crops to subside the food security issue. Energy crops are typically grown on marginal lands to avoid competition for land and water with food crops (Singh et al, 2020). Third-generation biomass utilizes microalgae because the organisms can grow under non-demanding conditions like brackish & saline water and processing factories can be set on non-arable land. The most studied eukaryotic microalgae are *Chlamydomonas reinhardtii* and *Chlorella vulgaris* spp. Fourth-generation biomass is genetically modified microalgae with modified metabolism to speed up the rate of reaction and biofuel production (Goria et al, 2022).

Co-digestion of animal manure with plant biomass is also widely used for animal waste management. SZBP (analyzed reactor in this research) is a mesophilic, wet reactor that is fed

with pig manure, sweet sorghum and maize silage (both are 1st generation biomasses and high polysaccharide content). The ideal mesophilic fermenter temperature ranges from 35°C to 37°C for the microbiota to perform AD (Fikri, 2019). The optimal biomass C/N ratio is from 20 to 35 (Dadaser-Celik et al, 2016). Pig manure holds excessive nitrogen and limited carbon content due to its protein content. It also contains the main degrading microbes. Nitrogen is a pollutant that has three stable forms in nature: organic nitrogen, nitrate and ammonium. They can pollute the water bodies, soil, and underground drinking water. Besides, agricultural biomass has a rate-limiting degradation of lignocellulose (Gaworski, 2017), the reason for the aforementioned agricultural biomass generation. Co-digestion can balance to the optimal C/N.

The AD reactor contains a complex microbiome, which breaks down the biomass to produce biogas, a mixture of CH₄ and CO₂. AD is a type of microbial resource management, therefore it is important to acquire knowledge from microbial communities to harness their unique abilities for stable operation, prevent system failures (Carballa et al, 2015) and for development of future sustainable technology practices (Read et al, 2011). Since numerous microbes are uncultivable (Solden et al, 2016), the accessible high throughput sequencing is necessary using genetic material, i.e. DNA and RNA extracted from the AD content. Metagenomics and metatranscriptomics tools can provide taxonomic and functional pictures of the entire microbial community and allow the discovery of unknown microbial species. This multi-omics approach has been carried out in this study, in order to understand the important features of the anaerobic microbiome working in a typical mesophilic AD reactor.

2. Literature Review

2.1 Biogas AD metabolic pathways and involved microbes

The biological degradation of biomass to biogas during AD is divided into four overlapping phases: hydrolysis, acidogenesis, acetogenesis and methanogenesis. The biochemical processes of the microbes are linked together in that the byproduct of one phase is the substrate for the next phase. The phases are simultaneous processes and their analysis is complex.

Hydrolysis is the breaking down of complex large biopolymers to simpler, smaller water-soluble compounds. The main components in biomass are carbohydrates, proteins and lipids. The polysaccharides are hydrolyzed to produce oligosaccharides, monomeric sugars and a variety of aromatic compounds, proteins yield peptides and amino acids. The acidogenic bacteria convert amino acids, glucose and lipids to organic acids, CO₂ and H₂. During acetogenesis, organic compounds are converted to volatile fatty acids (VFAs), CO₂ and H₂

(Nwokolo et al, 2020). The final phase of methanogenesis uses these substrates to produce methane and carbon dioxide.

Bacteroidetes and Firmicutes are the most abundant bacterial groups across biogas reactors (Wirth et al, 2021; Campanaro et al 2016 & 2020; Kampmann et al, 2012). Bacteroidetes play an important role in degrading proteins from manure and polysaccharides from plant substrates. Firmicutes primarily digest polysaccharides from plant biomass (Kampmann et al, 2012). In co-fermentation AD, Bacteroidetes have a relatively lower abundance (Gullert et al, 2016). Hydrolysis is a limiting factor for biogas production. A variety of enzymes and enzyme complexes, called carbohydrate-active enzymes participate in this important process, a brief review of CAZymes is presented in the next section.

2.1.1 Carbohydrate-active enzymes are crucial during hydrolysis

Functional analysis of transcriptome data is an approach to show microbes' pathways in the hydrolysis or synthesis phase. Carbohydrate-active enzymes (CAZymes) are important catalyzers that act on carbohydrates. CAZymes are gathered in CAZy database (<http://www.cazy.org>) in the following categories: glycoside hydrolase (GH), glycosyl transferase (GT), carbohydrate-binding module (CBM) and carbohydrate esterase (CE).

GHs catalyze the hydrolysis of glycosidic bonds between carbohydrate residues. The enzyme group is conserved in prokaryotes with exo- or endo-acting specificity. CBMs are non-catalytic proteins that bind to soluble carbohydrates (Ezeilo et al, 2017). CBMs co-occur with GHs. The multienzyme complex requires the transportation and utilization of carbohydrates as energy sources within the bacterial cell. Bacterial extracellular solute-binding protein (ESBP) utilize ABC transporters to move solutes through the cytoplasmic membrane (Tam & Saier, 1993). ESBP interacts more efficiently with C5 sugar substrates such as xylan (complex polysaccharide) containing hemicellulose and pectin (Poudel et al, 2018).

GTs play important roles in glycosylation. GTases are conserved in prokaryotes to catalyze activated monosaccharides to form glycosidic linkages with proteins, lipids, or carbohydrates (Amin et al, 2021; Ezeilo et al, 2017). Among the N-linked and O-linked glycosylation, the enzyme-involved N-glycosylation uses polyisoprenoid alcohols, made up of maximum 25 isoprene units, for cell wall synthesis and cellular metabolisms (Skorupinska-Tudek et al, 2008). 5-carbon isoprene biosynthesis is the building block based on the mevalonate (MVA) pathway and 1-deoxy-D-xylulose-5-phosphate pathway. Isoprenoid production is a conserved pathway among Bacteria and Archaea domains. Interestingly,

archaeal membranes are made up of isoprenoid chains and glycerol, linked by ether bonds. Archaea commonly utilizes MVA pathway to synthesize cellular membrane that is essential for cell growth.

Species from Bacteroidetes, Actinobacteria, Firmicutes, Proteobacteria, Acidobacteria, Planctomycetes predominate the GH families and are responsible for polysaccharide degradation in biogas microbial community (Geng et al, 2021). CAZy database is part of the Pfam database (Mistry et al, 2020). The CAZymes analysis in the current study provides more insights into the rate-limiting step of hydrolysis (Wirth et al, 2021) and places the carbohydrate hydrolytic potential of the microbial community in the process of biomethane production to exceedingly high rank.

2.1.2 *Methanogen metabolisms and syntrophic relationships*

There are three major molecular pathways of methanogenesis, and four classes of methane-producing archaea endowed with the ability of performing one or more of these methanogenesis pathways. The methanogens are grouped as hydrogenotrophic, methylotrophic, acetoclastic and mixotrophic.

The only enzyme that is present in all types of methanogenic molecular routes is methyl-coenzyme M reductase (Mcr) encoded with highly conserved gene cluster of *mcrBDCGA*. The enzyme catalyzes the last step in methanogenesis by reducing the methyl group to methane (Thauer et al, 2008). Mcr can be activated with coenzyme 430 (Chen et al, 2020) and assembled by McrD and McrC respectively (Gendron & Allen, 2022). [4Fe-4S] cluster is used for protein and electron transfer coupling, substrate activation, and binding (Kirschning, 2022). Hydrogenotrophic methanogens use H₂ (electron donor) for CO₂ reduction to produce CH₄. They belong in the orders Methanobacteriales and Methanomicrobiales. H₂ is oxidized by soluble F₄₂₀-reducing hydrogenase, Frh, and soluble F₄₂₀-nonreducing dehydrogenase, Mvh, with heterodisulfide reductase complex, SH. Flavin-based electron bifurcation reduces ferredoxin and heterodisulfide of coenzyme-M and coenzyme-B (CoM-S-S-CoB) closely participate in the process. Subunits α (McrA), β (McrB) and γ (McrG) catalyze the compound's formation. Reduction of ferredoxin takes place in the first step of methanogenesis. CO₂ is reduced to a cofactor-bound formyl group. The methyl group is transferred to CoM-SH and reduction of CoM with CoB-SH recycle heterodisulfide in the last step of methanogenesis. CO₂ is used in the biosynthesis of cells, reduced ferredoxin is required in the process.

Some methanogens use energy-converting hydrogenase like Eha, which catalyzes sodium

motive force-driven ferredoxin reduction and hydrogen. Membrane-bound tetrahydromethanopterin S-methyltransferase (Mtr) translocates the sodium ions across the membrane and leads to the buildup of sodium motive force. It can be used by ATP synthase as the energy source (Kurth et al, 2020). Hydrogenotrophic methanogens' energy conservation takes place exclusively in methyl transfer reactions.

Methylotrophic methanogens utilize small, methylated carbon compounds such as methanol and methylated amines as substrates. There are three components in the enzyme systems. Substrate-specific methyltransferase converts methyl groups to corrinoid protein. The second methyltransferase transform three-quarter of methyl groups to CH₄ and one-quarter to CO₂. This is used to generate reducing equivalents for heterodisulfide as electron acceptor. Energy conservation only happens in membrane-bound electron transport chain. The membrane-bound methyltransferase catalyzes a reverse reaction by dissipating the sodium motive force (Kurth et al, 2020).

Aceticlastic methanogens use primarily or exclusively acetate produced during acidogenesis. Acetate enters the cell through the acetate transporter and activates acetyl-CoA to generate carbonyl group to lead to CO₂ and CH₄ generation by conversion via the acetyl-CoA decarboxylase/synthase. Compared to hydrogenotrophic methanogens, aceticlastic methanogens also uses Mtr for energy conservation but there are more sodium and hydrogen ions translocated per round. Additionally, aceticlastic methanogens require initial ATP investment for activation of acetyl-CoA metabolism (Kurth et al, 2020).

Methanogens form syntrophic relationship with fermenting bacteria that produce H₂, CO₂, acetate and other C-1 compounds (Kampmann et al, 2012). Acetogens (e.g. Clostridia) or syntrophic acetate oxidation (SAO) produce H₂ for the hydrogenotrophic methanogens. H₂-activating membrane-associated hydrogenases (i.e. in Methanomicrobiales) underline the importance for a highly efficient interspecies H₂ transfer between these syntrophic partners. H₂ metabolism is delicately balanced to maintain low H₂ concentration, which favors CH₄ formation. Moreover, an adequate supply of acetate supports both the growth of aceticlastic methanogens and hydrogenotrophic methanogens with the help of syntrophic acetate oxidizers (Wirth et al, 2012). Mixotrophic methanogens can utilize several pathways to generate methane. Members of the Methanosarcina genus belong in this mixotrophic methanogen group, converting acetate to methane (Nishio & Nakashimada, 2013) and form syntrophic relationship with other microbes. Various methanogens are present in biogas reactors (Wirth et al, 2021; Campanaro et al 2016 & 2020). Instead of competition between methanogens, they can live together in the biogas producing microbial communities as they utilize several pathways to

digest chemically dissimilar substrate molecules.

2.2 Chemical fermentation parameters and core microbiome

2.2.1. Chemical fermentation parameters

The most important chemical AD parameters are pH, volatile organic acids (VOAs), total inorganic carbon (TIC), total ammonia nitrogen (TAN), total solid (TS), volatile solid (VS) and biological methane potential (BMP).

In mesophilic (30-40°C) reactor, optimal pH is between 6.5 to 8.0 and methanogens prefer slightly alkaline environment. (Vijayaraj & Varatharajulu, 2015). pH is sensitive towards TAN, temperature, volatile fatty acids, etc., that ultimately affects the growth of biogas microbes.

Specific VOAs and TIC are among the crucial AD fermentation indicators. VOAs is the ratio of acetic acid to propionic acid during biogas production. TIC is the buffer capacity and carbonate carbon concentration. VOA/TIC ratio serves as an early warning indicator for acidification, which could lead to system failure if persisted for an extended time period. The range of optimal VOAs/TIC from agricultural-industrial wastes is from 0.1 to 0.3 (Szilágyi et al, 2021).

The total ammonia nitrogen (TAN) is an important indicator of ammonia-nitrogen level as high concentration can also lead to AD inhibition. It also affects the stability of buffer capacity and pH, which is essential for microbial growth. pH, VOAs and TAN are the main factors concerning AD inhibition (Capson-Tojo et al, 2020). AD instability leads to an increase of volatile fatty acids that decreases pH. Hence, the free ammonia reduction results in continuous fermentation to a certain extent. However, ammonia level reduces methane production (Wang, 2016) and pig manure contains high ammonia level. Free ammonia can passively diffuse into the cells and cause potassium deficiency and proton imbalance (Rui et al, 2015). AD inhibitions were reported at TAN levels between 1.5-3 g/L (Rossi et al, 2022). Although methanogens are the most sensitive community members to AD parameter perturbation, hydrogenotrophic methanogens are one of the most resilient microbes to high TAN concentrations (Capson-Tojo et al, 2020). TIC and TAN can affect the pH of the reactor. TAN's acceptable range in AD is from 1.5 to 2.0 g/L in pig manure and maize silage reactor, which has stable operating conditions (Wirth et al, 2021).

TS or dry matter content is related to the feedstock composition (Capson-Tojo et al, 2020). TS content exhibits the best digestibility performance between 7.4-9.2% (Budiyo et al, 2010). VS is used to measure the organic matter in the digester. Co-digestion containing pig manure and food waste with 80-98% VS is an excellent substrate in AD (Wang et al, 2020).

Biological methane potential (BMP) is the biodegradability of substrates in AD based on VS. BMP test systems include a positive control to check the performance of the inoculum. The optimal inoculum-to-substrate ratio (ISR), constant temperature with gentle mixing is required in addition to replacing air in the headspace of the reactor with inert gas prior to incubation. BMP test shows the maximal biogas potential of the substrate, therefore actual biogas yield of an industrial, large scale, continuously fed biogas plant would be lower than indicated by the BMP test. The chemical energy of the substrate is divided between methane production and microbial biomass growth in 9 to 1 ratio. Moreover, BMP usually cannot detect acute toxicity of inhibitors, especially at high ISR. Usually VS-based ISR is between 1 to 4 depending on the substrate degradability (Koch et al, 2020). In spite of these limitations, BMP is a powerful and useful tool to determine the expectable maximum biogas yield.

Monitoring of chemical fermentation parameters is useful for the biogas industry. The daily observations are relatively simple and inexpensive. Nevertheless, the stable operation and undisturbed functional activity of the microbial community requires at least periodical checking of the microbiome abundance, activity and networking. The members of the microbial community are the key, most important hard-working actors. Their wellbeing and functional activity is fundamental for the safe and efficient operation of the AD reactors (Kirkegaard et al, 2017; Rui et al 2015).

2.2.2 The core microbiome

Core microbiome refers to the set of indispensable essential microbial taxa in the biogas reactor. They are associated with functional and genomic tasks in microbial ecology. The concept of core microbiome rapidly developed as the number of taxonomic investigations accumulated and allowed comparative studies. The evaluated thresholds are based on microbial occurrence, abundance or occurrence-abundance (Neu et al, 2021). Empirical analyses of human microbiome (Neu et al, 2021) built the framework for further applications such as AD microbiomes (Rui et al, 2015; Wirth et al, 2012) using abundance cutoff values and arbitrary occurrence.

Early studies commonly used 16S rRNA gene amplicon sequencing to identify microbial taxonomy. In some cases, operational taxonomic units or sequence clusters of shared nucleotide identification were established. In AD studies, core Archaea were members of the genera *Methanosaeta* (now renamed as *Methanothrix*), *Methanoculleus*, *Methanospirillum* and core Bacteria included *Clostridium*, *Synergistetes*, *Sedimentibacter*, Firmicutes, Bacteroidetes.

The core microbiomes are resistant in AD perturbations (Rui et al, 2015; Kirkegaard et al, 2017; Rivière et al, 2009). More insights into the keystone population were understood but those approaches contained limitations. 16S rRNA gene amplicon sequencing uses primers that would introduce bias. The primers bind to regions 100% conserved across all bacteria (Gupta et al, 2019). The 16S rRNA approach cannot identify unknown species due to reference dependency (Stolze et al, 2016). Different metrics used to quantify core microbiome in practice are also susceptible to bias, such as low abundance but functionally important microbes, arbitrary cutoff abundance or occurrence values, impact of sequencing depth and sampling coverage, etc (Neu et al, 2021). Therefore, developing standardized metrics is crucial at testing specific hypotheses about the ecology and function of the core microbiome.

To overcome the restriction in practice and quantification approach, novel bioinformatics tools and core microbiome metrics have been introduced. Machine learning guided genome-centric metagenomics and metatranscriptomics analysis are reference independent approach and not limited in identifying unknown species. Classifying core microbiome with 100% occurrence is the most common approach.

2.3 Multi-omics

Mapping out the complete blueprint of the biogas microbiome network is difficult, but metagenomics can widen the knowledge in this field. The limitations on microbial cultivation technologies challenge our understanding of the microbial world. Only 0.1-1% of microorganisms present in the environment can be cultivated at present (Solden et al, 2016). The remaining uncultivable and undetectable representatives are in the *Candidatus* phyla. Most of the Earth's yet-uncultivable microbes are like the dark matter in the universe, namely microbial dark matter (MDM) (Solden et al, 2016). Moreover, cultivated microorganisms may function differently in pure cultures than within the biogas community, which poses additional problem when we want to understand what they do for the community and how (Palkova, 2004; Dickson et al, 2014; Altowayti et al, 2020; Anguita-Maeso et al, 2020).

Scientists have addressed the missing information from uncultured microorganisms from the benefits of genomic sequencing technologies in the past decade. Previously, shotgun sequencing and assembly resulted in separate genomes from shared features (i.e. *k*-mer frequencies, coverage homology). Single-cell genomics can complement this ambiguity but suffers from amplification biases in genomic coverage, leading to less complete genome (Hedlund et al, 2014). On top of that, the utility of big data of MDM triggered several

challenges (Jiao et al, 2020). 16S rDNA or 16S rRNA amplicon sequencing are implemented to answer and study the microbial community, core microbiome pathways and ecology of food chains (Rivière et al, 2009; Straub et al 2020). Other approaches include using high-throughput sequencing reads for aligning DNA reads/protein sequencing (comparing against databases), read-based metagenomics with *k*-mer frequencies (a type of sequence features) and universal SCGs (a species classification method) to read depths (Sun et al, 2021).

Amplicon sequencing, high-throughput sequencing and read-based metagenomics cannot identify the unknown species in the microbiome because these approaches heavily depend on reference databases (Stolze et al, 2016; Calusinska et al, 2018). Sequencing data is directly aligned with reference genomes, species-specific sequences or marker gene sets for taxonomy annotations. Regardless of these strategies, the reference database is incomplete and requires high-quality genomes. The sequencing and assembly restrictions need to be resolved if the goal is to analyze and understand any microbiome, including those working and surviving in the AD reactors.

To overcome all the above limitations, a set of highly advanced bioinformatics tools is used to investigate the microbial community. In recent years, strategies and algorithms have been developed to facilitate independence from reference databases. Assembly-based MAGs are assembled based on a group of scaffolds with similar features from metagenomes. The features are tetranucleotide frequencies, complimentary marker genes, abundances, taxonomic alignments and coverage to resemble the microbial genome (Yang et al, 2021). Single-copy genes are typically used to quantify MAGs with marker genes for a specific genome because they should occur once. However, some of the limitations remain in the uneven distribution of the universal marker genes and their presence is low (<10%). Lineage reference has a more refined estimation of quantifying microbial genome completeness and contamination for automated assignments in large-scale genome sequencing. Substantial genome completeness is $\leq 70\text{-}90\%$ and medium contamination is $5\text{-}\leq 10\%$. The limitation of using this criterion is for lineages that are undergoing genome reduction to quantify genome quality (Parks et al, 2015). Only the MAGs with high genome completeness and low contamination are selected for downstream analysis. Yet Minimum Information about a Single Amplified Genome (MIMAG) and Minimum Information about a Metagenome-Assembled Genome (MISAG) were also initiated to assembly quality and estimation of genome completeness and contamination (Bowers et al, 2017). Based on these standards, several studies have proven MAGs to be useful (Stolze et al, 2016; Campanaro et al 2016 & 2020). 134 publicly available metagenome-assembled genomes have been derived from 1,635 MAGs (criteria: $>50\%$ C, $\leq 5\%$

contamination; nearly half of genomes had >90% C). These studies established the first biogas database for assembly and binning without reliance on a reference database. The identified genomes were shown to have distinct functional properties from former studies and complex microbial networks. This approach frees the need for culture dependence and MAGs perfected the complex microbiome annotations and functions in the digester. Improving the binning accuracy of MAGs remains a continuous task.

There are two common types of reference-independent pipelines for microbial community analysis. First, traditional unsupervised clustering depends on co-abundance (reads mapping coverage to the contig) and sequence features (i.e. *k*-mer frequency). Most softwares rely on this clustering method (Bair, 2013). Second, semi-supervised clustering takes advantage of must-link and cannot-link constraints joining contig pairs. Must-link constraint requires two contigs *must* place in the same cluster. Cannot-link constraint calls two contigs *must not* be in the same cluster (Bair, 2013). Information map and weighed *k*-mers contributes to siamese neural network (machine learning), which helps then to process the distributions to obtain fine-grade MAGs (Pan et al, 2022).

Scientists are deciding whether one or combined techniques are more holistic to investigate microorganisms. Combining the formerly studied and used techniques with bioinformatics tools are important in exploring the roles of uncultivable microbes in the population of MDM. Scientists are trying to close the knowledge gap of unknown or partially known microbes in the biogas digesters. They are also shrinking as the number of deep metagenomics studies increase.

These powerful methods can decipher and discover the phylogeny and metabolism of the biogas microbial community (Nobu et al, 2015; Hassa et al, 2018). When the known microbes are detected, the remaining microbes are resolved to newly discovered microorganisms through a set of analysis approach. The semi-supervised constraint with a machine learning system could enhance the binning performance of contigs to obtain MAGs (Pan et al, 2022). Studying the core microbiome relies on this analysis sets to identify key players, including unknown microbes.

2.4 DNA versus transcript abundance

Genome-centric metagenomics-based metatranscriptomics analysis provides more insights into the real-life microbial activity relative to the snapshot group picture of the microbial community participants, answering the question: who is there? In previous research,

metagenomics analysis has not been paralleled with metatranscriptomics (Wirth et al, 2021; Campanaro et al 2016 & 2020; Kampmann et al, 2012). The metagenome data (genomic DNA sequencing) can provide taxonomy and gene information. The metatranscriptome data (mRNA sequencing) are reads that are mapped back to the metagenome data to assemble the actual transcript and functional activity. The trivial assumption is that DNA abundance should be parallel with RNA abundance (De Vrieze et al, 2016). However, this is a widely accepted misconception. By way of explanation, microbes of high DNA abundance may not display high functional transcript activity and vice versa.

mRNA abundance or transcript activity is expressed in both transcript per million (TPM) and read per kilobase per million (RPKM) values, which is calculated differently from DNA abundance based on RPKM denomination. Connection of the taxonomy with gene activity gives a more accurate relationship between the microbiome abundance and active genes & functions.

Methanogens have a low abundance but produce unproportionally high methane yield in stable operating biogas reactors. This is due to two reasons, transcription apparatus and the ability of 'cheating' thermodynamics. The vital Mcr present in all methanogens has evidence of a linear relationship between methanogenesis and *mcrA* transcript to gene ratio. Transcript/gene ratio indicates physiological activity rather than gene abundance. 18-fold increase of *mcrA* transcript activity is parallel with the highest observed methanogenesis in pure culture form (Freitag & Prosser, 2009). However, methanogens may have different transcriptional activity in different environments. Moreover, methane production is dependent on temperature, ammonia, acetate, partial H₂ pressure, etc (Capson-Tojo et al, 2020). Apart from the transcription activity, methanogens also show different thermodynamics thresholds compared to other microorganisms. Hydrogenotrophic methanogens have energy yields as low as -10 kJ/mol (Hoehler et al, 2001), in situ almost free energy requirements. Since many microorganisms are MDM, the current known minimum requirement to sustain life is -20 kJ/mol (Schink, 1997). Methanogens overcame this thermodynamic energy barrier, for instance, linking the first and last step of hydrogenotrophic pathway to regenerate terminal electron donor compounds to overcome energy unfavorable reactions (Buan, 2018). This circular respiration is known as Wolfe cycle. With minimal energy conservation, the hydrogenotrophic methanogens can transcribe the Mcr gene and produce methane. Based on the connection between metagenome and metatranscriptome data, population (including MDM) and activity studies are identified in the biogas microbe's community.

3. Working hypothesis and goals

As discussed above, there remain limitations and challenges in fully understanding the microbiome of biogas plants. The AD community contains uncultivable biogas microbes (MDM). Understanding their taxonomy, functions and networking is a complex process. The problems include the changes and synonyms of taxonomy and nomenclatures in microorganisms. There is also a dependency on reference databases in sequencing analysis, which requires high-quality genomes. The goal is to analyze the microbial composition and function that is crucial for biogas production. Machine learning-guided metagenomics and metatranscriptomics can investigate the DNA abundance with transcript activities, genes, metabolic pathways, key enzymes and core microbiomes.

With the availability of reference-independent MAGs in genome-centric metagenomics complemented with metatranscriptomics data, we can understand both the identity and roles of microorganisms based on lineage-specific markers. We also have to keep in mind that microbes are interrelated in the food web of the biogas reactor.

The core microbiome is crucial for biogas production as they are the key players in metabolic pathways and resist perturbations in altering environments. The core microbiome includes microbes that are high in occurrence. There are microorganisms that are rare in presence but may be important in the microbial community of biogas. This study not only identifies core microbiome taxonomy and function but also relates core microbiome with industrial chemical fermentation parameters for a comprehensive view of biogas production. The correlation could be indicators in situations of process imbalance. The core microbiome could resist environmental changes but is a gradual adaptation process. The tolerance is developed in time longer than the feeding rate. Precision microbial management is based on the correlation and indicators of chemical fermentation parameters. The MDMs are important for their roles in the overall process of biogas production.

In summary, the thesis is intended to reveal the understanding of microbial composition in the biogas plant with genomics and transcriptomics data using the semi-supervised machine learning system. The microbial composition includes the core microbiome of taxonomy, functional identification and correlation with chemical fermentation parameters. Key enzymes, dominant phyla and metabolic pathways should be investigated. The methanogens should be further analyzed for their production pathways and activities in the AD community. MDMs are addressed with the availability of genomics and transcriptomics data.

4. Materials and Methods

4.1 Sludge collection

The microbial analysis was based on the fermentation effluent collected from Zöldforrás Ltd. in Szeged, Hungary. SZBP sludge samples were collected at four seasonal intervals during one-year period. Sampling took place in October and January of 2020, and April and July of 2021. Three samples were collected at each interval for biological replicates. The input substrate contained 55% pig manure, 30% maize silage, and 15% sweet sorghum. SZBP is a biogas plant next to the pig farm, so the pig manure is directly transferred to the fermenter. The maize silage is produced by putting chopped maize, lactic acid bacteria, urea, salt, water, and mineral mixes into pit. After 45 to 60 days of ensilaging, the maize silage is ready to be used by the biogas microbiome.

4.2 Chemical fermentation parameters and experiments

4.2.1 VOAs and TIC

15mL of sludge in a Falcon tube was centrifuged for 10 min at 30,000 rpm/min. To prepare the sample for analysis, 5 g was diluted with distilled water to make 20 g. Then 0.1 N H_2SO_4 is used for titration. A Pronova FOS/TAC 2000 Version 812 09.2008 automatic titrator was used to test the VOAs, TIC (Pronova, Berlin, Germany).

4.2.2 TAN

Added 14.5 mL of distilled water to biogas sludge for dilution. The Merck Spectroquant Ammonium test (1.00683.0001) (Merck, Billerica, USA) determined the total ammonia nitrogen content (TAN = ammonium ions and dissolved ammonia).

4.2.3 TS and VS

To determine TS, substrates were kept at 105°C for an overnight period for weight constant. VS are the total organics in the sludge, which is the amount of material lost during incinerated. The remaining contents were therefore put into the incinerator at 550°C for 2 hours for the final mass.

4.2.4 BMP

BMP estimates the methane potential and biodegradability of biomass (Szilágyi et al, 2021). Tests were carried out in 160 mL reactor vessels (Wheaton glass serum bottle, Z114014 Aldrich) containing 60 mL fluid phase at mesophilic temperature ($37\pm0.5^\circ\text{C}$). One set of measurements determined the baseline methane production of inoculum, i.e. filtered AD biomass and another was fed with the α -cellulose (positive control, C8002 Sigma-Aldrich) at

an inoculum VS ratio of VDI 4630 standard (2 substrate: 1 inoculum; Verein Deutscher Ingenieure 4630, 2006).

4.3 Sequencing of metagenomics and metatranscriptomics

The three sampling replicas (2 mL aliquots) were used for DNA and RNA isolation with ZymoBIOMICS DNA/RNA miniprep kit (R2002, Zymo Research, Irvine, USA). After lysis with bead beating (bead size: 0.1 mm; Vortex Genie 2; beating speed: max.; time: 15 min), Zymo Research kit protocol was followed for DNA and RNA purification. The isolated DNA and RNA replicas were then pooled together for sequencing.

Illumina sequencing platform recommendations were followed (Illumina Inc., USA). NEBNext Ultra II Library Prep Kit was used prepared the genomic DNA samples for sequencing. Illumina sequencer used NextSeq High Output Kit v2 (300 cycles) to conduct paired-end metagenomics.

Metatranscriptomics RNA samples was prepared with paired-end libraries using Zymo-Seq RiboFree™ Total RNA Library Kit, which includes the universal rRNA depletion. Illumina NextSeq 550 sequencer mRNA (translated to cDNA) using NextSeq High Output Kit v2 (300 cycles). *bcl2fastq* software (v2.17.1.14, Illumina) was used to base-call the primary data.

4.4 Metagenomics binning and assembly

Fastp (v0.23.2, length required: 150) filtered the raw sequences and then quality checked with FastQC (v0.11.8). The filtered sequences were assembled by Megahit v1.2.9 (Li et al, 2015) with 4 samples, minimum 1500 contig length (*k*-mer size was between the interval of 21-141 bases). The 4 samples assembled into contigs were used to reconstruct the metagenome-assembled genomes. Contig database was created using Anvi'o (v7 *hope*) (Eren et al, 2015).

The deep siamese neural network using advanced co-assembly of the binning workflow was combined with the semi-supervised approach for a machine learning-guided software. Semibin (v1.1.1) was used to reconstruct the genomes (Pan et al, 2022). dRep (v2.2.3) was used for quality filtration and dereplication of MAGs. Within dRep, CheckM (v1.1.6) programs were used (dereplicate: sa 0.95, S_algorithm fastANI, con 5, comp 10 (Olm et al, 2017; Park et al, 2015)).

Open reading frames (ORF) were identified by Prodigal and InterProScan v5.31-70 was

used to annotate the functional gene coding sequences on Pfam database (Finn et al, 2017). GTDBTk (v1.0.2) was used to assign the microbial taxonomy with the Genome Taxonomic Database (GTDB: Release 202) (Parks et al, 2021). CoverM (v0.6.1, method: rpkm) calculated the MAGs' reads per kilobase per million values (RPKM) (<https://github.com/wwood/CoverM>). To build the phylogenomics tree using a set of single-copy core genes (SCGs) of 120 bacteria and 122 archaea, GTDBTk (classify_wf) and IQTree (v2.2.0.3) with the number of bootstraps at 1000, maximum iteration at 1000 and stopping rule at 1000 build were used.

4.5 Metatranscriptomics mapping

Extraction of ORF from MAGs (bedtools: getfasta) resulted in MAG-specific gene calls. These were combined to create Salmon index (parameter of keepDuplicates, <https://github.com/COMBINE-lab/salmon>). Transcript per million (TPM) of the read counts were generated by Salmon (v1.8.0) with GC bias correction (gcBias) and quasi-mapping.

4.6 Metagenomics and metatranscriptomics visualization

The parameters graph is made through R (R Core Team, 2021), and Microsoft Excel (Microsoft Corporation, 2018). R packages, namely phyloseq (McMurdie & Holmes, 2013), ggplot2 (Wickham, 2016) and MicroViz (Barnett et al, 2021), were used to visualize RKPM, TPM, core microbiome heatmap correlations, KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) and Pfam heatmaps. Visualization of core microbiome percentage package was R package VennDiagram (Chen & Boutros, 2011). The taxonomy and phylogenetic tree were made to graph using Interactive Tree of Life, iTOL v5 (Letunic & Bork, 2021). Pfam (contains CAZymes and gene databases) (Mistry et al, 2020) was visualized using Circos (Krzywinski et al, 2009).

5. Results and Discussions

5.1 Substrates and Chemical Fermentation Parameters

The considerations for choosing biogas plant for this study included digesters operating successfully for 5 years with essentially the same substrate composition. This guarantees an adapted and stable microbial community in the reactor. During the experiment, we monitored the anaerobic reactor (SZBP) in every 3 months. The chemical fermentation parameters were measured four times (Fig. 1). In the experiment, the VOAs were 2.82 g/L and TIC is 17.20 g/L

(Fig. 1A, C, D, standard deviations are indicated in Fig. 1), which gives the average value of 0.16 VOAs/TIC. The ratio is within the optimal range of 0.1-0.3 (Szilágyi et al, 2021). In the observed values, July 2011 TIC has the lowest and halved buffer capacity. The average TAN concentration was 1.72 g/L (Fig. 1E), which is acceptable for AD (mentioned in Section 2.2). The pH of the reactor was 7.63 throughout the experimental period (Fig. 1A). The TS content was 6.03% (Fig. 1F), with minor fluctuations in TS in October and January. These are normal conditions as TS value may vary between 2.64-18.40% (Section 2.2). VS content in SZBP was 73.59% and stable.

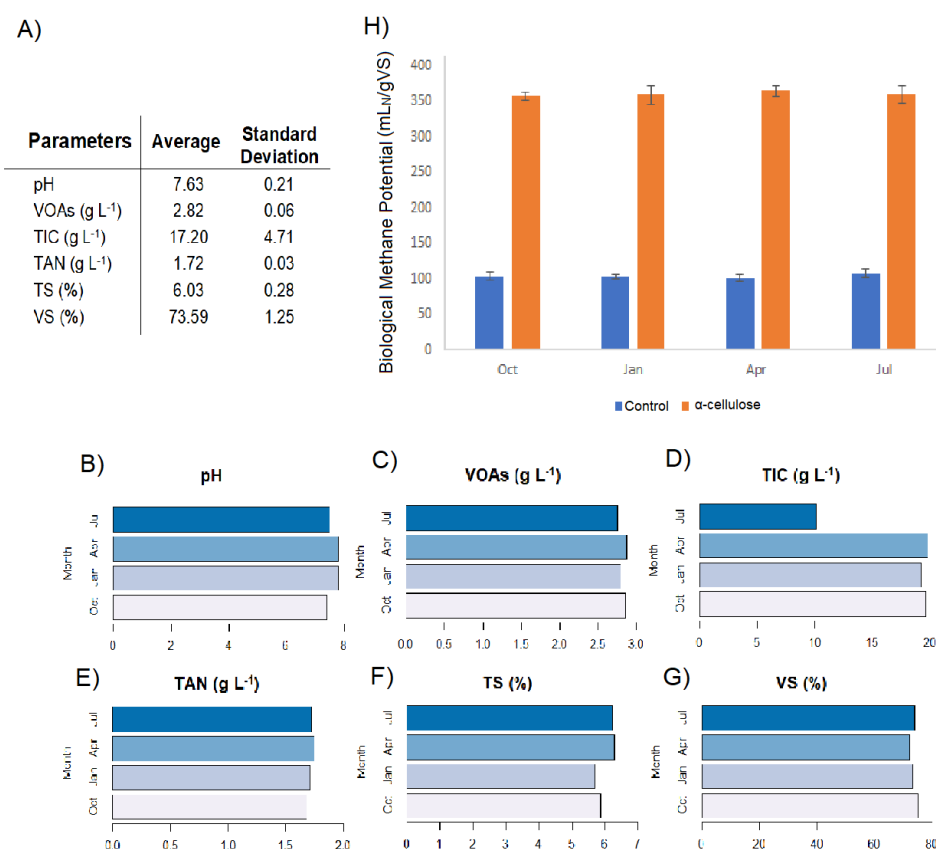


Fig. 1. Anaerobic digestion (AD) parameters of mesophilic agriculture biogas reactor. The bar plot colors represent chemical fermentation parameters samples retrieved in October 2020, January 2021, April 2021 and July 2021. (A) Average and standard deviation of the four months AD parameters. (B)-(G) pH, volatile organic acid (VOAs, g L⁻¹), total inorganic carbonate buffer (TIC, g L⁻¹), total ammonia-nitrogen concentration (TAN, g L⁻¹), total solids (TS, %) and volatile solids (VS, %) of the monthly measurements. H) Biological Methane Potential (BMP, mL_N/gVS) of control group and α -cellulose group in the four sampling times.

The methane measurement showed that the control group produced 103.95 mL_N/gVS in October, 102.94 mL_N/gVS in January, 101.39 mL_N/gVS in April and 107.05 mL_N/gVS in July (Fig. 1H). The α -cellulose (standard substrate for positive control) group was determined with BMP at 357.15 mL_N/gVS in October, 359.18 mL_N/gVS in January, 365.03 mL_N/gVS in April and 359.72 mL_N/gVS in July (Fig. 1H). α -cellulose group produced similar methane yield in the 3-month intervals. The control group had 3-fold lower methane potential compared to α -

cellulose, which has been expected as in other study (Szilágyi et al, 2021).

5.2 Genome-centric metagenome data

There were 92.59 million filtered DNA sequences (Table 1). The filtered reads were assembled by Megahit and created 77,156 contigs (N50 = 7362). The binning procedure was performed by neural network-based binning process that resulted in 199 MAGs. 165 out of 199 MAGs passed the threshold (completeness $\geq 1\%$, contamination $< 5\%$), set to prevent low genome completion and high contamination microbes in the multi-omics analysis.

Table 1. DNA raw sequences and filtered sequences.

DNA	Total raw sequences	Total filtered sequences
Oct	18,818,954	15,679,826
Jan	18,575,390	16,445,728
Apr	26,354,002	24,709,486
Jul	38,689,062	35,746,324
Total	102,437,408	92,581,364

5.2.1 DNA and transcript abundances present dissimilar predominant microbes

The MAGs were compared to GTDB (Parks et al, 2021) for taxonomy annotation of known or partially known microbes and Bio-Gas Microbiome database (Campanaro et al, 2020) (mentioned in Section 2.3 & 4.4). The binning results showed 11 MAGs belonging in the domain Archaea and 154 MAGs in the domain Bacteria (Fig. 2A). The high microbiome similarity between the months also supported the uniformity of class abundances found in metagenomics data (Fig. 2B and E). The most abundant taxonomic groups in Bacteria domain was in the phylum Firmicutes, class Clostridia and Bacilli confirming previous studies using animal manure and maize silage substrates (Wirth et al, 2012; Kampmann et al, 2012). Next in the abundance line were phyla Bacteroidetes, Proteobacteria, and Actinobacteria (Kakuk et al, 2021) (Fig. 2B, E).

High transcript abundance of hydrogenotrophic methanogens, Methanomicrobia and Methanobacteria, were observed in the biogas reactor. Methanomicrobia also had high DNA abundance (Fig. 2B, E). The mixotrophic methanogens, i.e. Methanosarcina, had one of the highest DNA and transcript abundance (Fig. 2B, E). They can utilize acetate, H₂, methanol or methylated amines as substrates, which are produced by upstream microbes (mentioned in section 2.1.2). Despite the similarities in microbial composition using metagenomics data, there were also differences in their activity based on metatranscriptomics data. The principal component analysis was observed to have 5.1% DNA abundance (Fig. 2C) and 5.3% RNA

abundance differences between months (Fig. 2F). The PCA plot shows the small differences between months. It is noteworthy that Methanomicrobia had elevated gene activity in time and decreased DNA abundance (Fig. 2B, E). There was more transcript abundance of methanogens than shown in DNA abundance (Fig. 2B, E).

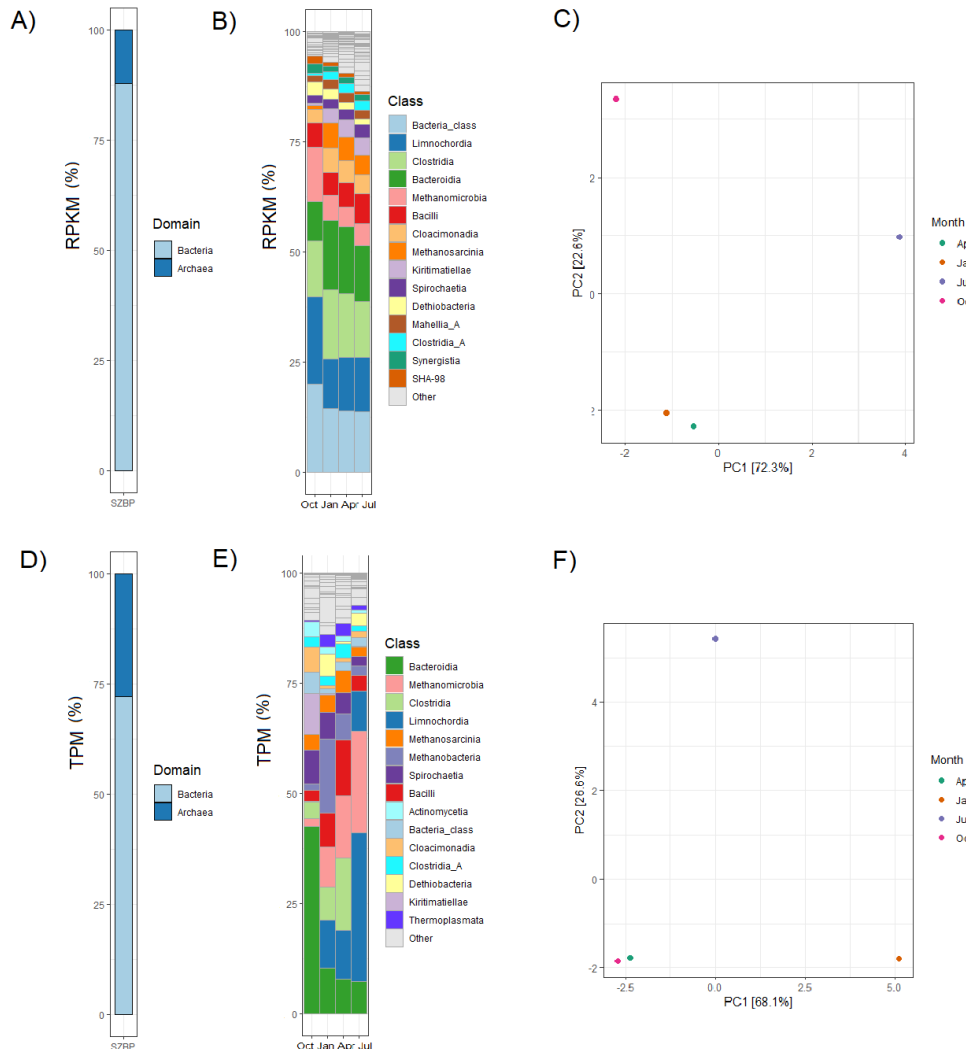


Fig. 2. Genome-centric results. (A) and (B) DNA domain level, (D) and (E) RNA of top 15 abundant phylum level taxonomic distribution of the microbiome. (C) DNA and (F) RNA principal component analysis of the microbial community of all the four-month sludge samples based on DNA and RNA. The colors represent the corresponding months.

Bacteroidia in October and Limnochordia in July showed the highest RNA abundance and one of the lowest DNA abundances in the four sampling time points (Fig. 2B, E). Bacteroidia has been commonly found in degradation of organic materials like plant materials (Wirth et al, 2020). Limnochordia were found to form syntrophic relationship (Section 2.1.2) with Methanosarcinia in 20 biogas plants (Calusinska et al, 2018). Actinomycetia, Clostridia, Bacilli, Limnochordia and others have important implications in plant materials degradation using cellulolytic enzymes (Lewin et al, 2016) (Section 2.1). These results suggest that DNA

abundance cannot reflect the total transcript activity of microbes.

The most significant result from Fig. 2 is that Archaea had low DNA abundance and showed at least two-fold increase in transcript activity. Methanomicrobia, Methanosarcina, Methanobacteria and Thermoplasmata were the most active Archaeal classes. Methanomicrobia was the second most active methanogen, but these methanogens were in the 5th position according to DNA abundance.

The class Thermoplasmata were present in the RNA abundance but not in the top 15 DNA based metagenome abundant taxa (Fig. 2B, E). Thermoplasmata contains mixotrophic methanogen using formaldehyde and acetate. The total methanogens activity is much higher than DNA presence. Therefore, the DNA and transcript abundance observations show different data. Metagenomics should be complemented with metatranscriptomics to analyze microbial activity in biogas AD.

5.2.2 Microbial tree of life in biogas plant and the MDM

The taxonomic assignment showed that the most common phylum is Firmicutes A with 42 species, Bacteroidota with 24 species, and Firmicutes G with 13 species (Fig. 3), which is in accordance to other studies (Campanaro et al, 2016). There were 66 species (40% of the total) found in Firmicutes A, B, D, F, and G phylum out of 165 total species in the microbiome. The species in the phylum Firmicutes are extremely important in metabolic pathways such as fatty acid utilization, carbohydrates' degradation, Wood-Ljungdahl Pathway (WLP) or SAO. Bacteroidetes were less abundant, they participated primarily in protein and polysaccharide degradation (see Section 2.1). The higher Firmicutes to Bacteroidetes ratio is due to the substrate mix composition fed to the biogas reactor.

Biogas MDM was observed by the normalized RPKM (DNA abundance) and TPM (transcript abundance) data (Fig. 3). There was a total of 24 potential MDMs marked with “_species” ending in the figure. For further MDM analysis, the recommendation by Parks et al (2015) for $\geq 90\%$ completeness and $\leq 5\%$ contamination was followed. All 24 MAGs had $\leq 5\%$ contamination. In genome completeness, 4 MAGs were near ($\geq 90\%$), 1 MAG was substantial ($\geq 70-90\%$), 6 MAGs were moderate ($\geq 50-70\%$) and 14 MAGs were partial ($< 50\%$). The 4 MAGs genera with near genome completeness and highest metabolism (KEGG) activity were in the following descending order: *Methanosarcina* (MAG_165, 98.69% completeness) for methanogenesis, *DMER64* (MAG_363, 100% completeness) for citrate cycle, *Miniphocibacter* (MAG_88, 95.73% completeness) for glycolysis and *Fermentimonas*

(MAG_666, 90.97% completeness) for glycolysis. The possibly unknown glycolysis microbes both had medium-low level DNA and transcript abundance from Firmicutes_A and Bacteroidota phyla (Fig. 3, Section 2.1).

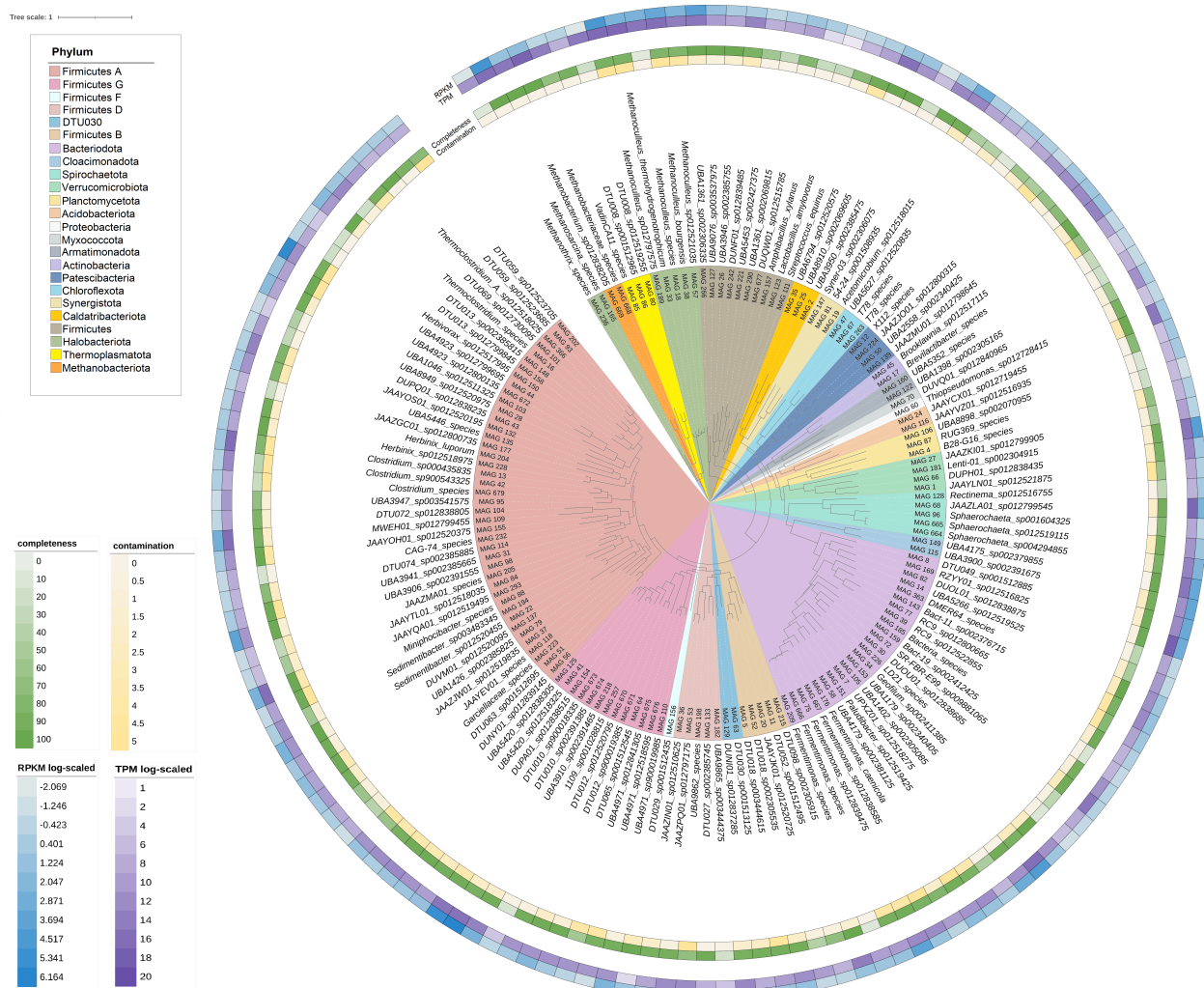


Fig. 3 Phylogenetic assignments of 165 MAGs. The tree was built using iTol with taxonomic phylum species level identification, log-scaled RPKM, and TPM values, DNA completeness and DNA contamination in different color gradients.

With the availability of Pfam data in this study, *DMER64_species* has the highest gene activity for large subunit ribosomal protein for protein & peptide bond synthesis (Schmeing, 2013) and peroxiredoxin as an antioxidant for cytoprotection (Knoops et al, 2016). *DMER64_species* predicted stable cellular metabolism, continuous gene activity, which protects ammonia-stress from methanogens in the biogas microbiome. For instance, *DMER64* genus in Bacteroidetes phylum is a partially known microbe. It was proposed to have syntrophic relationship with methanogens through magnetite-mediated direct interspecies electron transfer (DIET) through anaerobic batch tests. In DIET, electrons are transferred from one cell to another by conductive pili, specific cytochromes or conductive materials, like magnetite. This mechanism differs from transfer via reduced molecules like hydrogen or

formate. Magnetite can relieve ammonia inhibitory effects during VFA degradation to aid methane production (Lee et al, 2019). However, no physiological activities are known.

The Archaea domain includes phyla Halobacteriota, Thermoplasmatota, and Methanobacteriota (Fig. 3). At lower taxonomic levels: hydrogenotrophic methanogens *Methanoculleus*, *VadinCA11*, *Methanobacteriaceae*, and *Methanobacterium spp.*, acetoclastic methanogen *Methanotherix sp.*, methylotrophic methanogen *Methanomassiliicoccaceae DTU008 sp.*, and mixotrophic methnogen *Methanosarcina sp.* have been detected. Ammonia-tolerant species included *Methanoculleus*, *VadinCA11* and *Methanosarcina spp.* (Ting et al, 2020). Hydrogenotrophic methanogens usually have the highest phylogenetic diversity (Wirth et al, 2012; Campanaro et al, 2016). The presence of ammonia tolerant methanogens and ammonia shielding syntrophy *DMER64_species* indicate that the community was getting ready to combat toxic ammonia levels (Fig. 1A).

The *Methansarcina_species* (MAG_165), mentioned earlier, is a potential unknown species with one of the highest transcript abundances (Fig. 3). It was the only mixotrophic methanogen in the recovered data. *Methanoculleus bourgensis* and *Methanoculleus thermohydrogenotrophicum* are common hydrogenotrophic methanogens found in the digester that share similar RPKM and TPM (Fig. 3). The two species were reported to share common essential genes and methanogenic pathways (Kougias et al, 2017), which allows resistance to environmental perturbations. The results confirmed that when metagenomics data, which is used to analyze microbial identity, are complemented with metatranscriptomics data, microbial activity may be associated to the species.

5.2.3 Positive correlations between chemical fermentation parameters and core microbiome

The core microbiome analysis and correlation with chemical fermentation parameters are based on 165 MAGs identified in metagenomics analysis. The core microbiome is the essential community in the AD processes. 86.7%, i.e. 143 species of the total 165 microbial species belonged in the core microbiome present during the entire experimental period (Fig. 4A). The correlation of microorganisms and chemical fermentation parameters, i.e. relationship between process indicators, and microbiomes tell us how the composite substrate feed determined the composition and functional activity of the adapted AD microbial community. The composition relatedness between specific members of the core microbiome is shown in relationship, the “what operational parameter influences primarily the abundance and activity of which member(s) of the community” question can be answered. Bacteroidota species

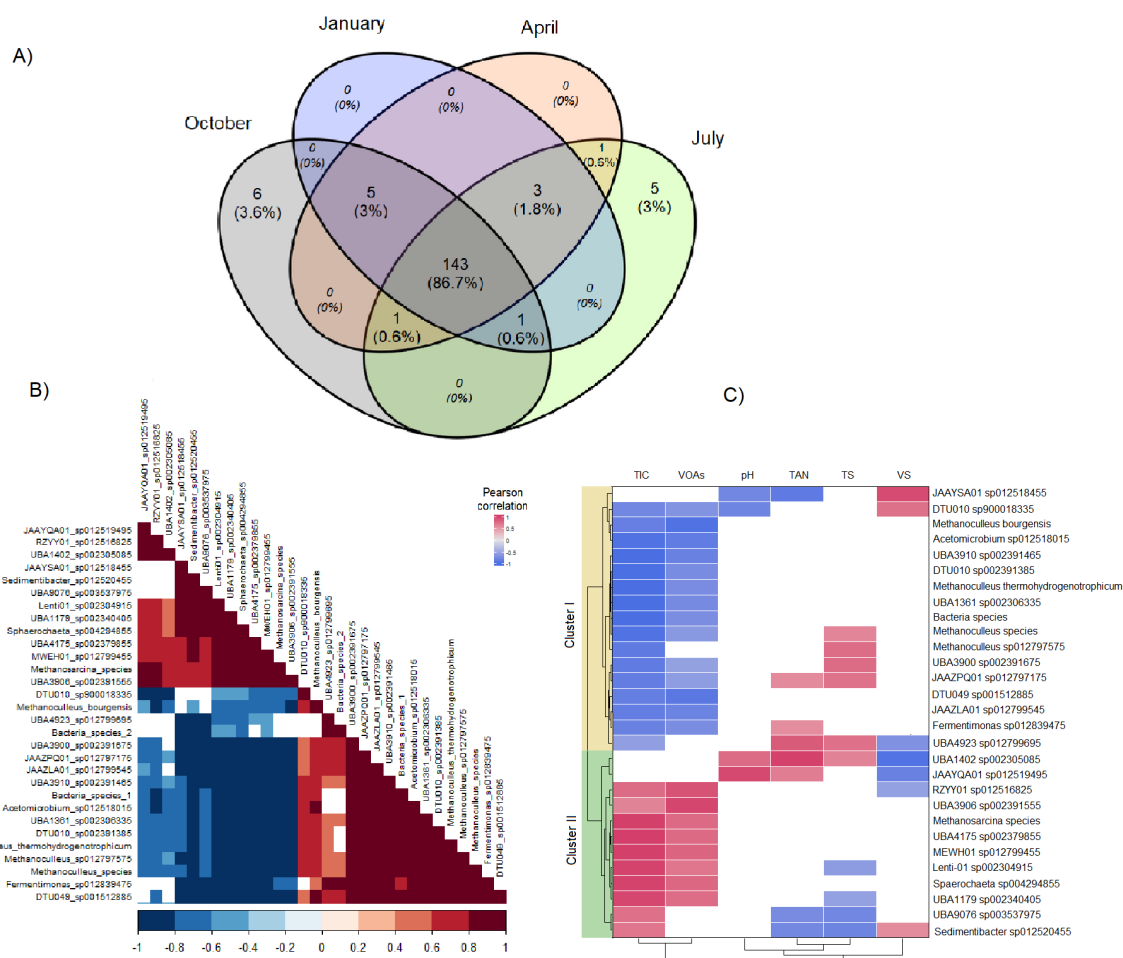


Fig. 4 Core microbiome analysis A) Venn diagram of core microbiome, month-specific, and month-related microbial community. B) Correlation groups of core microbiomes. C) Core microbiome correlation with chemical fermentation parameters in clusters. Colors in B) and C) mean positive and negative Pearson correlations, respectively. White boxes indicate no specific correlation between the chemical process parameters and the given taxon in the community.

correlate with chemical fermentation parameters TIC and VOAs (Fig. 4C). Depending on the positive or negative correlations between the chemical process parameters and members of the microbiota, two clusters could be distinguished. Remarkably, *Methanoculleus spp.* are in Cluster I, and *Methanosarcina_species* (refer to Section 5.2.2) is in Cluster II. The clusters of methanogens with bacterial groups may point to specific syntrophic relations as discussed in Section 2.1.2.

Methanoculleus spp. have been reported being sensitive to the TS chemical process parameter. Previous report showed that the increase of TS contents from 5% to 20% led to the decrease of *Methanoculleus* spp. from 7.63% to 2.91% genus-level (Yi et al, 2015). Our results did not corroborate this observation (Fig. 4C). The SZBP reactor maintained an average TS = 6.03%, near the optimal range of 7.4-9.2% (Budiyono et al, 2010) and the *Methanoculleus* spp. did not correlate neither positively nor negatively with TS. *Methanoculleus bourgensis* has

higher gene content and larger genome than *Candidatus M. thermohydrogenotrophicum* in the MDM (Kougias et al, 2017). Fig 3 shows a higher transcript activity by *M. bourgensis* than *M. thermohydrogenotrophicum* based on TPM value.

Cluster II contains mainly *Methanosarcina* species and *Firmicutes A* species (Fig. 4C). *Methanosarcina*_species correlated with TIC and VOAs. The mixotroph *Methanosarcina* sp. is tolerant to the toxic environment by forming irregular cell clumps (Conklin et al, 2006) for their successful survival. They can utilize both hydrogen and acetate pathways from the available organic acids. *Firmicutes A* spp. were mostly associated with TIC and VOAs. Some *Firmicutes* spp. were also related to other parameters, namely pH, TAN, TS, and VS. *Firmicutes* are the main polysaccharide degraders and the most diverse phyla in AD. TIC buffer capacity is related to pH and ammonia toxic concentrations. Clusters of positively correlated microbes are susceptible towards specific chemical fermentation parameters.

5.3 Metatranscriptomics: functional analysis and taxonomy roles

52.61 million filtered RNA sequences (Table 2) were mapped back using Salmon to MAGs and calculate TPM values for transcriptomics data (Section 2.4). Microbial functional roles could be revealed through the data. Results in the following are based on metatranscriptomics data.

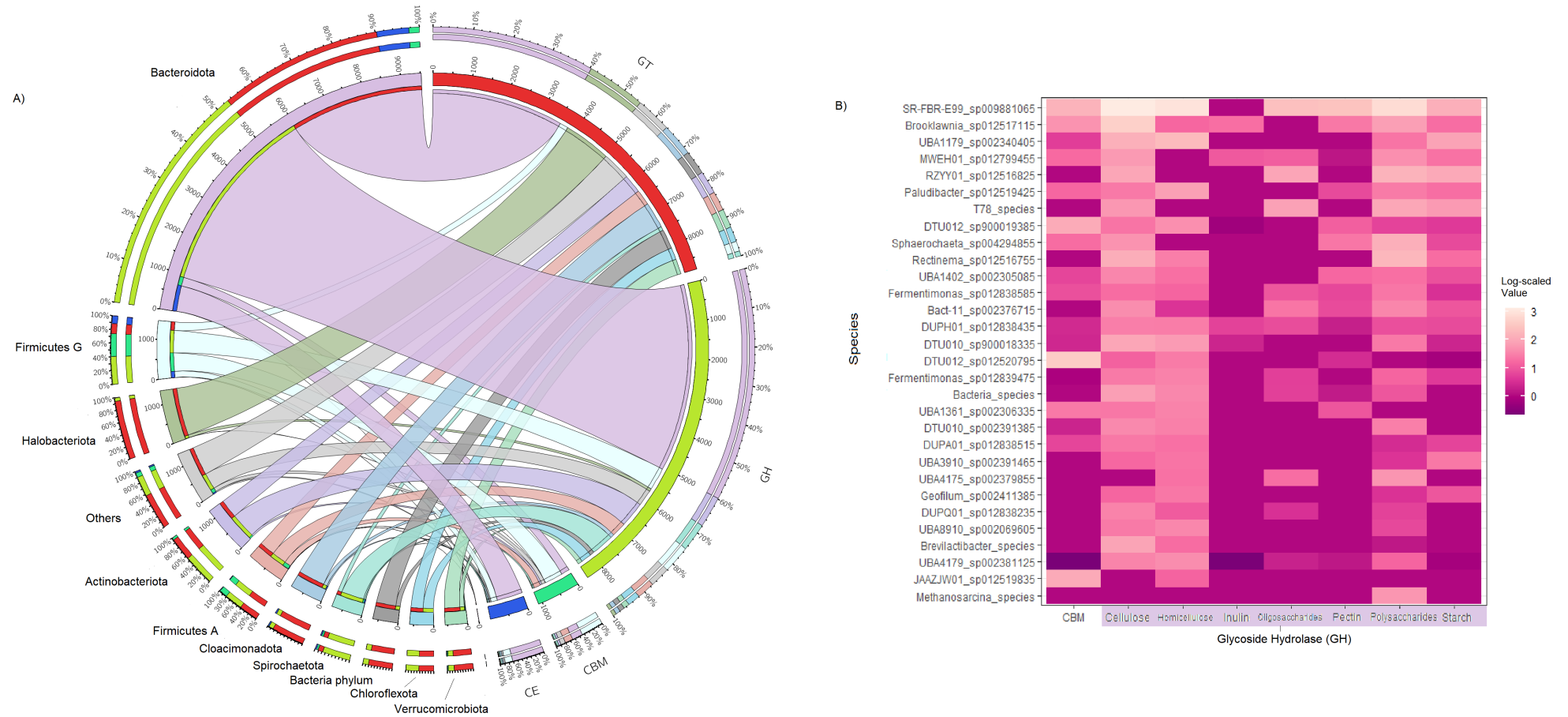
Table 2. RNA raw sequences and filtered sequences.

RNA	Total raw sequences	Total filtered sequences
Oct	16,416,766	1,383,216
Jan	17,884,478	20,642,438
Apr	23,538,936	19,242,282
Jul	17,356,234	11,343,300
Total	75,196,414	52,611,236

5.3.1 Carbohydrate-active enzymes affect downstream microbes

The metatranscriptomics data provided insight into the rate-limiting step of hydrolysis (Section 2.1.1). The top 30 most CAZymes active phyla are compiled in Fig. 5A. The most active phyla are Bacteroidota, Firmicutes G, Halobacteriota, Actinobacteriota, Firmicutes A, Cloacimonadota, Spirochaetota, Chloroflexota, Verrucomicrobiota etc. Bacteriota from Bacteroidota had the highest gene activity, followed by Methanomicrobia, Methanosarcina, Methanobacteria from Halobacteriota, and Limnochordia from Firmicutes G (Fig. 2E). GH and

Fig. 5 CAZymes analysis based on RNA sequencing A) Main microbiome phyla in CAZymes activity. B) Heatmap of carbohydrates-binding module (CBM) and glycoside hydrolase (GH).



GT were present in all top 30 phyla. GT was shown to be the most active enzyme group in CAZymes family. Bacteria domain Bacteroidota had 40% and Archaea domain Halobacteriota had 12.5% of GT activity. 5-carbon isoprene biosynthesis is also a top 11 metabolism pathway (Fig. 6A). The methanogens require isoprenoids for cell growth. N-glycosylation occurs at the external surface of the archaea cell surface, which explains high GT activity in archaeal Halobacteriota. Identical to archaea, the N-glycosylation of bacteria occurs across cell membranes (refer to Section 2.1.1). GT initiates the transport of nucleotide-activated sugars across membranes, resulting in the highest CAZymes activity of GT across all phyla (Fig. 5A).

The second most active CAZymes is GH. 60% Bacteroidota, 9% Actinobacteria, and 31% remaining phyla of GH activity (Fig. 5A). GH has diverse degrading functions of cellulose, hemicellulose, inulin, oligosaccharides, pectin, polysaccharide, and starch (Fig. 5B). Two species of Actinobacteria were identified in the taxonomy with high genome completeness and gene activity (Fig. 3). *Brooklawnia_sp012517116* and *Brevilactibacter_species* showed activity in both GH and CBM (Fig. 5B). Bacteroidota *SR-FBR-E99_sp009881065* has top CMB and GH activity. Cellulose, hemicellulose (contains xylan), and polysaccharides are the mostly GH-degraded substrates (Fig. 5B). Bacteroidota and Actinobacteria species had much higher CBM and GH activity compared to *Methanosarcina_species*, whose main function is methanogenesis (Fig. 5B). Hydrolysis of substrates was actively performed by Bacteroidota and Actinobacteria in the SZBP reactor.

5.3.2 Essential genes and metabolism in biogas production

The microbial metabolic pathways (KEGG) and gene activity (Pfam) were constructed from metatranscriptomics data. Interestingly, methanogenesis had the highest metabolism over the other housekeeping metabolism (Fig. 6A). This was also confirmed by the transcript activity (Fig. 2E) and taxonomy diversity (Fig. 3) of specific methanogens. Activity of upstream metabolism such as glycolysis and pentose phosphate pathways (PPP) have lower metabolic activity than methanogenesis (Fig. 6A). Microbes utilizing these upstream pathways are tightly connected. Glucose is metabolized by the glycolytic pathway in PPP to produce NADPH, which is the reducing coenzyme in enzymatic systems. In the experiments, high Firmicutes diversity and activity has been observed (Fig. 3) but methanogenesis remained the most active metabolism pathway (Section 5.2.2). Moreover, methanogens are archaea with membranes made up of isoprenoid and glycerol (Section 2.1.1). *C5_isoprenoid_biosynthesis* is observed to be highly active (Fig. 6A).

Pfam analyzes functional genes of microbial data that participate in metabolism pathways (KEGG). *ABC_transporter* and *Bacterial_extracellular_solute-binding_proteins* (Fig. 6B) are known to move solute across the membrane in the process of polysaccharide degradation (Section 2.1.1), which is observed in high degradation activity (Fig. 5B).

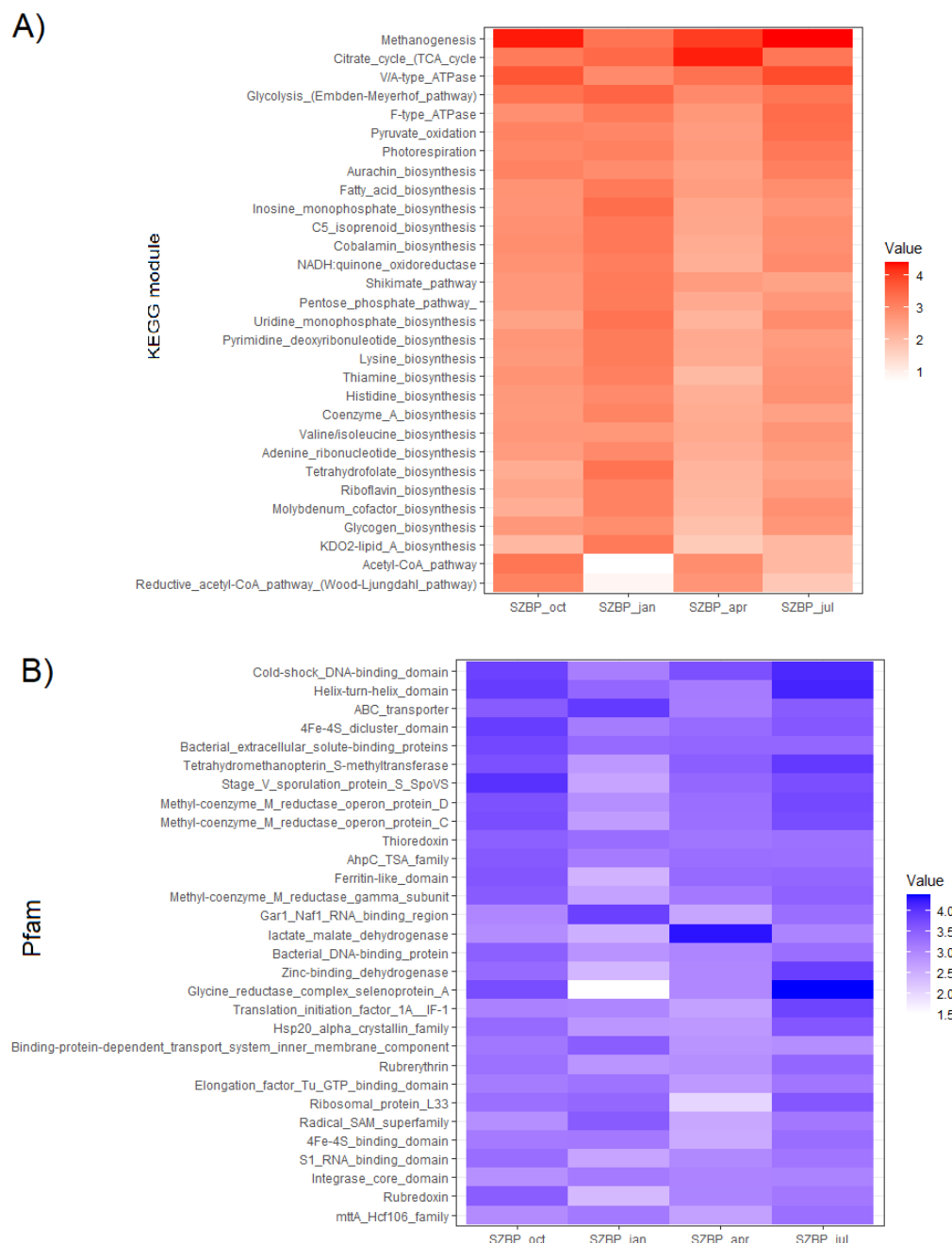


Fig. 6 Functional analysis of the SZBP microbiome. A) Unknown and known proportion in KEGG modules and Pfam data. B) The heat map of top 30 metabolic pathway KEGG modules. C) Top 30 Pfam transcribed genes in SZBP.

One of the most important enzymes in methanogenesis is Mcr (Section 2.1.2) and its high activity was detected as *Methyl-coenzyme_M_reductase_gamma_subunit*, *Methyl-coenzyme_M_reductase_operon_protein_D&C* (Fig. 6B). Mtr is responsible for sodium

transport (Section 2.1.2). *Tetrahydromethanopterin_S-methyltransferase* was also detected (Fig. 6B). *4Fe-4S_dicluster domain* and *4Fe-4S_binding_domain* had high transcript activity (Fig. 6B) and they work with Mcr simultaneously in methanogenesis (Section 2.1.2). Remarkably, the methanogens had the highest metabolism and gene activity to produce methane than any bacterial function present in the AD although some of them had high DNA abundance and diversity.

6. Conclusions

We can understand the taxonomy and functional roles of the microbes at details by employing the most up-to-date and precise bioinformatics approach known today, i.e. genome-centric metagenomics based metatranscriptomics analysis. The transcript data were mapped back to the metagenome data to verify transcript activity. This strategy offers novel horizon in understanding the relationship among members in a complex microbial community, which cannot be achieved by relying solely on DNA abundance data.

All chemical fermentation parameters were within the optimal ranges. Under these conditions, the control group, supplied with the standard substrate α -cellulose, showed 3-fold higher methane potential, indicating that the microbial community operated properly and reliably. Archaea had two-fold higher transcript abundance compared to DNA abundance. The outstanding transcript abundant Archaea classes were Methanomicrobia, Methanobacteria and Thermoplasmata. Methanogens with highest DNA and transcript abundances were Methanomicrobia and Methanosarcina.

40% of species were classified in the Firmicutes phylum, in line with previous studies. The evolution to increase the Firmicutes to Bacteroidetes ratio is a frequently observed phenomenon in biogas digesters. There were four potential MDM taxa found in our database, namely *Methanosarcina*, *DMER64*, *Miniphocibacter* and *Fermentimonas* spp. *DMER64_species* has been found to have large subunit ribosomal protein and peroxiredoxin transcript activity. It is noted, that *DMER64* genus has been proposed to have syntrophic relationship with methanogens (Lee et al, 2019). The possibly unknown *Methanosarcina_species* belonged in the core microbiome in Cluster II with Firmicutes A spp. *Methanosarcina_species* was found sensitive to TIC and VOAs, the important industrial chemical process indicators of appropriately operating digesters. Correlation analysis revealed that Cluster I included *Methanoculleus* spp., Firmicutes and Bacteroidota. SZBP TAN concentration was 1.72 g/L, which was a warning sign and some process inhibitions were reported between 1.5-3.0 g/L (Rossi et al, 2022). Bacteroidota spp. are sensitive to TIC and

VOAs, which means fluctuations of main chemical fermentation parameters could affect their population and activity. The correlation between core microbiome and chemical fermentation parameters uncovered a novel relationship indication in cases of process stability.

The core microbiome is resistant towards certain AD perturbations, but if chemical fermentation parameters fluctuate too far or beyond optimal ranges, process inhibition would occur. Therefore, calculating the correlation between core microbiome and chemical fermentation parameters is important.

The metatranscriptomics data provided insight into the rate-limiting step of hydrolysis. CAZymes are important because they are the enzymes produced in hydrolysis and affect the downstream microbes. Bacteroidota and Firmicutes can degrade polysaccharides, but other phyla are also capable for such activity like Actinobacteriota, Cloacimonadota, Spirochaetota, Chloroflexota, Verrucomicrobiota, etc. ABC transporter and ESBP were noted to be involved in polysaccharide degradation through Pfam analysis.

Methanogenesis showed the highest metabolic activity compared to other, e.g. housekeeping, pathways. This is an extremely remarkable novel feature of the AD community. One of the most important enzymes in methanogenesis is Mcr, which has been found as the most active gene: *Methyl-coenzyme_M_reductase_gamma_subunit*, *Methyl-coenzyme_M_reducatsse_ operon_protein_D&C*. *4Fe-4S_dicluster domain* and *4Fe-4S_binding_domain* work with Mcr to participate and facilitate methanogenesis. Mtr is responsible for sodium transport in energy conservation of hydrogenotrophic and acetoclastic methanogens, which is shown in the highly active *Tetrahydromethanopterin_S-methyltransferase*. The activity of methanogenic Archaea outcompeted Bacteria at transcript level, metabolism and gene activity.

The study of biogas microbiome is a complex process. Precision microbial community management is important for stable biogas operation in the industrial scale biogas production reactors together with correlation of chemical fermentation parameters with core microbiome. For academic basic research the exploration of the largely unknown “microbial dark matter” is an exciting scientific challenge with obvious projections for practical applications.

Acknowledgments

I would like to thank Prof. Dr. Kornél L. Kovács for supporting my research to join the exciting field of bioenergy research and development. I am also extremely grateful for Dr. Roland Wirth guiding me the way into bioinformatics, ‘omics approach, and for the time we

spend discussing the results and suggestions on the manuscript.

Bibliography:

1. Altowayti, WA., Almoalemi, H., Shahir, S., & Othman, N. (2020). Comparison of culture-independent and dependent approaches for identification of native arsenic-resistant bacteria and their potential use for arsenic bioremediation. *Ecotoxicology and Environmental Safety*, 205, 111267. <https://doi.org/10.1016/j.ecoenv.2020.111267>
2. Amin, K., Tranchimand, S., Benvegnu, T., Abdel-Razzak, Z., & Chamieh, H. (2021). Glycoside hydrolases and glycosyltransferases from hyperthermophilic archaea: Insights on their characteristics and applications in biotechnology. *Biomolecules*, 11(11), 1557. <https://doi.org/10.3390/biom11111557>
3. Anguita-Maeso, M., Olivares-García, C., Haro, C., Imperial, J., Navas-Cortés, et al. (2020). Culture-dependent and culture-independent characterization of the olive xylem microbiota: Effect of SAP Extraction Methods. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01708>
4. Awasthi, MK., Singh, E., Binod, P., Sindhu, R., Sarsaiya, S., Kumar, A., et al. (2022). Biotechnological strategies for bio-transforming biosolid into resources toward circular bio-economy: A review. *Renew. Sustain. Energy Rev.*; 156, 1–20. <https://doi.org/10.1016/j.rser.2021.111987>
5. Bair, E., (2013). Semi-supervised Clustering Methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(5), 349–361. <https://doi.org/10.1002/wics.1270>
6. Barnett, D., Arts, I., Peders, J. (2021). microViz: an R package for microbiome data visualization and statistics. *Journal of Open Source Software*, 6(63), 3201, <https://doi.org/10.21105/joss.03201>
7. Berghuis, BA., Yu, FB., Schulz, F., Blainey, P.C., Woyke, T., et al. (2018). Hydrogenotrophic methanogenesis in Archaeal Phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *BioRxiv*, <https://doi.org/10.1101/391417>
8. Bowers, RM., Kyrpides, NC., Stepanauskas, R., Harmon-Smith, M., Doud, D., et al. (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. <https://doi.org/10.1038/nbt.3893>
9. Buan, NR. (2018). Methanogens: Pushing the boundaries of biology. *Emerging Topics in Life Sciences*, 2(4), 629–646. <https://doi.org/10.1042/etls20180031>

10. Calusinska, M., Goux, X., Fossépré, M., Muller, E. E., Wilmes, P., et al. (2018). 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. *Biotechnology for Biofuels*, 11(1). <https://doi.org/10.1186/s13068-018-1195-8>
11. Campanaro, S., Treu, L., Kougias, PG., De Francisci, D., Valle, G., et al. (2016). Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel Binning Strategy. *Biotechnology for Biofuels*, 9(26). <https://doi.org/10.1186/s13068-016-0441-1>
12. Campanaro, S., Treu, L., Rodriguez-R, LM., Kovalovszki, A., Ziels, RM., et al. (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(25). <https://doi.org/10.21203/rs.2.18473/v2>
13. Capson-Tojo, G., Moscoviz, R., Astals, S., Robles, Á., & Steyer, J.-P. (2020). Unraveling the literature chaos around free ammonia inhibition in anaerobic digestion. *Renewable and Sustainable Energy Reviews*, 117, 109487. <https://doi.org/10.1016/j.rser.2019.109487>
14. Carballa, M., Regueiro, L., & Lema, JM. (2015). Microbial management of anaerobic digestion: Exploiting the microbiome-functionality Nexus. *Current Opinion in Biotechnology*, 33, 103–111. <https://doi.org/10.1016/j.copbio.2015.01.008>
15. Chen, H., & Boutros, PC. (2011). Venndiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12(1). <https://doi.org/10.1186/1471-2105-12-35>
16. Chen, H., Gan, Q., & Fan, C. (2020). Methyl-coenzyme M reductase and its post-translational modifications. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.578356>
17. Cloy, JM., & Smith, KA. (2022). Methane and nitrous oxide emissions from land. *Reference Module in Earth Systems and Environmental Sciences*. 145-153. <https://doi.org/10.1016/b978-0-12-822974-3.00023-9>
18. Dadaser-Celik, F., Azgin, ST., & Yildiz, YS. (2016). Optimization of solid content, carbon/nitrogen ratio and food/inoculum ratio for biogas production from food waste. *Waste Management & Research*, 34(12), 1241–1248. <https://doi.org/10.1177/0734242x16659922>
19. De Vrieze, J., Regueiro, L., Props, R., Vilchez-Vargas, R., Jáuregui, R., et al. (2016). Presence does not imply activity: DNA and RNA patterns differ in response to salt

- perturbation in anaerobic digestion. *Biotechnology for Biofuels*, 9(1). <https://doi.org/10.1186/s13068-016-0652-5>
20. Drula, E., Garron, ML., Dogan, S., Lombard, V., Henrissat, B., et al. (2021). The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Research*, 50(D1). <https://doi.org/10.1093/nar/gkab1045>
 21. Dickson, RP., Erb-Downward, JR., Prescott, HC., Martinez, FJ., Curtis, JL., et al. (2014). Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. *Journal of Clinical Microbiology*, 52(10), 3605–3613. <https://doi.org/10.1128/jcm.01028-14>
 22. Ezeilo, UR., Zakaria, IL., Huyop, F., & Wahab, RA. (2017). Enzymatic breakdown of lignocellulosic biomass: The role of glycosyl hydrolases and lytic polysaccharide monooxygenases. *Biotechnology & Biotechnological Equipment*, 1–16. <https://doi.org/10.1080/13102818.2017.1330124>
 23. Fikri Hamzah, MA., Md Jahim, J., & Mohamed Abdul, P. (2019). Comparative start-up between mesophilic and thermophilic for acidified palm oil mill effluent treatment. *IOP Conference Series: Earth and Environmental Science*, 268(1), 012028. <https://doi.org/10.1088/1755-1315/268/1/012028>
 24. Finn RD., Attwood TK., Babbitt PC., Bateman A., Bork P., et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research* 2017; **45**: D190–D199. <https://doi.org/10.1093/nar/gkw1107>
 25. Freitag, TE., & Prosser, JI. (2009). Correlation of methane production and functional gene transcriptional activity in a peat soil. *Applied and Environmental Microbiology*, 75(21), 6679–6687. <https://doi.org/10.1128/aem.01021-09>
 26. Gaworski, M., Jabłoński, S., Pawlaczyk-Graja, I., Ziewiecki, R., Rutkowski, et al. (2017). Enhancing biogas plant production using pig manure and corn silage by adding wheat straw processed with liquid hot water and steam explosion. *Biotechnology for Biofuels*, 10(1). <https://doi.org/10.1186/s13068-017-0922-x>
 27. Gendron, A., & Allen, KD. (2022). Overview of diverse methyl/alkyl-coenzyme M reductases and considerations for their potential heterologous expression. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.867342>
 28. Geng, A., Jin, M., Li, N., Zhu, D., Xie, R., et al. (2021). New insights into the co-occurrences of glycoside hydrolase genes among prokaryotic genomes through network analysis. *Microorganisms*, 9(2), 427. <https://doi.org/10.3390/microorganisms9020427>

29. Gorla, K., Kothari, R., Singh, HM., Singh, A., & Tyagi, VV. (2022). Biohydrogen: Potential Applications, approaches, and hurdles to overcome. *Handbook of Biofuels*, 399–418. <https://doi.org/10.1016/b978-0-12-822810-4.00020-8>
30. Gupta, S., Mortensen, MS., Schjørring, S., Trivedi, U., Vestergaard, G., et al. (2019). Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology*, 2(1), 291. <https://doi.org/10.1038/s42003-019-0540-1>
31. Güllert, S., Fischer, MA., Turaev, D., Noebauer, B., Ilmberger, N., et al. (2016). Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnology for Biofuels*, 9(1), 121. <https://doi.org/10.1186/s13068-016-0534-x>
32. Hassa J., Maus I., Off S., Pühler A., Scherer P., et al. (2018) Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions and functional relationships of microbial communities residing in biogas plants. *Applied Microbiology and Biotechnology*, 102, 5045–5063. <https://doi.org/10.1007/s00253-018-8976-7>
33. Hedlund, BP., Dodsworth, JA., Murugapiran, SK., Rinke, C., & Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of Extremophile “Microbial Dark matter.” *Extremophiles*, 18(5), 865–875. <https://doi.org/10.1007/s00792-014-0664-7>
34. Hoehler, TM., Alperin, MJ., Albert, DB., & Martens, CS. (2001). Apparent minimum free energy requirements for methanogenic archaea and sulfate-reducing bacteria in an anoxic marine sediment. *FEMS Microbiology Ecology*, 38(1), 33–41. <https://doi.org/10.1111/j.1574-6941.2001.tb00879.x>
35. Jiao, JY., Liu, L., Hua, ZS., Fang, BZ., Zhou, EM., et al. (2020). Microbial dark matter coming to light: Challenges and opportunities. *National Science Review*, 8(3), nwaa208. <https://doi.org/10.1093/nsr/nwaa280>
36. Kakuk, B., Bagi, Z., Rákhely, G., Maróti, G., Dudits, D., et al. (2021). Methane production from green and woody biomass using short rotation willow genotypes for Bioenergy Generation. *Bioresource Technology*, 333, 125223. <https://doi.org/10.1016/j.biortech.2021.125223>
37. Kanehisa, M. & Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Infection, Genetics and Evolution*, 44, 313–317. <https://doi.org/10.1016/j.meegid.2016.07.022>.

38. Kampmann, K., Ratering, S., Kramer, I., Schmidt, M., Zerr, W., et al. (2012). Unexpected stability of Bacteroidetes and Firmicutes communities in laboratory biogas reactors fed with different defined substrates. *Applied and Environmental Microbiology*, 78(7), 2106–2119. <https://doi.org/10.1128/aem.06394-11>
39. Kirkegaard, RH., McIlroy, SJ., Kristensen, JM., Nierychlo, M., Karst, SM., et al. The impact of immigration on microbial community composition in full-scale anaerobic digesters. *Science Reports*. 2017; 7, 1–11. <https://doi.org/10.1038/s41598-017-09303-0>
40. Kirschning, A. (2022). On the evolution of coenzyme biosynthesis. *Natural Product Reports*. 39, 2175-2199. <https://doi.org/10.1039/d2np00037g>
41. Knoobs, B., & Argyropoulou, V. (2016). Multiple roles of peroxiredoxins in inflammation. *Molecules and Cells*, 39(1), 60–64. <https://doi.org/10.14348/molcells.2016.2341>
42. Koch, K., Hafner, SD., Weinrich, S., Astals, S., & Holliger, C. (2020). Power and limitations of biochemical methane potential (BMP) tests. *Frontiers in Energy Research*, 8. <https://doi.org/10.3389/fenrg.2020.00063>
43. Kougiass, PG., Campanaro, S., Treu, L., Zhu, X., & Angelidaki, I. (2017). A novel archaeal species belonging to methanoculleus genus identified via de-novo assembly and metagenomic binning process in biogas reactors. *Anaerobe*, 46, 23–32. <https://doi.org/10.1016/j.anaerobe.2017.02.009>
44. Krzywinski, M., Schein, J., Birol, I., Conner, J., Gascoyne, R., et al. (2009) Circos: an information Aesthetic for Comparative Genomics. *Genome Research*. 19, 1639-1645. <https://doi.org/10.1101/gr.092759.109>.
45. Kurth, JM., Op den Camp, HJ., & Welte, CU. (2020). Several ways one goal—methanogenesis from unconventional substrates. *Applied Microbiology and Biotechnology*, 104(16), 6839–6854. <https://doi.org/10.1007/s00253-020-10724-7>
46. Lee, J., Koo, T., Yulisa, A., & Hwang, S. (2019). Magnetite as an enhancer in methanogenic degradation of volatile fatty acids under ammonia-stressed condition. *Journal of Environmental Management*, 241, 418–426. <https://doi.org/10.1016/j.jenvman.2019.04.038>
47. Letunic, I., & Bork, P. (2021). Interactive tree of life (itol) V5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293-W296. <https://doi.org/10.1093/nar/gkab301>
48. Li, D., Liu, CM., Luo, R., Sadakane, K., & Lam, TW. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de

- Bruijn graph. *Bioinformatics*; 31, 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
49. Li, Y., Alaimo, CP., Kim, M., Kado, NY., Peppers, J., et al. (2019). Composition and toxicity of biogas produced from different feedstocks in California. *Environmental Science & Technology*, 53(19), 11569–11579. <https://doi.org/10.1021/acs.est.9b03003>
 50. McMurdie PJ, & Holmes S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4): e61217. <https://doi.org/10.1371/journal.pone.0061217>
 51. Microsoft Corporation. (2018). *Microsoft Excel*. Retrieved from <https://office.microsoft.com/excel>
 52. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, GA., et al. (2020). Pfam: The Protein Families Database in 2021. *Nucleic Acids Research*, 49(D1), D412-D419. <https://doi.org/10.1093/nar/gkaa913>
 53. Naik, SN., Goud, VV., Rout, PK., & Dalai, AK. (2010). Production of First and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 14(2), 578–597. <https://doi.org/10.1016/j.rser.2009.10.003>
 54. Neu, AT., Allen, EE., & Roy, K. (2021). Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy of Sciences USA*, 118(51), e2104429118. <https://doi.org/10.1073/pnas.2104429118>
 55. Nishio, N., & Nakashimada, Y. (2013). Manufacture of biogas and fertilizer from solid food wastes by means of anaerobic digestion. *Food Industry Wastes*, 121–133. <https://doi.org/10.1016/b978-0-12-391921-2.00007-x>
 56. Nobu, MK., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, SG., et al. (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME Journal*; 9, 1710–1722. <https://doi.org/10.1038/ismej.2014.256>
 57. Nwokolo, N., Mukumba, P., Obileke, KC., & Enebe, M. (2020). Waste to energy: A focus on the impact of substrate type in biogas production. *Processes*, 8(10), 1224. <https://doi.org/10.3390/pr8101224>
 58. Olm, MR., Brown, CT., Brooks, B., Banfield, JF. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal*, 11, 2864-2868. <https://doi.org/10.1038/ismej.2017.126>
 59. Pan, S., Zhu, C., Zhao, XM., & Coelho, LP. (2022). A Deep Siamese neural network improves metagenome-assembled genomes in microbiome datasets across different

- environments. *Nature Communications*, 13(1), 2326. <https://doi.org/10.1038/s41467-022-29843-y>
60. Parks, DH., Imelfort, M., Skennerton, CT., Hugenholtz, P., & Tyson, GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25, 1043–1055. <http://www.genome.org/cgi/doi/10.1101/gr.186072.114>
 61. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 62. Read, S., Marzorati, M., Guimarães, BC., & Boon, N. (2011). Microbial Resource Management Revisited: Successful Parameters and new concepts. *Applied Microbiology and Biotechnology*, 90(3), 861–871. <https://doi.org/10.1007/s00253-011-3223-5>
 63. Rivière, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., et al. (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME Journal*, 3, 700–714. <https://doi.org/10.1038/ismej.2009.2>
 64. Thauer, RK., Kaster, AK., Seedorf, H., Buckel, W., Hedderich, R. (2008) Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nature Reviews in Microbiology*, 6, 579–591. <https://doi.org/10.1038/nrmicro1931>
 65. Rossi, E., Pecorini, I., Ferrara, G., & Iannelli, R. (2022). Dry anaerobic digestion of the organic fraction of municipal solid waste: Biogas production optimization by reducing ammonia inhibition. *Energies*, 15(15), 5515. <https://doi.org/10.3390/en15155515>
 66. Rui, J., Li, J., Zhang, S., Yan, X., Wang, Y., et al (2015) The core populations and co-occurrence patterns of prokaryotic communities in household biogas digesters. *Biotechnology for Biofuels*, 8, 158. <https://doi.org/10.1186/s13068-015-0339-3>
 67. Poudel, S., Giannone, RJ., Basen, M., Nookaew, I., Poole, FL., et al. (2018). The diversity and specificity of the extracellular proteome in the cellulolytic bacterium *caldicellulosiruptor bescii* is driven by the nature of the cellulosic growth substrate. *Biotechnology for Biofuels*, 11(1), 80. <https://doi.org/10.1186/s13068-018-1076-1>
 68. Saleem, M. (2022). Possibility of utilizing agriculture biomass as a renewable and Sustainable Future Energy Source. *Heliyon*, 8(2), e08905 <https://doi.org/10.1016/j.heliyon.2022.e08905>
 69. Sarkar, N., Ghosh, SK., Bannerjee, S., & Aikat, K. (2012). Bioethanol production from Agricultural Wastes: An overview. *Renewable Energy*, 37(1), 19–27. <https://doi.org/10.1016/j.renene.2011.06.045>

70. Schink B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews*, 61, 262-280. <https://doi.org/10.1128/mmbr.61.2.262-280.1997>
71. Schmeing, TM. (2013). Ribosome structure. *Encyclopedia of Biological Chemistry*, 128–135. <https://doi.org/10.1016/b978-0-12-378630-2.00400-x>
72. Singh, D., Sharma, D., Soni, SL., Sharma, S., Kumar Sharma, P., et al. (2020). A review on feedstocks, production processes, and yield for different generations of biodiesel. *Fuel*, 262, 116553. <https://doi.org/10.1016/j.fuel.2019.116553>
73. Skorupinska-Tudek, K., Wojcik, J., & Swiezewska, E. (2008). Polyisoprenoid alcohols—recent results of structural studies. *The Chemical Record*, 8(1), 33–45. <https://doi.org/10.1002/tcr.20137>
74. Solden, L., Lloyd, K., & Wrighton, K. (2016). The bright side of Microbial Dark matter: Lessons learned from the uncultivated majority. *Current Opinion in Microbiology*, 31, 217–226. <https://doi.org/10.1016/j.mib.2016.04.020>
75. Stolze, Y., Bremges, A., Rummig, M., Henke, C., Maus, I., et al. (2016) Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. *Biotechnology for Biofuels*, 9, 1–18. <https://doi.org/10.1186/s13068-016-0565-3>
76. Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., et al. (2020). Interpretations of environmental microbial community studies are biased by the selected 16S rna (gene) amplicon sequencing pipeline. *Frontiers in Microbiology*, 11, 550420. <https://doi.org/10.3389/fmicb.2020.550420>
77. Sun, Z., Huang, S., Zhang, M., Zhu, QY., Haiminen, N., et al. (2020). Challenges in benchmarking metagenomic profilers. *Nature. methods*, 18, 618-626. <https://doi.org/10.1101/2020.11.14.382994>
78. Szilágyi, Á., Bodor, A., Tolvai, N., Kovács, KL., Bodai, L., et al. (2021) A comparative analysis of biogas production from tomato bio-waste in mesophilic batch and continuous anaerobic digestion systems. *PLoS One*, 16, 1–17. <https://doi.org/10.1371/journal.pone.0248654>
79. Tam, R., & Saier, MH. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews*, 57(2), 320–346. <https://doi.org/10.1128/mr.57.2.320-346.1993>
80. Ting, HN., Lin, L., Cruz, RB., Chowdhury, B., Karidio, I., et al. (2020). Transitions of microbial communities in the solid and liquid phases during high-solids anaerobic

- digestion of organic fraction of municipal solid waste. *Bioresource Technology*, 317, 123951. <https://doi.org/10.1016/j.biortech.2020.123951>
81. U.S. Energy Information Administration - EIA - independent statistics and analysis. Where greenhouse gases come from - U.S. Energy Information Administration (EIA). (n.d.). Retrieved January 18, 2023, from <https://www.eia.gov/energyexplained/energy-and-the-environment/where-greenhouse-gases-come-from.php>
 82. Vijayan, D.S., Mohan, A., Nivetha, C., Sivakumar, V., Devarajan, P., et al. (2022). Treatment of pharma effluent using anaerobic packed bed reactor. *Journal of Environmental and Public Health*, 2022, 4657628. <https://doi.org/10.1155/2022/4657628>
 83. Wang, Z., Jiang, Y., Wang, S., Zhang, Y., Hu, Y., et al. (2020). Impact of total solids content on anaerobic co-digestion of pig manure and food waste: Insights into shifting of the methanogenic pathway. *Waste Management*, 114, 96–106. <https://doi.org/10.1016/j.wasman.2020.06.048>
 84. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
 85. Wirth, R., Kovács, E., Maróti, G., Bagi, Z., Rákhely, G., et al. (2012). Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnology for Biofuels*, 5(1), 41. <https://doi.org/10.1186/1754-6834-5-41>
 86. Wirth, R., Pap, B., Dudits, D., Kakuk, B., Bagi, Z., et al. (2021). Genome-centric investigation of anaerobic digestion using sustainable second and third generation substrates. *Journal of Biotechnology*, 339, 53–64. <https://doi.org/10.1016/j.jbiotec.2021.08.002>
 87. Yang, C., Chowdhury, D., Zhang, Z., Cheung, W.K., Lu, A., et al. (2021). A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Computational and Structural Biotechnology Journal*, 19, 6301–6314. <https://doi.org/10.1016/j.csbj.2021.11.028>
 88. Yi, J., Dong, B., Jin, J., & Dai, X. (2014). Effect of increasing total solids contents on anaerobic digestion of food waste under mesophilic conditions: Performance and Microbial Characteristics Analysis. *PLoS ONE* 9(7), e102548. <https://doi.org/10.1371/journal.pone.0102548>