Molecular characterization of microbial communities in small-scale biodigesters using whole genome shotgun metagenomic approach

by

Matamela Tendani

18008492

Research dissertation submitted in fulfilment of the requirements for the award of
a Master of Science Degree in Biochemistry

To The

Faculty of Science, Engineering, and Agriculture

Department of Biochemistry & Microbiology

University of Venda

Thohoyandou, Limpopo

South Africa

Supervisor: Dr L.M Mathomu

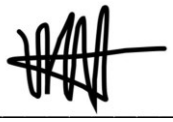
Co-supervisor: Prof T. Matambo (UNISA)

August 2025

© University of Venda

DECLARATION

I, the undersigned, **Matamela Tendani (18008492)**, declare that this thesis with title "Molecular characterization of microbial communities in small scale biodigester using whole genome shotgun metagenomic approach" for the award of Master of Science degree in Biochemistry is my original work and has not been previously submitted to this or any other university or institution for degree. The document does not contain other persons' writing unless expressly acknowledged and referenced accordingly.



Matamela T

© University of Venda

ACKNOWLEDGEMENT

First and foremost, I would like to thank the almighty God who has been with me throughout this journey and brought me this far.

I would also like to thank the National Foundation of Research (NRF) and DSI-TIA for funding this project. It wouldn't have been an easy journey without the support.

My deepest gratitude to my supervisor, Dr LM Mathomu, for allowing me to do my MSc project in the laboratory; his support and guidance helped me finish my project. Thank you to Professor NE Maluta for your unveiling support during this project. To my co-supervisor, Professor T. Matambo, thank you for agreeing to be my co-supervisor and for your support. Most importantly, my mentor, Dr R Nkuna (UNISA), showed me a way of analyzing shotgun metagenomic data. I appreciate your guidance and motivation.

To Dr Nekhubvi, your input and suggestions during our Biogas consortium meetings have made me who I am today.

I also extend my special thanks to Prof NE Madala for your motivation and guidance. It was a great pleasure to be a tutor for your modules. You have played a more significant role in my academic and life journey.

Most of the figures in this project were constructed with the guidance of Mr. Adivhaho Khwathisi. I want to extend my gratitude; you have played a crucial role in this project. I wish you well in your academic career.

I want to acknowledge BioNanotech and Mass Spectrometry Research Group in the Department of Biochemistry and Microbiology.

To all biogas consortium, thank you for teaching me various aspects of both research and life, Miss Nethavhanani (persistence and dedication), Miss Phaphana, Miss Nemaridili (Bacterial culturing), Miss Tsotetsi, my mentee (MultiQC, Trimmomatic), Madzivhandila Vhukudo and Mr Mothiba.

My mom took care of my daughter while I was studying. Your support is highly appreciated. Finally, I sincerely thank Mr. Nthangeni for your love, encouragement, and support.

© University of Venda

University of Venda
Creating Future Leaders

DEDICATION

To my mother, Thivhakoni Dorah Mapholi, and daughter, Natalie Ngaipfi Nthangeni,

I am doing this for you.

© University of Venda

ABSTRACT

Increasing attention has been directed towards the role of microbial communities in anaerobic digestion (AD). During AD, microorganisms degrade organic waste, such as animal excrement, through four key phases: hydrolysis, acidogenesis, acetogenesis, and methanogenesis, resulting in digestate and biogas production. Advances in metagenomic techniques, surpassing traditional culture-based methods, have enabled a deeper characterization of the microbial populations involved in this process. However, a knowledge gap remains regarding small-scale biodigesters, particularly those in rural areas. This study employs whole-genome shotgun metagenomics integrating data from metagenomic reads, contigs, and metagenome-assembled genomes (MAGs) can achieve a more comprehensive understanding of microbial taxonomy and functional capabilities in biodigesters. This study investigates the microbial communities in small-scale biodigesters, using taxonomic data from various sources to reveal the functional roles of the microorganisms. The research identified 4 superkingdoms, 115 phyla, 107 classes, 189 orders, 322 families, 738 genera, and 2046 species. Bacteria dominated ranging from 79 to 89%, whereas archaea accounted for 11 to 20% of the community. The eukaryotic microbial's relative abundance was less than 1%. Both read-based and contig-based classifiers found *Proteobacteria*, *Firmicutes,* and *Bacteroidetes* to be the most abundant bacterial phyla, followed by *Actinobacteria* and *Chloroflexi* in cow dung-fed biodigesters. The archaeal community was dominated by *Euryarchaeota*, with *Methanomicrobiales* and *Methanobacteriales* as the predominant orders. Functional analysis reveals that most identified microorganisms contribute indirectly to methanogenesis by playing crucial roles in hydrolysis and acidogenesis through genes involved in the metabolism of organic molecules broken down during the four phases of anaerobic digestion. RAST analysis annotated 11 methanogen-related enzymes and their abundance across different samples. Future research should focus on cultivating and monitoring these identified microbes to determine optimal conditions for maximizing biogas production in laboratory settings, thus advancing environmental bioaugmentation strategies.

Keywords: Microbial Diversity, Cow dung, Anaerobic Digestion, Biodigester, Whole Genome Shotgun sequencing, Bioinformatics

© University of Venda

PUBLICATIONS ORIGINATING FROM THIS DISSERTATION

Poster/Presentations

1. **Matamela, T., Matambo, T., Mathomu, L.M (2024).** Molecular characterization of microbial diversity in small-scale biodigesters using whole genome shotgun metagenomic approach. UNIVEN Research Open Day 2024. Thohoyandou: South Africa. **Poster**
2. **Matamela, T., Nethavhanani T., Maluta N.E., Mathomu, L.M (2024).** Toward Sustainable Biogas Production. AJ-Core Workshop of Africa-Japan Collaborative Research on Environmental Science. UNISA: South Africa**. Presentation**

Publications

3. **Matamela T., Matambo, T., Mathomu, L.M.** Microbial population in a four-point crossed process of hydrolysis, acidogenesis, acetogenesis, and methanogenesis: A review of bioaugmentation and characterization techniques, **In Preparation.**
4. **Matamela T., Matambo, T., Mathomu, L.M.** Utilizing Whole-Genome Shotgun Metagenomics to Unveil the Hidden World of Microbial Communities and functional capability in Small-Scale Biodigesters in Rural Limpopo, **In Preparation**.

© University of Venda

University of Venda
Creating Future Leaders

Table of contents

DECLARATION i
ACKNOWLEDGEMENT ii
DEDICATION iii
ABSTRACT iv
PUBLICATIONS ORIGINATING FROM THIS DISSERTATION v
LIST OF FIGURES viii
LIST OF TABLES ix
LIST OF ABBREVIATIONS x
CHAPTER 1: BACKGROUND AND STUDY DESIGN 1
1.1. Background 1
1.2. Study design 3
1.2.1 Problem Statement 3
1.2.2. Study Rationale 4
1.2.3. Aims and Objectives 5
CHAPTER 2: LITERATURE REVIEW 6
2.1. Biochemical process of substrate conversion in anaerobic digestion 6
2.2. Methanogenesis pathways and related genes: The Final Phase of Anaerobic Digestion 8
2.3. Next Generation Sequencing- a commonly used characterization technique for microbial communities in biodigesters 9
2.4. Metagenomic approach for characterization of microorganisms in biodigesters 10
2.5. Bioinformatics analysis of metagenomic sequencing data 11
2.6. Background of MEGAHIT, a metagenomics assembler tool 14
2.7. Visualization of metagenomic data 15
CHAPTER 3: MATERIALS AND METHODS 16
3.1. Ethical Clearance 16
3.2. Geographic and status information of the biodigester site 16
3.3. Sample collection and whole genome sequencing 18
3.4. Pre-processing of raw metagenomic data 18
3.4.1. Quality control of the generated raw sequencing reads 18
3.4.2. Trimming of raw data 18
3.4.3. Taxonomic classification of reads 19
3.5. Taxonomic classification utilizing long sequence-based tools 19
3.5.1. Metagenomic sequence assembly 19

© University of Venda

..... 20
a) Meta QUAST, as a quality assessment tool 20
b) Taxonomic classification of generated contigs 20
3.6. Metagenomic binning, assessment of bins quality, and construction of MAGs 20
(a) Grouping of contigs using MetaBAT2 20
(b) Quality assessment of bins using CheckM 20
(c) Dereplication of the generated bins 21
(d) Taxonomic classification of MAGs 21
3.7. Unveiling the function capability of microbial communities in small-scale biodigesters 21
3.8. V isualization of metagenomic data 21
CHAPTER 4: RESULTS 23
4.1. Whole genome shotgun (WGS) sequencing statistics 23
4.2. Whole-genome shotgun raw sequence read quality assessment using FastQC 23
4.3. Trimming and removal of adapters from raw sequences 24
4.4 Taxonomic profile utilizing read-based tool Kaiju 29
4.5. Classification of microbial communities based on read-based tool 32
4.6. Reference-independent whole genome assembly of filtered reads 38
4.7. Taxonomic assignment of contigs 38
4.8. Reconstruction of the metagenome-assembled genome (MAGs) 44
4.9. Taxonomic classification of MAGs utilizing GTDB-tk 45
4.10. Functional characterization of microbial population in small-scale biodigesters 47
4.11. Overall functional potential across the samples for bacteria and archaea 48
CHAPTER 5: DISCUSSION AND CONCLUSION 50
CHAPTER 6: REFERENCES 55

© University of Venda

LIST OF FIGURES

Figure 1: Systemic diagram of Anaerobic Digestion process……..... 11

Figure 2: Overview of two metagenomic approaches and bioinformatics analysis …….…….…… 14

Figure 3: Sampling locations of the project and geographical co-ordinates ……...…... 17

Figure 4: Metagenomic workflow of the study……... 18

Figure 5: FastQC statistics for Dzwerani samples before and after trimming…….…… 26

Figure 6: FastQC statistics for Agriculture digester samples before and after trimming …… 27-28

Figure 7: Relative abundance stacked graph of microbial communities at a kingdom level……..…. 30

Figure 8: Microbial communities at a phyla level present in digestate of Dzwerani and Univen Agriculture farm samples……...…..……. 31

Figure 9: Relative abundance of microbes in class level of taxonomic classification……..…. 33

Figure 10: Relative abundance of bacterial and archaeal microbial order classified using read-based tool…….……... 34

Figure 11: Relative abundance of bacterial and archaeal microbial at a family level……. ….. 35

Figure 12: Microbial community composition at the phylum level based on CAT analysis…… 40

Figure 13: Clustered bar graph of Microbial communities at the class level identified using CAT…….…… 41

Figure 14: The top 20 microbial communities by relative abundance at the order level clustered bar graph ……..…… 42

Figure 15: Number of identified taxa in each bins generated using MetaBAT2……. …..……..... 43

Figure 16: Phylogenetic tree of bacterial and archeal species distribution across the biodigesters……. 45

Figure 17: Functional capability of Euryarchaeota phyla……. 46

Figure 18: Relative abundance of hits annotated to the different methanogenesis related genes…… 48

© University of Venda

LIST OF TABLES

Table 1: Statistics of generated raw reads after sequencing with Illumina Nextseq500/550 system…… 23

Table 2: Quality assessment of reads generated before and after pre-processing using Trimmomatic…….……..……...……. 25

Table 3: Overview of relative abundance of microbial communities (excluding phyla, genus. 36

Table 4: MetaQUAST statistics for the sample assemblies generated using the de novo-based tool MEGAHIT……... 37

Table 5: Overview number of microbes identified based on taxonomic hierarchy utilizing long read-based classifiers (CAT)……. 39

© University of Venda

University of Venda
Creating Future Leaders

LIST OF ABBREVIATIONS

AD:	Anaerobic Digestion
NGS:	Next-Generation Sequencing
FTR:	Formylmethanofuran transferase
FDH:	Formate dehydrogenase
MCR:	Methyl-CoM reductase
GHG:	Greenhouse gas
rRNA:	Ribosomal Ribonucleic Acid
LCFA:	Long-Chain Fatty Acids
SCFA:	Short-Chain Fatty Acids
MPN:	Most Probable Number
DGGE:	Denaturing Gradient Gel Electrophoresis
RFLP:	Restriction Fragment Length Polymorphism
FISH:	Fluorescent In-Situ Hybridization
qPCR:	quantitative real-time Polymerase Chain Reaction
VFAs:	Volatile Fatty Acids
SDS:	Sodium Dodecyl sulfate
VOCs:	Volatile Organic Compounds
HRT:	Hydraulic Retention Time
OTUs:	Operational Taxonomic Units
ACE:	Abundance Coverage-based Estimator of species richness
Bp:	Base Pair

© University of Venda

rRNA:	Ribosomal RiboNucleic Acid
GC:	Guanine-Cytosine
PCR:	Polymerase Chain Reaction
HPLC:	High-Performance Liquid Chromatography
BLAST:	Basic Local Alignment Search Tool
WGS:	Whole Genome Shotgun
DNA:	Deoxyribonucleic acid
RNA:	Ribonucleic acid
PacBio:	Pacific Biosciences (PacBio),
UV:	UltraViolent
QC:	Quality Control
MG-RAST:	MetaGenomic Rapid Annotation using Subsystem Technology
MAGs:	Metagenomic-Assembled Genomes
HGT:	Horizontal Gene Transfer
RDP:	Ribosomal Database Project
DBG:	de Bruijn Graphs
QUAST:	QUality Assessment Tool
NCBI:	National Center for Biotechnology Information
MA:	missasssembly
LCA:	Last Common Ancestor
MASH:	MinHash
ANI:	Average Nucleotide Identity
MetaBAT:	Metagenome Binning with Abundance and Tetra-nucleotide frequency

© University of Venda

CHAPTER 1: BACKGROUND AND STUDY DESIGN

1.1. Background

The capacity of microorganisms to break down biomass feedstocks, such as agricultural, industrial, municipal, and animal waste from livestock, has caught the attention of researchers investigating biomethane production (Rahman *et al*., 2021; Singh and Singh, 2022). This innovation has resulted in a decrease in the amount of food waste and the carbon footprint, which is known to contribute to greenhouse gas (GHG) emissions from fossil fuels and open field burning (Senés-Guerrero *et al*., 2019; Uche *et al*., 2020; Mahmudul *et al*., 2022; Lee *et al*., 2024). The World Energy Council projects estimated that there will be over 6 million tonnes per day of waste generated by 2025. According to the 2014 Framework for Climate and Energy, the European Council has committed to reducing GHG emissions by at least 40% by 2030 compared to 1990 by increasing the use of twenty-seven shares of renewable energy (Zhang *et al*., 2019). In 2019, their action plan was unveiled to achieve zero net emissions of GHG by 2050 in European Union countries (Menzel *et al*., 2020).

Anaerobic Digestion is a promising technology for safely disposing of waste while generating energy (methane) without adverse environmental effects (Kainthola *et al*., 2019; Mukhuba *et al*., 2020; Niya *et al*., 2024). In this process, complex communities of microorganisms (from the bacteria, fungi, and archaea domains) decompose biodegradable materials through their collective metabolic activities in four sequential processes (Hydrolysis, Acidogenesis, Acetogenesis, and Methanogenesis) to produce biogas (Noraini *et al*., 2017; Laiq Ur Rehman *et al*., 2019; Bonugli-Santos *et al*., 2022). Biogas primarily consists of methane (50–75%) and carbon dioxide (25–50%), with trace amounts of other gases such as hydrogen sulfide (0–0.5%), nitrogen (0–10%), and oxygen (0–2.5%) (Zain and Mohamed, 2018; Pyzik *et al*., 2018; Diagi *et al*., 2019; Kim *et al*., 2022; Arelli *et al*., 2023). The first three phases are dominated by bacteria, i.e., hydrolytic in the first phase, acidogens in the second, and acetogens in acetogenesis. Other microbes can play a role in more than one phase. The final release of methane is specialized by the kingdom of archaea known as methanogens (Rahman *et al*., 2021). Most of the phases occurs in the absence of molecular oxygen, with the exception of hydrolysis, which can tolerate small amounts of oxygen (Pramanik *et al*., 2019; Fu *et al*., 2023; Belone *et al*., 2024).

The stable operation and efficient biogas production rely on the complex microbial communities in biodigesters. Therefore, a thorough understanding of these communities' structure and function, taxonomy, metabolism, and the relationships between system function,

© University of Venda

biodiversity, and operational conditions is crucial for enhancing biogas productivity (Zhang *et al*., 2019). Future enhancements in anaerobic biodigester performance to increase biomethane yield could be achieved through bioaugmentation by manipulating microorganisms and using them as inoculants or enzymes in laboratory-scale studies. The interplay of microbial populations in biodigesters has led to innovative approaches in the scientific field.

Key microbial populations responsible for specific metabolic processes in biodigesters have been identified using traditional culture-dependent techniques (Brumfield *et al*., 2020; Qurbani *et al*., 2024). However, these methods only reveal some microbial communities in biodigesters since some cannot be isolated and cultured (Uyaguari-Diaz *et al*., 2016; Sudarikov *et al*., 2017; Dabban *et al*., 2024). This limitation hampers scientific understanding of microbial communities' function, enzymes, genes, and activity in biogas production (Vanwonterghem *et al*., 2014). Attention has shifted to metagenomics, which allows microbial populations to be studied directly in their environment without growing them in artificial conditions that mimic their natural habitats. (Mallawaarachchi *et al*., 2024; Sharma *et al*., 2024). The rapidly growing metagenomic interest has led scientists to reveal previously unobserved microbes in their diverse and complex habitats, such as the soil (Brumfield *et al*., 2020), human gut (McGregor *et al*., 2023), marine (Akter *et al*., 2023), biodigesters (Muñoz-Ramírez *et al*., 2024).

The two metagenomic approaches include amplicon-based, which uses the 16S rRNA gene as a marker for bacterial and archaeal species and is amplicon-independent (Brumfield *et al*., 2020). Overrepresentation of artificial diversity may be observed when solely relying on the 16S rRNA marker because most bacterial species have multiple 16S rRNA genes. The amplicon-based approach introduces bias association primers with polymerase chain reaction (PCR) since some microorganisms are left out during the amplification process, subsequently in sequencing (Escobar-Zepeda *et al*., 2018; Forry *et al*., 2025). The approach is limited to resolving a substantial fraction of diversity in the microbial community because only insight into taxonomic composition is provided, thus rendering a comprehensive understanding of the entire microbial genes/functions (Jünemann *et al*., 2017).

The shortcomings of amplicon-based techniques have led to adopting whole genome shotgun (WGS) sequencing (Kang *et al*.,2015; Pérez-Llano *et al*., 2021; Jain *et al*., 2024). This technique has made it possible to analyze microorganisms, including those that traditional methods and 16S rRNA amplicon sequencing cannot detect and identify (Chinwendu *et al*., 2020; Kimisto *et al*., 2023). WGS offers insights into microbial diversity, genetic and evolutionary relationships, microbial morphology and function, and the interaction of microbes

with the environment (Li *et al*., 2022; Niya *et al*., 2024). The WGS approach can help answer questions such as 'What are the microbes doing?' by functional annotating the identified genes and reconstructing the metabolic pathways.

Bioinformatics grew similarly to the rapidly growing metagenomic data by introducing software and tools to analyze the massive data to answer research questions. Beginning with tools that allow pre-treatment of data, constructing draft genomes representing microorganisms, and grouping draft genomes based on shared similarities in binning and functional annotation of genes. Taxonomic classification of microbial populations can be achieved directly from pretreatment of raw sequencing data or after construction of draft genome via assembly of paired end or single reads. It is essential to select a tool based on its sensitivity, precision, speed, and the algorithm implemented.

1.2. Study design

1.2.1 Problem Statement

The shift from traditional culture-based techniques to advanced methodologies such as metagenomics has significantly enhanced the understanding of microbial structure and function across various environments, including biodigesters. Whole Genome Shotgun (WGS) sequencing generates extensive data, necessitating the development of sophisticated tools for accurate classification. Taxonomic classification of microbial communities in biodigesters can be achieved using two primary approaches: read-based and genome-based methodologies. Read-based annotation involves directly comparing individual short reads to a comprehensive reference database. While this method provides rapid results, it is prone to errors and spurious annotations. In contrast, identifying novel taxa reliably requires assembling reads into contiguous sequences (contigs). Furthermore, metagenome-assembled genomes (MAGs), derived through the binning process, offer a more accurate means of classifying microbial communities compared to read- and contig-based approaches. However, the success of assembly and binning depends on sufficient genome sequence coverage, which can lead to the underrepresentation of rare microbial species. This approach not only reveals the hidden diversity of microbial communities but also facilitates the early detection of potential issues, allowing for the identification of factors contributing to suboptimal gas yields and preventing complete digester failure. Insights gained from these analyses provide critical knowledge to optimize biogas production. Given the increasing global demand for electricity, there is an urgent need for alternative, cost-effective, and sustainable energy sources. This study

© University of Venda

contributes to advancing scientific understanding while aligning with the objectives of the National Development Plan 2030, particularly Goal 7, which emphasizes access to affordable and sustainable energy.

1.2.2. Study Rationale

Assessing the microbial community and understanding gene function in anaerobic digestion (AD) within biodigesters are crucial for optimizing biogas production through bioaugmentation. Scientific progress has enabled the direct study of microorganisms in their natural habitats via metagenomics, eliminating the need for culturing. This advancement has spurred the development of bioinformatics pipelines and tools to analyze extensive metagenomic datasets, facilitating the comprehension of microbial communities in biodigesters.

Taxonomic classification of metagenomic sequencing data has been achieved through two approaches: read-based, which entails mapping individual short reads to known reference databases, and genome-based, which involves assembling short reads to reconstruct genomes and grouping resulting contigs based on shared characteristics. Recent studies have utilized read-based (Campanaro *et al*., 2016; Saetta *et al*., 2023) and genome-based tools (Deng *et al*., 2022) to classify microbial communities into taxa. However, relying solely on these methods may lead to less accurate characterizations due to challenges such as highly conserved regions in reads hindering correct discrimination, horizontal gene transfer between microorganisms, and misannotated sequences in reference databases.

Grouping genomes based on shared characteristics during binning offers insights into microbial taxa and functions. However, some sequences may be filtered out due to unmapping in the reference database, resulting in incomplete characterization of microbial community structure.

The utilization of annotation pipelines in metagenomic sequencing is crucial for retrieving diverse microbial populations, especially rare taxa. These pipelines ensure the precise annotation of sequencing reads by leveraging information from contigs and metagenome assembled genomes (MAGs). Employing an integrated pipeline facilitates a thorough characterization of microbial community structure across the four phases of anaerobic digestion (AD). Additionally, functional annotation of metabolic genes offers insights for future investigations targeting the optimization of AD phases via bioaugmentation, aiming to enhance methane production efficiency.

© University of Venda

1.2.3. Aims and Objectives

1.2.3.1 Aim

To unveil the diversity and functional capabilities of microbial populations in small-scale biodigesters using a metagenomic approach.

1.2.3.2. Specific Objectives

(a) To evaluate and enhance the quality of raw sequences generated from whole genome shotgun metagenomics using various tools in the galaxy metagenomic pipeline

(b) To thoroughly investigate the taxonomic composition of microbial communities within anaerobic biodigesters using read, assembly, and binning-based classifiers.

(c) To predict and annotate the functional genes present within the microbial communities

© University of Venda

CHAPTER 2: LITERATURE REVIEW

2.1. Biochemical process of substrate conversion in anaerobic digestion

Biodigesters employ anaerobic digestion (AD) to break down various biodegradable substances like food waste, animal manure, and agricultural residue into methane in the absence of oxygen (Mao *et al*., 2015; Noraini *et al.*, 2017; Diag *et al*., 2019; Kainthola *et al*., 2019; Kim *et al*., 2022). This process consists of four main phases (Hydrolysis, Acidogenesis, Acetogenesis, and Methanogenesis), where diverse microbial communities interact interchangeably to produce mainly methane (CH_4) and carbon dioxide (CO_2) (**Fig. 1) (**Kimisto *et al*., 2023; Shabbirahmed *et al*., 2024). Other trace amounts produced include nitrogen (N_2), Hydrogen sulfide (H_2S), and hydrogen (H_2) dependent on the feedstock used (St-Pierre and Wright, 2014; Meegoda *et al*., 2018; Ram & Nikhil, 2022).

In the first stage, microorganisms such as Bacteroides sp., Butyrivibrio sp., Clostridia sp., Fusobacterium sp., Micrococci sp., Selenomonas, and Streptococcus sp. break down complex organic matter proteins, polysaccharides (cellulose and lignocellulose, hemicellulose, cellobiose), and lipids like triglyceride esters) into amino acids, simple sugars, and long-chain fatty acids (LCFA) (Laiq Ur Rehman *et al*., 2019; Pasalari *et al.,* 2021). In addition to microbial activity, hydrolysis in this stage can be facilitated by mechanical, chemical, and enzymatic processes that further assist in the breakdown of complex substrates. Various extracellular enzymes like proteases, lipases, cellulases, amylases, cellobioses, and xylanases facilitate this breakdown (Ijoma *et al*., 2021).

Studies indicate that the same microbes active in the hydrolysis stage are also active in the acidogenic stage. For example, Long-Chain Fatty Acids (LCFA) must first be fermented by specific hydrolytic bacteria into Short-Chain Fatty Acids (SCFA), typically C_3-C_6 compounds (Duan *et al*., 2021; Janesch *et al*., 2023). Acidogenic bacteria then convert these SCFA, sugars, and amino acids into organic acids, carbon dioxide (CO_2), hydrogen (H_2), and ammonia (Laiq Ur Rehman *et al*., 2019). These hydrolytic and acidogenic bacteria are sometimes referred to as fermentative bacteria. There exists a complex interplay between acidogenic bacteria and methanogens; thus, intermediates like acetic acid, hydrogen, carbon dioxide, and other single-carbon compounds (formic acid, methyl sulfide, methanol, methylamines) can be directly utilized for methanogenesis (ManyiLoh *et al*., 2013; Ram & Nikhil, 2022). Most of these microbes are facultative anaerobes; they can survive in the presence and absence of oxygen.

© University of Venda

Commonly known microbes that are significant players in this phase are those that belong to the genus Pseudomonas, Clostridium, Bacillus, Flavobacterium, and Micrococcus.

Common acidogenic microbes in this stage include Anaerocellum thermophilum, Anaerococcus, Bacillus cereus, Aminobacterium, Bacteroides, Butyrivitro, Caldicellulosiruptor, and Cellulomonas (Ram & Nikhil, 2022).

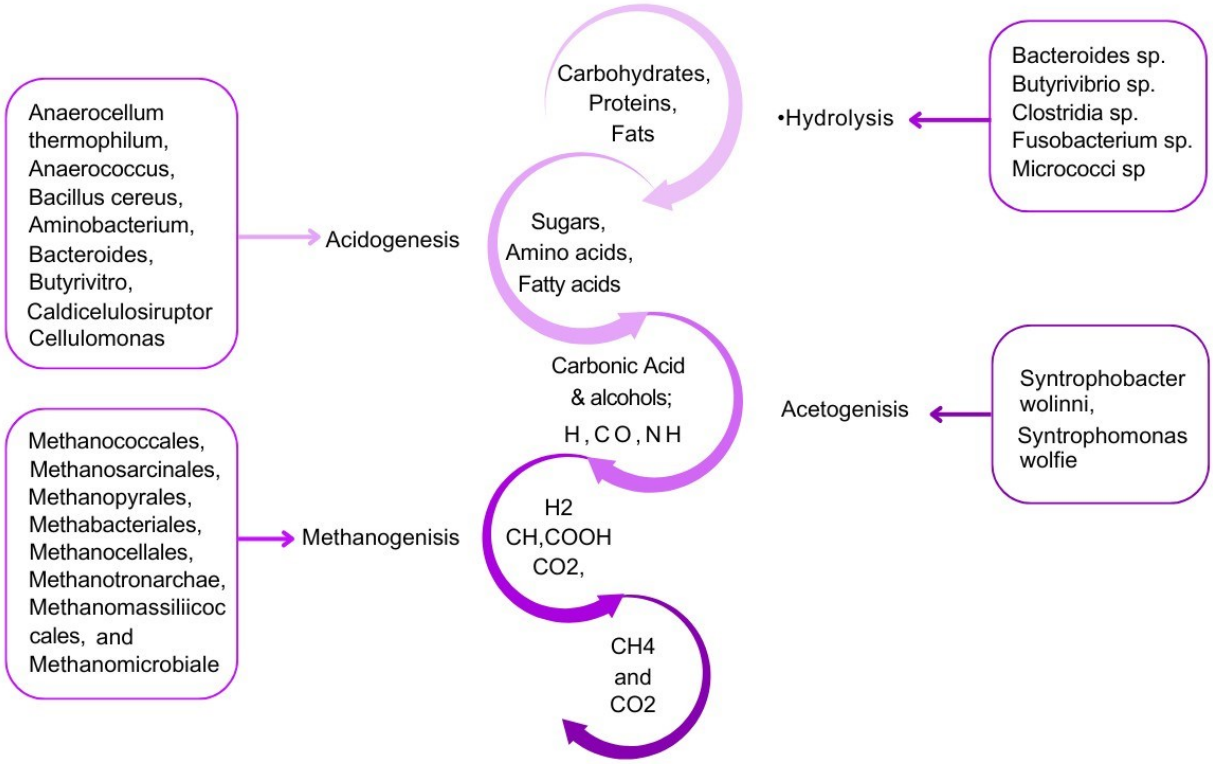


Figure 1: Diagram illustrating the phase division of anaerobic digestion flow. Each phase is characterized by a distinct microbial community specialized in biodegrading molecules (Modified from www.researchgate.net , assessed on 3rd August 2023).

In the third phase, mutually dependent bacteria ferment organic acids into acetate, and carbon dioxide is converted into acetic acid (Janesch *et al*., 2023). *Clostridium aceticum* was among the first microorganisms to demonstrate this conversion of CO_2 to acetic acid. Here, organic acids like butyric acid and propionic acid are transformed into acetic acid by *Syntrophobacter wolinni* and *Syntrophomonas wolfei*, respectively (Sharma *et al*., 2023). Alcohol produced during acidogenesis is also oxidized to yield substrates for methanogenesis.

© University of Venda

Finally, acetate is converted to methane by methanogenic archaea in the fourth phase. Some of the cited methanogens involves *Methanobacteria*, *Methanocella*, *Methanococcus*, *Methanomicrobia*, *Methanopyra*, *Methanosarcina* and *Methanomassiliicoccus*.

2.2. Methanogenesis pathways and related genes: The Final Phase of Anaerobic Digestion

In anaerobic digestion, archaea play a pivotal role, particularly in the process of methanogenesis. Archaea differ from bacteria in several fundamental aspects, including their genetic makeup, metabolic pathways, and environmental tolerance. Both archaea and bacteria are single-celled microorganisms; however, archaea possess unique lipid membranes, which are more resistant to extreme conditions such as high temperatures and acidic environments. Additionally, archaea's ribosomal RNA (rRNA) sequences are more similar to those of eukaryotes than to bacteria, highlighting their distinct evolutionary lineage (Woese *et al*., 1990). Unlike bacteria, archaea lack peptidoglycan in their cell walls, instead utilizing other compounds, such as pseudomurein, which further distinguishes them. Archaea are also well adapted to thrive in extreme environments, including the anoxic conditions found in biodigesters, where they are responsible for the production of methane through methanogenesis (Miller *et al*., 2019).

Methanogenesis involves specific pathways mediated by distinct methanogens and key enzymatic reactions. Hydrogenotrophic methanogens, including *Methanobacteria*, *Methanocella, Methanococcus, Methanomicrobia*, and *Methanopyra*, reduce carbon dioxide (CO_2) using hydrogen (H_2). This process is facilitated by enzymes such as formylmethanofuran dehydrogenase (fmd), formylmethanofuran-H4MPT formyltransferase (ftr), N(5), N(10)methenyltetrahydromethanopterin cyclohydrolase (mch), and F420-dependent methylenetetrahydromethanopterin reductase (mer). These steps results in the transfer of the methyl group from methyl-tetrahydromethanopterin to coenzyme M via methyltransferase, followed by methane synthesis catalyzed by methyl-CoM reductase (mcr). Acetoclastic methanogens, primarily *Methanosarcina*, ferment acetate by converting it into acetyl-CoA through acetate kinase (ack), phosphate acetyltransferase (pta), or acetyl-CoA synthetase (acs). The acetyl-CoA is then converted via acetyl-CoA decarbonylase, reducing the methyl group to methane and oxidizing the carbonyl group to carbon dioxide (Ferry, 2011). In the methylotrophic pathway, methanol and methylamines are metabolized by *Methanosarcina* and *Methanomassiliicoccus*. These methanogens transfer methyl groups to corrinoid proteins using methyltransferase. The methyl groups are subsequently processed through methanogenic

pathways, involving enzymes such as coenzyme B-coenzyme M, heterodisulfide reductase (hdr) and hydrogenase (hmd), to produce methane (Deppenmeier, 2002). The final stage of methanogenesis in all pathways involves the reduction of intermediates to methane by methylCoM reductase (mcr), with coenzyme F430 and N-7 mercapto heptanoyl threonine phosphate (HS-HTP) playing essential roles (Gao *et al*., 2022; Ram and Nikhil, 2022).

2.3. Next Generation Sequencing- a commonly used characterization technique for microbial communities in biodigesters

Next Generation Sequencing (NGS) has enabled comprehensive studies of microbial populations in diverse environments, including biodigesters (Senés-Guerrero *et al*., 2019; Pillai *et al*., 2022), wetlands (Mellado and Vera, 2021; Ballesteros *et al*., 2023), and oceans (Flaviani *et al*., 2018; Mazur-Marzec *et al*., 2024). This technological advancement has revolutionized genomics, bioinformatics, and microbial ecology. NGS allows for the rapid sequencing of nucleic acids, such as DNA and RNA, processing millions of nucleotide fragments simultaneously, in contrast to the traditional Sanger sequencing method (Liu *et al*., 2012).

Next-Generation Sequencing (NGS) technology has revolutionized the field of genomics and microbial ecology, including the study of microbial populations in biodigesters. NGS encompasses a range of modern sequencing technologies that enable rapid, high-throughput sequencing of DNA and RNA, offering a significant advantage over traditional Sanger sequencing through its ability to simultaneously process millions of DNA fragments (Fukuda *et al*., 2016). Key NGS platforms, such as Illumina, Roche 454, Ion Torrent, Pacific Biosciences (PacBio), and Oxford Nanopore, offer unique capabilities, including high accuracy, long read lengths, and real-time sequencing (Zhang *et al* 2021; Satam *et al*., 2023). In the context of biodigesters, NGS facilitates microbial community profiling by sequencing 16S rRNA genes to elucidate microbial composition and diversity. At the same time, metagenomics provides a comprehensive view of the genetic potential within the community by sequencing the entire DNA content of a sample. Metatranscriptomics further enhances this understanding by sequencing RNA to reveal active metabolic pathways and microbial responses to varying conditions.

Additionally, NGS enables comparative genomics, detection of rare taxa, and functional gene analysis, all of which contribute to a detailed understanding of microbial dynamics (Cabezas *et al*., 2015). For instance, comparative genomics allows for the analysis of genetic differences

and similarities between various microbial strains, aiding in identifying specific traits that contribute to the efficiency of anaerobic digestion (Satam *et al*., 2023). Detection of rare taxa, facilitated by the high sensitivity of NGS, is crucial as these low-abundance microorganisms can have significant functional roles within the microbial community, potentially impacting the stability and performance of biodigesters. Functional gene analysis provides insights into the presence and expression of genes involved in critical processes such as cellulose degradation, hydrogen production, and methane formation, thereby uncovering the molecular mechanisms underpinning digestion.

NGS's high resolution, speed, cost-effectiveness, and ability to generate comprehensive data allow for in-depth analysis of complex microbial ecosystems. NGS data can be integrated with bioinformatics tools to predict metabolic pathways, model microbial interactions, and simulate system responses to various operational conditions (Zhang *et al*., 2019). This integration enhances our fundamental understanding of microbial ecology in biodigesters and informs the development of strategies for microbial management and process optimization. For instance, researchers can devise targeted bioaugmentation strategies to enhance biogas production by identifying vital microbial players and their functional roles. Furthermore, continuously monitoring microbial communities using NGS can provide early warning signs of process imbalances, enabling timely interventions to maintain optimal biodigester performance. Consequently, NGS technologies significantly enhance the ability to optimize and manage biodigesters, leading to improved biogas production, more efficient waste treatment processes, and greater overall sustainability in the management of organic waste.

2.4. Metagenomic approach for characterization of microorganisms in biodigesters

Metagenomics, a methodology employed for the direct examination of genetic material amassed from microbial communities inhabiting natural environments, offers insights into microbial diversity, structure, function, evolutionary relationships, and interactions (Mathieu *et al*., 2022; Kim *et al*., 2022; Niya *et al*., 2024). The efficacy of this approach hinges significantly upon the proper execution of sample extraction, a pivotal preliminary step for delineating microbial community structures from their habitats (Navgire *et al*., 2022; Liu *et al*., 2025). To avert the risk of erroneous outcomes and potential loss of pertinent information regarding microbial communities, it is imperative to meticulously select extraction methods commensurate with the characteristics of the sample under investigation (Upadhyay *et al*., 2023). Extraction of total microbial nucleic acid necessitates achieving high concentrations,

integrity, and a representative spectrum of all cells within the environment for subsequent library preparation and sequencing (Thomas *et al*., 2012, Fukuda *et al*., 2016). Maximizing DNA yield often mandates physical separation and isolation of cells from the sample, thereby dictating the choice between direct and indirect extraction methods (Thomas *et al*., 2012). In direct extraction, microorganisms are initially lysed via chemical and physical means to liberate total genomic DNA efficiently, albeit at the cost of DNA purity. Conversely, indirect extraction mandates prior separation and isolation of cells to procure high-purity nucleic acid.

Subsequent to DNA isolation, assessment of quality typically entails visual inspection of agarose gel electrophoresis bands, with quantification and purity evaluation performed via UV spectrophotometry at wavelengths of 260 nm and 280 nm utilizing instruments such as Nanodrop or qubit (Upadhyay *et al*., 2023). Pure DNA is subsequently utilized for DNA library preparation before sequencing (Zhang *et al*., 2021). NGS is emerging as the preeminent choice globally due to its cost-effectiveness and high throughput. The technology enables the parallel synthesis of millions of sequencing templates, facilitating comprehensive genomic analysis.

The choice between amplicon-based (targeted) and shotgun metagenomics hinges on research objectives (Yoshitake *et al*., 2021). Amplicon-based approaches, employing conserved taxonomic markers like 16S rRNA, facilitate taxonomic classification and pathogen detection but may introduce bias and overlook certain microorganisms (Navgire *et al*., 2022). Conversely, shotgun metagenomics, devoid of PCR bias, enables whole-genome sequencing for a holistic view of microbial communities, including novel taxa (Vanwonterghem *et al*., 2014). Advanced sequencing technologies have broadened the applicability of shotgun metagenomics, aligning data output with the complexity of microbial communities, even in environments like anaerobic digesters. Sequencing reads, typically 300 base pairs in length with NGS, undergo pre-processing and analysis using various bioinformatic tools post sequencing.

2.5. Bioinformatics analysis of metagenomic sequencing data

The advancement of methodologies for characterizing microbial communities within biodigesters has facilitated the integration of bioinformatics technologies into metagenomic investigations (Mathieu *et al*., 2022). The resultant sequence data is intricate, comprising billions of reads, necessitating computational pipelines and tools for extracting meaningful insights regarding the microbial consortia under examination (Achudhan *et al*., 2023). Initial quality assessment of metagenomic sequencing reads is imperative, encompassing evaluations of raw read count, read length, per-base sequence quality, GC content, and adaptor presence.

© University of Venda

Tools like Cutadapt or FASTQC are commonly employed for this purpose, with FASTQC being particularly recommended for metagenomic datasets (Almeida *et al*., 2019). Subsequent pre-processing involves the removal of adapters, chimeras, linkers, and non-specific sequences introduced during extraction and sequencing, typically executed using software like Trimmomatic and Sickle, with Trimmomatic being favoured due to its efficiency and reliability (Gao *et al*., 2021). The resultant trimmed reads are input for various taxonomic analysis tools such as KRAKEN2, metaphlAn2, and KAIJU, which utilize k-Mers to assign taxa based on available databases.

Bioinformatic analyses encompass a range of procedures, including Operational Taxonomic Units (OTUs) clustering analysis (for amplicon-based metagenomics), microbial diversity/similarity analysis, alpha diversity analysis, prediction of functional gene patterns (for shotgun metagenomics) **(Fig. 2)**. Each analysis requires corresponding tools and platforms, with integrated pipelines like Kbase, Galaxy, DRAGEN metagenomic, MG-RAST, UCHIME, QUIIME, ANASTASIA, SHAMAN, UCLUST-RDP, and MUTHOR being commonly employed, offering user-friendly interfaces for biologists with limited programming skills (Achudhan *et al*., 2023).

Software such as MOTHUR, MUSCLE, INFERNAL, and CLUSTALW are typically utilized for aligning multiple clean sequences in OTU analyses, followed by clustering at 97% sequence similarity (Upadhyay *et al*., 2023). USEARCH Software, UCLUST-RDP, and MEGA are frequently referenced tools for OTU analysis. Investigating microbial population diversity within biodigesters often employs OTU-based alpha diversity analysis, utilizing metrics such as ACE, Simpson diversity index, Shannon-Weaver diversity index, Chao1 richness estimator, and R software packages with the VEGAN library and RDP pipeline. Additionally, straightforward calculations using provided equations are often employed to estimate biological diversity rapidly. Furthermore, alpha rarefaction curves can be utilized to depict the relationship between species abundance and sequencing depth.

In the realm of bioinformatic analysis utilizing whole-genome shotgun data, the classification of taxa and functions can be accomplished through two main approaches: direct utilization of trimmed reads or through the de novo assembly of trimmed reads into contigs or scaffolds (**Fig.2**). Assembly is aided by software tools such as MetaSPADE, MEGAHIT, and Velvet (Zhang *et al*., 2019; Upadhyay *et al*., 2023). The assembly quality is evaluated using MetaQUAST, which furnishes metrics including contig count, length, GC content,

© University of Venda

misassemblies, and N50 value through comparative analyses with diverse databases (Olson *et al*., 2019).

To improve the precision of taxonomic and functional annotation, categorizing sequence fragments based on shared attributes such as GC content and coverage, a process termed binning, is indispensable. Notable tools tailored for binning procedures comprise MaxBin, Metagenome Binning with Abundance and Tetra-nucleotide frequencies (MetaBAT), CONCOCT, MetaWRAP, GroopM, and MyCC (Scholz *et al*., 2012; Olson *et al*., 2019). In this process, reads from individual or multiple samples are grouped back to their genomes. The former generates complex datasets when different strains of the same species are present, resulting in more fragmented sequences. Although the binning of individual samples provides high-quality genomes due to the samples' less complex state, the bins' processing is required before analysis. This can be achieved by tools such as dRep, which identifies the identical genomes and filters the best quality in each replicate set (Lindner *et al*., 2024). This results in generating metagenomic-assembled genomes (MAGs), representing discrete microbial populations. The generated MAGs provide maximum accuracy for downstream analysis. Challenges encountered during binning may stem from the intricacy of metagenomic data, encompassing fragmented sequences, uneven microbial representation, acquisition of partial or incomplete sequences, and Horizontal Gene Transfer (HGT) phenomena among microorganisms (Zhou *et al*., 2022).

For subsequent analysis of bins, genomes should attain a minimum completeness of ≥ 50% (indicative of the number of unique single-copy genomes within the bin) while maintaining ≤ 10% contamination levels (reflecting the estimation of single-copy genomes appearing in multiple copies). Tools such as CheckM serve this particular purpose effectively. Pertinent attributes of a queried genome can be discerned through the utilization of tools like Prokka, which integrates various existing software modules including prodigal (for Coding Sequence (CDS) prediction), RNAmmer (for Ribosomal RNA gene prediction), Aragorn (for Transfer RNA gene prediction), SignalP (for prediction of Signal leader peptides), and Infernal (for Noncoding RNA prediction), resulting in comprehensive prokaryotic annotation (Seemann *et al*., 2014).

RAST (Rapid Annotation using Subsystem Technology) also represents a web-based server facilitating gene prediction, functional assignment to predicted genes, and metabolic pathway reconstruction. Furthermore, it offers a platform for comparative analysis leveraging annotated

genomes curated within a SEED environment (Thomas *et al*., 2012). Some tools that can be used to gain insight into the microbial function and pathways include BLAST (Basic Local Alignment Search Tool) against databases COG, KEGG, MGRAST, MEGAN, and GeneDB. The length of the sequences of interest determines the tool to be utilized. For instance, BLAST is commonly used to annotate short contigs sequence of length, and the other tools can be used for long contigs (Zhang *et al*., 2019).

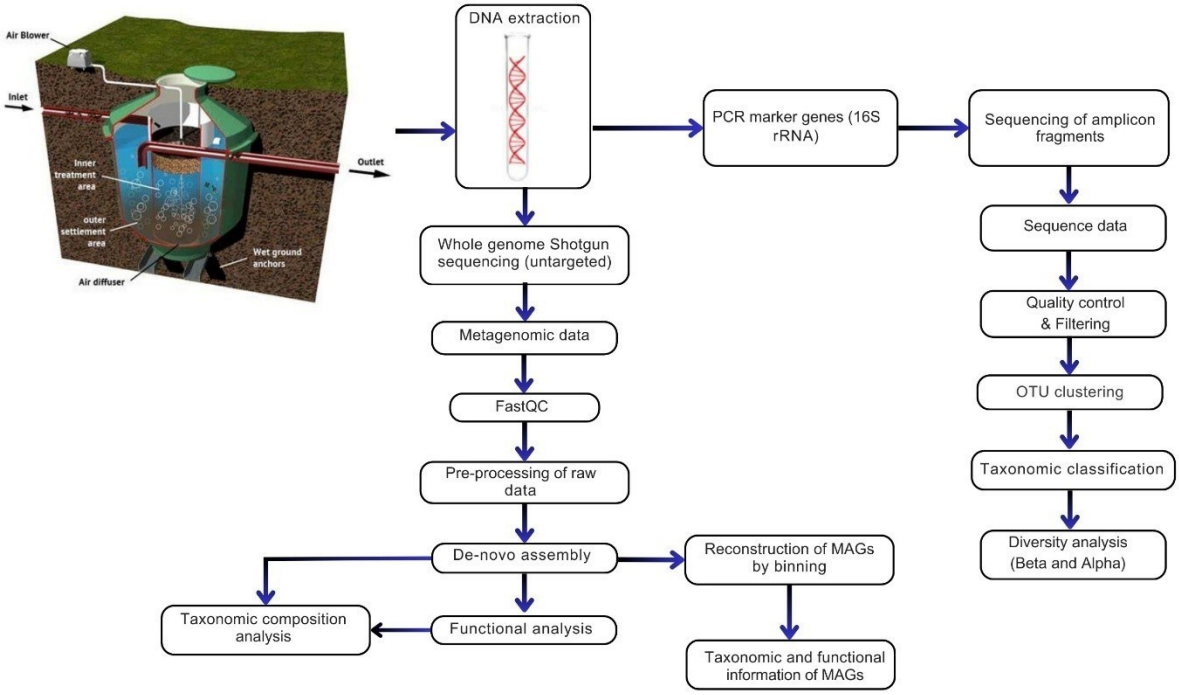


Figure 2: Comparison of two metagenomic approaches-Targeted Metagenomics (Amplicon Sequencing) and Whole Genome Shotgun (WGS) Metagenomics, along with their respective bioinformatics workflows.

2.6. Background of MEGAHIT, a metagenomics assembler tool

MEGAHIT is a fast, memory-requiring NGS metagenome assembler released in late 2014. The tool can efficiently assemble larger, more complex metagenomic datasets on a single server. It is a *de-novo* assembler, which means that the reads are assembled without the availability of the reference genome. It utilizes an algorithm known as de Bruijn Graphs (DBG), that is based on the composition of k-mers (all possible substrings of nucleotide bases of a certain length) (Li *et al*., 2016). In simple terms, metagenomic sequences are fragmented into short k-mer, which are then overlapped to form a network. This network is continuously navigated to form contigs (Van der Walt *et al*., 2017). Contigs are

contiguous sequences in which bases are known to have a high degree of certainty. This tool was used in the current study as suggested by Van der Walt *et al*., as it is inexpensive, has high accuracy, and few misassembles can occur compared with the other *de-novo* assemblers. In addition, it can assemble more complex datasets in less time.

2.7. Visualization of metagenomic data

The volume of data generated in metagenomic environments requires selecting appropriate tools for visually displaying these complex datasets to help researchers answer questions based on the research focus. Graphically displaying biological metagenomic datasets can be achieved via publicly available web-based software or a stand-alone application that requires knowledge of programming languages such as R and Python (Sudarikov *et al.*, 2017). Depending on the goal of the researcher, visualization can be applied to single or multiple metagenomes. Tools such as KRONA are usually used for taxonomic classification analysis of microbial populations in biodigesters, wherein different levels (domain, phylum, class, order, family, and genus) are visualized in a single diagram. Sometimes, a ternary plot is used to study the distribution of the unique and shared microbial taxonomic composition in different samples (Zhang *et al*., 2019).

© University of Venda

CHAPTER 3: MATERIALS AND METHODS

3.1. Ethical Clearance

Ethical Clearance was obtained from the University of Venda's human and clinical trials research ethical committee with an Ethical approval number, FSEA/24/BMY/23.

3.2. Geographic and status information of the biodigester site

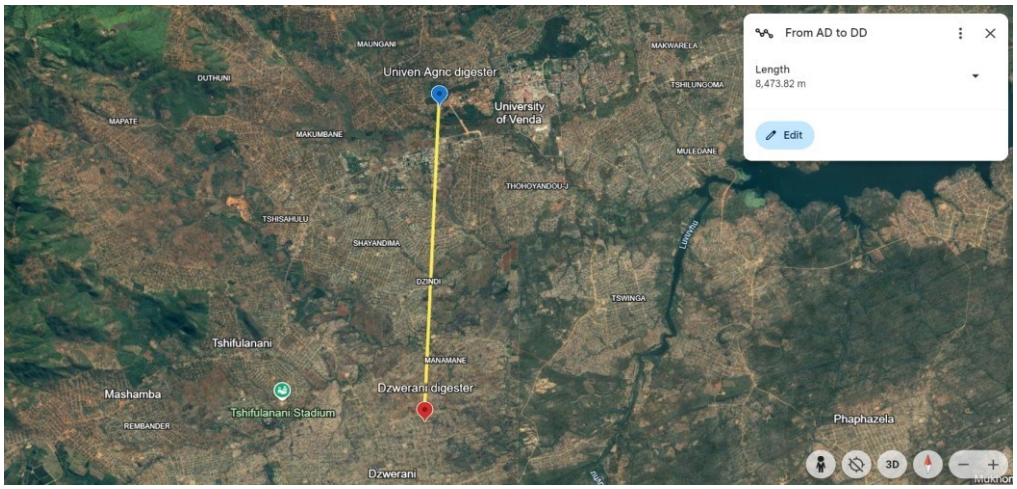
Two biodigesters used in this study, one was located in Dzwerani (23°03'20.3"S 30°26'00.8" E), a rural area, which is approximately 16.8 km away from the second digester located in the University of Venda (Univen) Agriculture farm (22°58'44"S,30°26'4"E) in Limpopo Province, South Africa as shown in Fig 3. The digesters had a capacity of 6 m^3 and generates methane gas for cooking purposes. Cow dung, used as the substrate for both digesters, was collected from free-ranging cows at the Dzwerani digester, while the other sample was obtained from cows used for research within the Univen Agriculture Farm. In both cases, the cows were grassing fed. The resulting sludge is utilized as biofertilizer for a small household farm nearby. Prior to commencing the project, all digestate content was extracted from both digesters. The removed digestate was then used as an inoculum to expedite biogas production. Initially, 4 x 20 L buckets of cow dung, mixed with water at a 1:1 ratio, were fed daily into the digester until biogas production commenced within 30 days. As the system began producing biogas, the quantity of feeding substrate decreased to 1 bucket per day.

© University of Venda

A



B.



C. **D.**





Figure 3: Sampling location and geographical co-ordinates (A) The location of the digesters in South Africa. (B) The geographical distance between the digesters is 8.473.82 m (C) Univen agric digester (D) Dzwerani digester. (A) **Univen Agric digester:** 22° 58' 44" S, 30° 26' 14" E (B) **Dzwerani digester:** (23°03'20.3"S30°26'00.8"E)

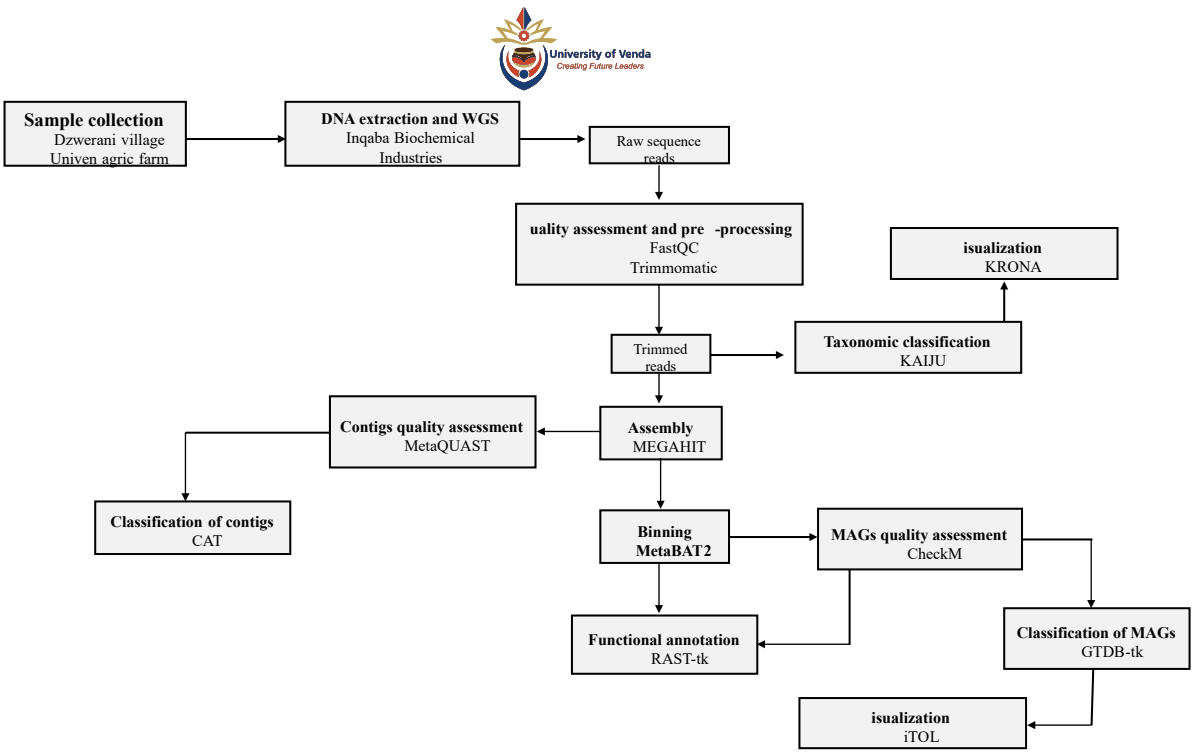


Figure 4: A representative workflow of sequencing and bioinformatics analysis**.**

3.3. Sample collection and whole genome sequencing

Digestate samples (two from the Dzwerani digester and three from the Univen agriculture digester) were collected from the outlet/slurry after the biodigesters had been reactivated for 30 days. After collection, the samples were kept chilled on ice and transported to the Biochemistry laboratory at the University of Venda. Upon their arrival, the samples were stored at -20 ℃. While chilled on ice, they were transported for genomic DNA extraction, library preparation, and shotgun sequencing using the Illumina NextSeq 500/550 system utilizing paired-end whole genome sequencing.

3.4. Pre-processing of raw metagenomic data

3.4.1. Quality control of the generated raw sequencing reads

The generated paired-end raw fast file sequencing reads were assessed for quality using FastQC (Quality Control) to check the per-base sequence quality, sequence composition, and GC contents.

3.4.2. Trimming of raw data

Chimera, adapters, and non-specific sequences were removed using the Trimmomatic tool in palindrome mode, primarily during the Illumina clip step. Truseq3 adapter sequences for

© University of Venda

paired-end Illumina NextSeq generated reads were removed in the initial step of trimming. Three operational parameters were employed to trim the reads post adapter removal: Sliding window trimming of the size of 4 with a minimum quality of 20, post tail crop length of 150, and trailing minimum quality of 10.

3.4.3. Taxonomic classification of reads

Filtered reads were assigned to their taxonomy utilizing Kaiju using the *RefSeq* Genomes (no euk) reference database. Kaiju is a tool that directly compares metagenomic pre-processed reads to the reference database of microbial proteins. The tool was used in the current study because of its sensitivity (it can detect distant homologies) as compared to other classifiers (Kraken, Clark) that are based on nucleotides (Menzel *et al*., 2016). Whole genome pre-processed paired-end reads were assigned to taxonomy at all levels (Superkindoms, phylum, order, family, class, genus, and species).

3.5. Taxonomic classification utilizing long sequence-based tools

3.5.1. Metagenomic sequence assembly

The clean paired-end reads were assembled using MEGAHIT. The quality of genome assembly was evaluated with meta QUAST (QUality ASssement Tool) in reference-independent mode. Various sequencing platforms generate reads of different lengths. For instance, the Illumina platform employed in this investigation produces fragmented short reads simultaneously. These reads necessitate assembly to reconstruct the genomes of microbial populations within the studied habitat. The resulting output, in the form of contigs, reflects the microorganisms present in the environment. Assembly methods can be categorized as either reference-guided or de novo. Reference-guided approaches utilize markers such as the 16S rRNA gene for gene amplification, followed by sequencing the resulting amplicon. Reference-based assembly relies on databases such as SILVA, Greengene, or the Ribosomal Database Project (RDP) to compare query sequences during the assembly process. De novo assembly, on the other hand, is an unbiased approach primarily employed when the sequence of interest is unknown. This method constructs assemblies from short reads without the assistance of reference genomes. It is significant as it enables the generation of novel sequence assemblies for previously uncharacterized genomes.

© University of Venda

a) Meta QUAST, as a quality assessment tool

It is a modified version of QUAST (QUality Assessment Tool) used to assess the quality of assemblies of metagenomic datasets by alignment with multiple close reference genomes. The tool automatically detects species by pulling the reference genomes from NCBI. MetaQUAST can detect chimeric contigs and present them as misassembles (MA) because of closely related species.

b) Taxonomic classification of generated contigs

The generated contigs were classified into their taxonomy using the Contigs Annotation Tool (CAT). In the initial step, the Open Reading Frame (ORFs) from contigs are predicted using Prodigal. Secondly, the protein translations of the ORF are compared to the non-reductant set of proteins in the NCBI database. This is done to identify the top hits (≥90% of the highest bit score of that ORF). Default settings are employed for homology search with DIAMOND. The taxonomic annotation of top hits is compared to derive individual taxonomy of ORF according to the determination of their Last Common Ancestor (LCA) in NCBI (Cambuy *et al*., 2016). The tool was used in the current study because of its robustness and accuracy in the taxonomy classification of longer metagenomic sequences (contigs).

3.6. Metagenomic binning, assessment of bins quality, and construction of MAGs

All contigs of ≥ 500 bases that belong to the same biological taxon were mapped into their respective groups using MetaBAT2, and the quality of the resulting bins was assessed using CheckM lineage_wf. The binning approach used in this study was unsupervised, meaning that contigs mapping does not rely on the availability of a reference genome. Binning aims to reconstruct draft genomes from assemblies or contigs.

(a) Grouping of contigs using MetaBAT2

MetaBAT2 is a fast and robust binning tool that uses both hybrid compositional and depth of coverage of sequence reads Guanine-Cytosine (GC contents) or an abundance approach to successfully assign contigs to the taxonomic group they belong to. MetaBAT2 produces accurate and high-quality genome bins on a large assembly of contigs within a short period.

(b) Quality assessment of bins using CheckM

CheckM was used in the current study to assess the quality of the generated bins by comparing the genomic bins to data sets containing genomic information of universal single copies marker

genomes for prokaryotes, specifically bacteria, and archaea. The representative genomes in bins were selected by CheckM filter tool in kbase platform. This was done by filtering only medium-high bins with at least ≤ 10% contaminations (estimation of how many single-copy genomes are present in multiple copies) and ≥ 50 % completeness.

(c) Dereplication of the generated bins

For accurate downstream analysis, the generated bins were dereplicated using the dRep tool using default settings, a robust tool that identifies identical genomes from a larger dataset and determines the genome with the highest quality. It utilizes two algorithms, MinHash (MASH) and Average Nucleotide Identity (ANI), for primary and secondary clustering at 90% and 99%, respectively, to decrease the time required to dereplicate bins while ensuring the generation of high-quality MAGs (Olm *et al*., 2017). Medium and high MAGs with at least ≤ 10% contaminations and ≥ 50 % completeness were retained and used for downstream analysis.

(d) Taxonomic classification of MAGs

The dereplicated MAGs were classified using the Genome Taxonomy Database Toolkit (GTDB-Tk). The tool first predicts the ORFs of the dereplicated MAGs using prodigal, then the resulting bacterial and archaeal single copy genes were retrieved and aligned using GTDB database. The concatenated alignment was then used to build a phylogenetic tree. An effective toolkit designed to classify bacterial and archaeal genomes by automatically placing them into specific domain protein reference trees. For taxonomic classification to be successful, the Metagenomic Assembled Genomes (MAGs) in question need to be of medium-high quality (consistent with community standards) (Chaumeil *et al*., 2020).

3.7. Unveiling the function capability of microbial communities in small-scale biodigesters

The best representative MAGs (high-quality), with ≥90% completeness and ≤ 5% contamination for both *Eurychaeta* phylum representing archaeal microbes and the most abundant bacterial MAG were subjected to Rapid Annotation using Subsystem Tool (RAST) server to identify anaerobic digestion phase-related genes methanogenesis-related genes.

3.8. Visualization of metagenomic data

KRONA was used to visualize the abundance of taxonomic groups of microbial communities of pre-processed reads. This visualization tool allows exploration (in hierarchical order) of

relative abundance and confidence within complex microbial communities classified (Ondov *et al.*, 2011). It allows a deeper understanding of the underlying classification. Interactive Tree of Life (iTOL) was used to visualize the generated phylogenetic trees from GTDB-Tk.

© University of Venda

CHAPTER 4: RESULTS

Whole Genome Shotgun (WGS) metagenomic sequencing was performed on five digestate samples collected from two small-scale biodigesters that produced biogas solely from cow dung. These biodigesters were located at Dzwerani and Univen Agriculture farm, Limpopo, South Africa. The digestate samples were labelled as Dzwerani Digester (DD) and Agriculture Digestate (AD), respectively. Microbial communities' taxonomic and functional potential in small-scale biodigesters were integrated to identify the essential microbes responsible for the anaerobic biodegradation of free-ranged cow dung and the subsequent production of biogas.

4.1. Whole genome shotgun (WGS) sequencing statistics

Sequencing was conducted using the Illumina NextSeq500/550 system. Altogether, 5 samples were analyzed in this study. On average, each sample produced 2.1 Gigabyte of data, equivalent to approximately 10-13 gigabases, with an approximate sequencing depth of 7 million raw reads for samples collected from a digester situated at Dzwerani and an average of 5 million raw reads, giving 33, 993, 584 million base pairs raw reads. All the samples had an average length of 35-151 base pairs, with the GC content ranging between 54-57 % (**Table 1**).

Table 1: Statistics of generated raw reads after sequencing with Illumina Nextseq500/550 system

	DD1	DD2	AD1	AD2	AD3
Generated raw reads	7,737,741	7,584,147	5,926,906	5,961,985	6 782 805
Length of sequences	35-151	35-151	35-151	35-151	35-151
Mean GC (%)	57	57	57	54	57

4.2. Whole-genome shotgun raw sequence read quality assessment using FastQC

Given the high rates of contamination and error linked to sample preparation and NGS, it is essential to evaluate and pre-process the resulting raw sequencing reads before proceeding with analysis. The present study utilized Illumina sequencing technology to generate paired-end reads with a total base pair length of 300 (2x150 each). FastQC, a quality control tool, was employed to evaluate the quality of the raw whole genome shotgun sequences obtained. This step is essential for downstream analysis, particularly for ensuring a high-quality metagenomic

© University of Venda

dataset that accurately reflects the abundance and diversity of microbial populations within the sample. FastQC assesses various aspects of the sequences, including total sequences generated, per-base sequence quality, GC content base composition, and read length. Evaluation of reads using FastQC is presented using three colors; green, yellow, and red to denote normal, abnormal, and unusual outcomes, respectively. This aids in quickly identifying potential issues with NGS data, allowing for pre-emptive resolution before proceeding with bioinformatics analysis. Per-base sequence quality analysis (**Fig. 5**, **Fig. 6** illustrates the quality scores on the y-axis and individual or position of bases on the x-axis.

4.3. Trimming and removal of adapters from raw sequences

Although no sequence was classified as poor quality, trimming and eliminating technical sequences, such as adapters and Illumina non-specific sequences (contaminants) introduced during sampling, extraction, and sequencing steps, is recommended. This process aims to acquire pure metagenomes containing only the essential sequences and mitigate challenges such as misassembling of reads, which could potentially compromise downstream analyses. In addition, it is advisable to trim them due to their lengths ranging from 35 to 151 base pairs, aiming to standardize the reads to 150 base pairs per sequencing depth of Illumina. The quality of sequences should meet a threshold of ≥ 20% at each base pair position. The number of base pairs drastically decreased after trimming, which may be attributed to a higher proportion of adapter contamination. Consequently, more reads and bases were removed, leading to a substantial reduction in total base pairs (**Fig. 5**). In contrast, some sequences from the agricultural digester fell below the recommended threshold before trimming, but few base pairs were excluded after trimmomatic (**Fig. 6**). This might be due to clean sample library preparation, which resulted in the introduction of fewer adapter sequences and consequently minimal loss of bases during trimming. Following Trimmomatic execution, 30,199,322 paired-end reads (R1 for forward and R2 for reverse) were obtained, and reads for samples collected at Dzwerani ranged between 7,517,625 and 7,679,092. Samples collected at the Univen agriculture farm biodigester ranged between 4,654,704 and 5459864 (**Table 2**). Additionally, 4,061,773 sequences had lengths less than 150 base pairs and were consequently excluded. Approximately 30,000 for each sample was filtered out in the pre-processing step due to low quality, reducing the average length to 150 base pairs. Most of the reads had a GC content of 57. Only one sample had a GC of 54.

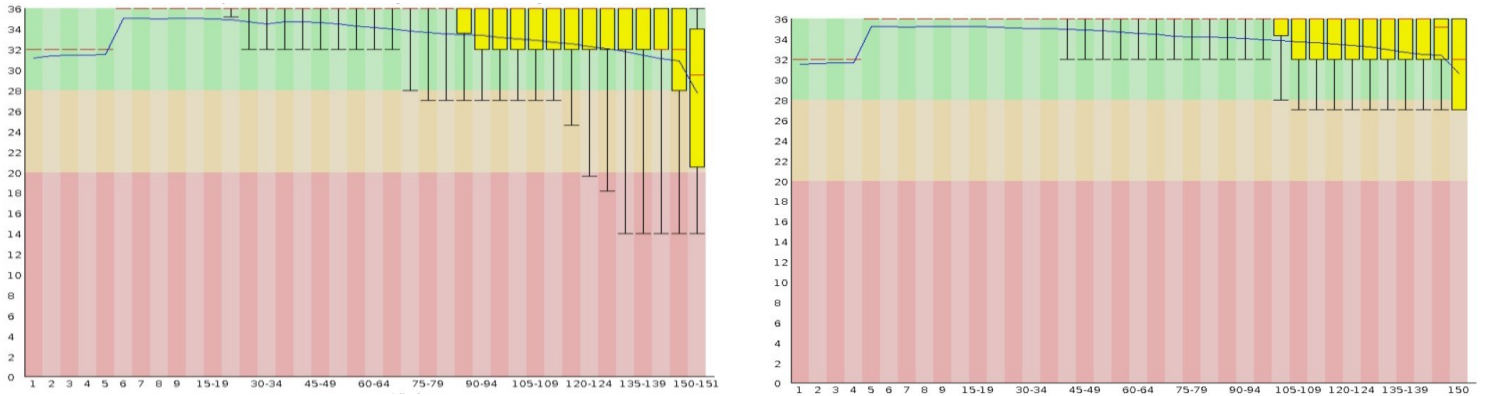
© University of Venda

Table 2: Quality assessment of reads generated before and after pre-processing using Trimmomatic.

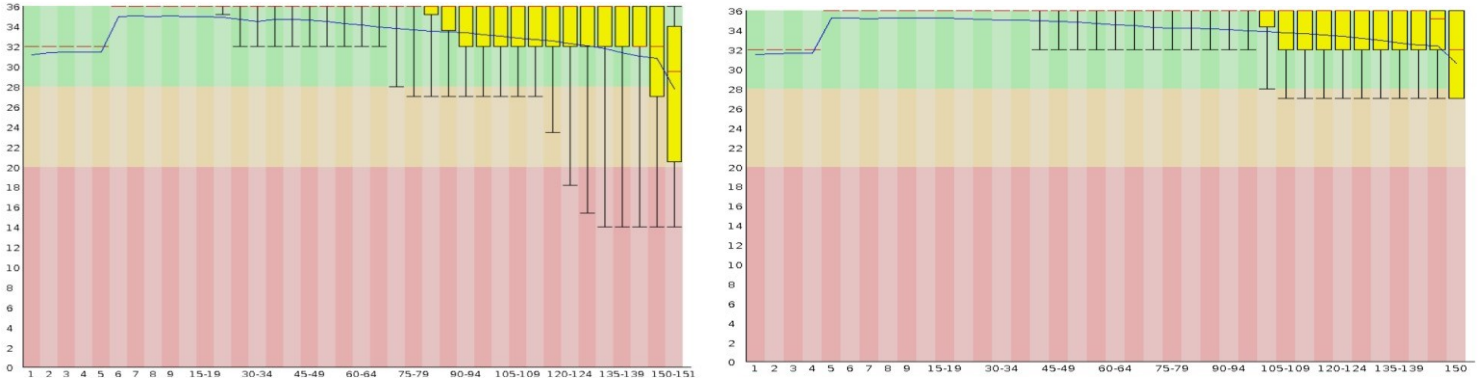
	DD1	DD2	AD1	AD2	AD3
Generated raw reads	7,737,741	7,584,147	5,926,906	5,961,985	5,961,985
Reads after trimming	7,679,092	7,517,625	5,895,426	5,738,697	5,927,914
Reads excluded	58,649	66,522	31,480	223,288	34,071
Mean GC	57	57	57	54	57

© University of Venda

A) DD1



B) DD2

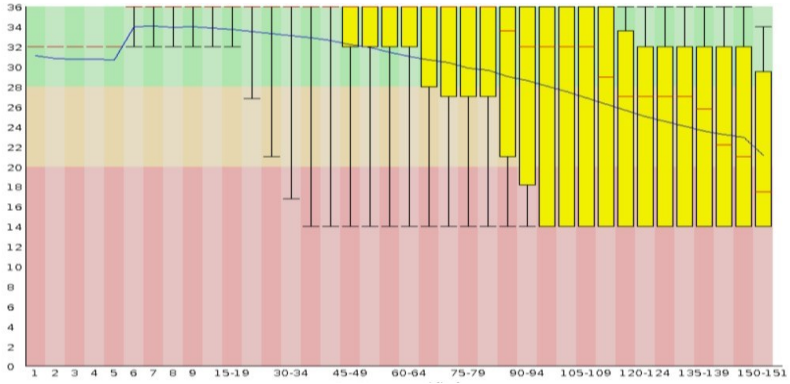


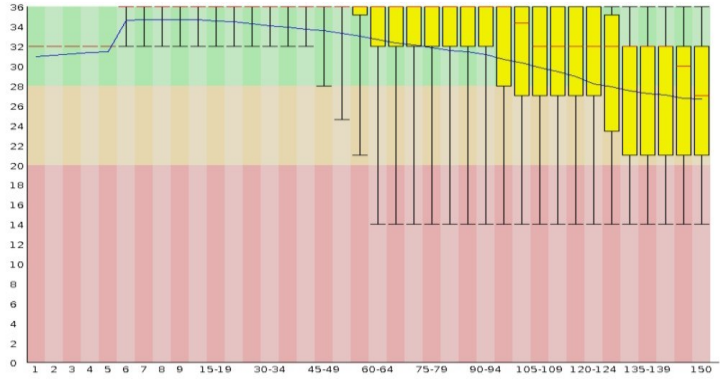
Position of base pairs

Figure 5: FastQC (Quality Control) statistics before **(left)** and after (**right**) trimming for Dzwerani samples. (A) DD1 (B) DD2 (C) DD3.

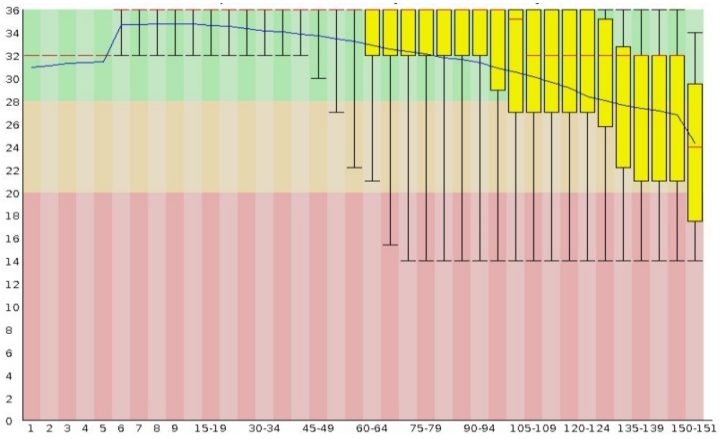
© University of Venda

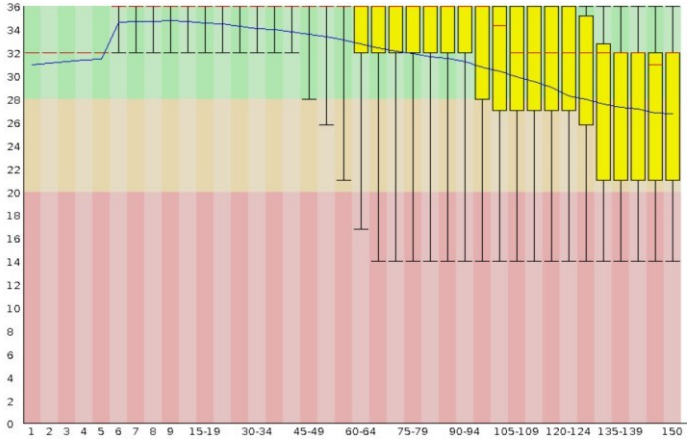
D. AD1





E. AD2





© University of Venda

F. AD3

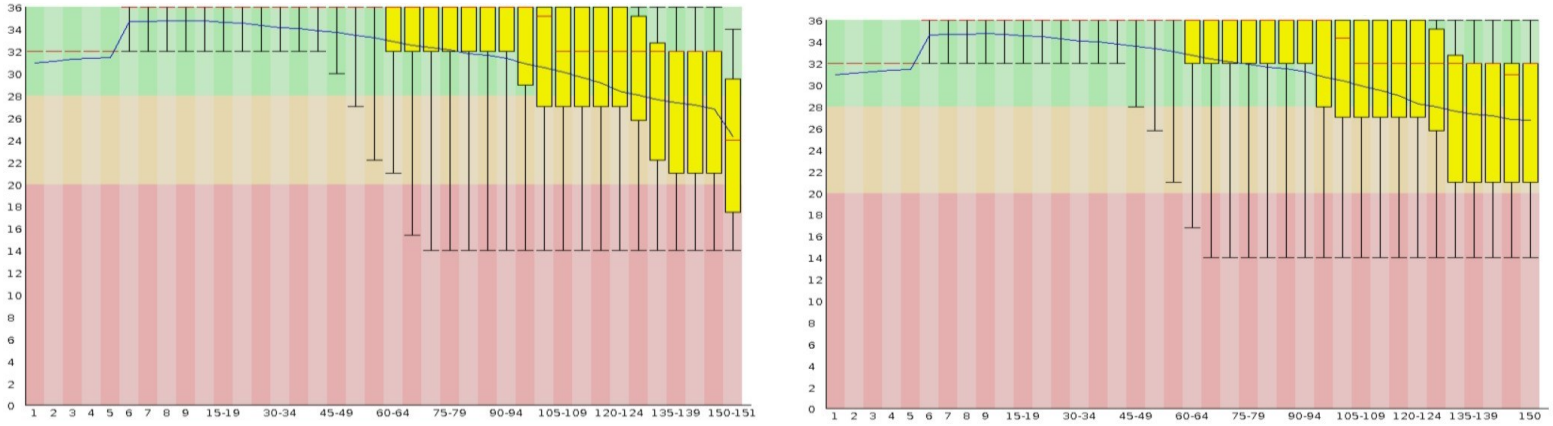


Figure 6: FastQC (Quality Control) statistics before (**left**) and after (**right**) for Univen Agriculture samples. (**D**) AD1 (**E**) AD2 (**F**) AD3

© University of Venda

4.4 Taxonomic profile utilizing read-based tool Kaiju

It is essential to perform taxonomic classification to elucidate the diverse microbial communities present in biodigesters. This approach enables the investigation of microbial functions and interactions, thereby revealing the mechanisms by which these microorganisms biodegrade the substrate (cow dung in our study). The taxonomic diversity of microbes from metagenomic reads generated from digestate samples collected at Dzwerani and Univen agricultural farm biodigesters (operating under normal environmental conditions) was assessed using a rapid taxonomic classifier, Kaiju.This is tool that assigns metagenomic pre-processed reads by direct comparison to the reference comprising the annotated protein-coding genes of a set of microbial genomes. At the domain level, all sequenced microbial communities were predominantly composed of bacteria (79-89%) and archaea (11-20%) (**Fig. 7**). High throughput whole genome shotgun (WGS) sequencing allowed for the identification of seven abundant bacterial phyla in all the samples; *Proteobacteria*, accounting for 30-37%, *Firmicutes*, representing 10-17%, and *Bacteroidetes*, comprising 7-14%, *Chloroflexi* (4-6%), *Plantomycetia* (3-4%), *Actinobacteria* (7-17%) and *Acidobacteria* (1%) which are commonly known as biodigester microorganisms (**Fig. 8**). The most predominately archaeal phylum present was *Euryarchaeota* (11-20%) in all datasets. Interestingly, some phylum, such as *Candidatus Thermoplasmatota* (0.1-0.6%) and *Candidatus Prometheoarchaeum* (0.1%), were also identified in small percentages of the archaeal phylum. Other metagenomic reads were classified as other archaea (0.02-0.05%). The analysis of the identified microorganisms revealed 18 distinct classes, 30 orders, 28 families, and 18 genera (**Table 3**). The study identified 56 bacterial and archaeal species, with the majority being bacterial species. This aligns with expectations, as bacteria are primarily responsible for the initial three phases of AD.

© University of Venda

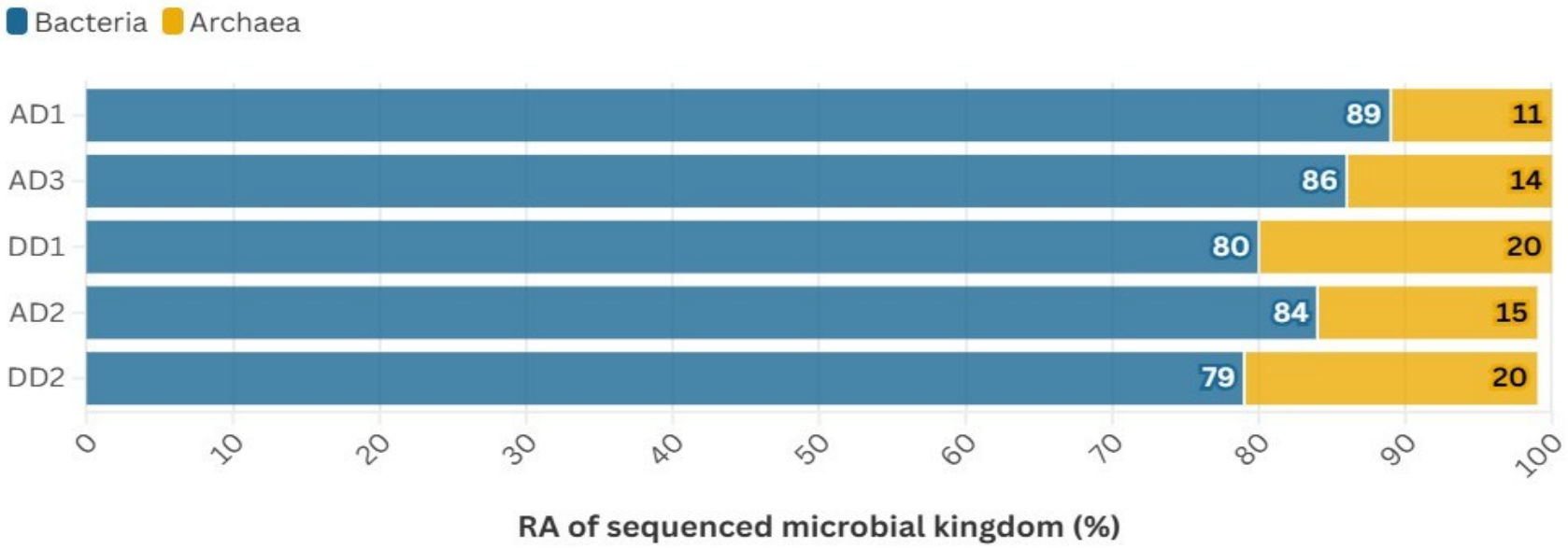


Figure 7: The microbial kingdom's relative Abundance (RA) in small-scale biodigesters. The taxonomic assignment was obtained from whole genome shotgun metagenomic sequence analysis. Bacteria were more dominant than archaea in all samples. The remaining 1% in DD2 and AD2 samples belonged to viruses and fungi. DD and AD denote Dzwerani and Agriculture digesters, respectively.

© University of Venda

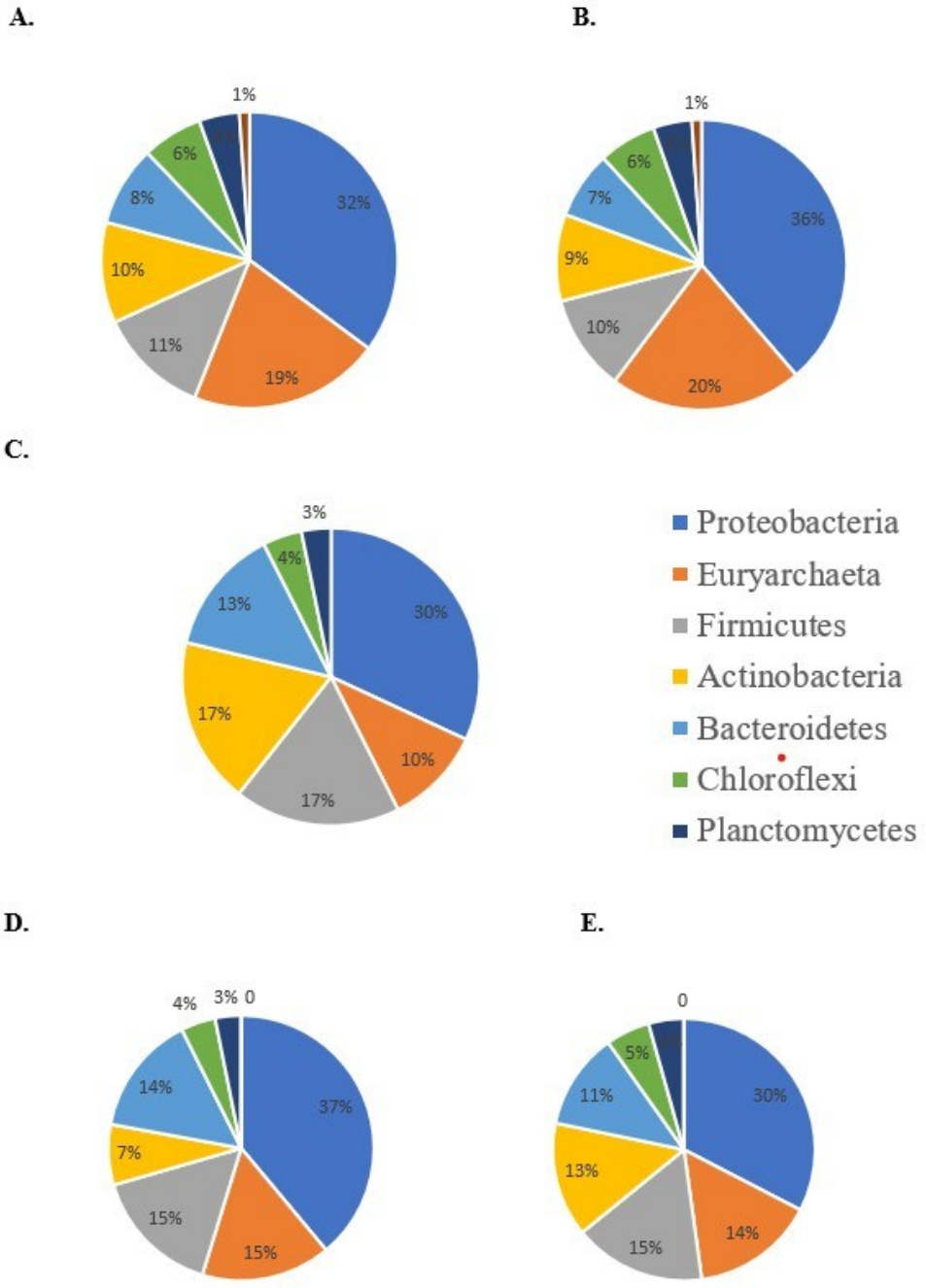


Figure 8: Microbial communities at a phyla level present in digestate of samples **A)** DD1, **B)** DD2, **C)** AD1, **D)** AD2, **E)** AD3 derived from small-scale biodigesters. KAIJU recovered a high percentage of Proteobacteria, followed by *Firmicutes* in all samples as compared to other phylum.

© University of Venda

4.5. Classification of microbial communities based on read-based tool

In the microbial communities of the two small-scale biodigesters analyzed, bacteria constituted the predominant group, surpassing archaea in abundance. This observation is consistent with the expected role of bacteria in driving the initial stages of anaerobic digestion, specifically hydrolysis, acidogenesis, and acetogenesis. At the phylum level, members of the *Proteobacteria* were the most prevalent. Taxonomic analysis at the class level identified 18 major bacterial classes in which the microbial community was abundant in the following 10 classes: *Gammaproteobacteria* (12-22%), *Betaproteobacteria* (2-8%), *Alphaproteobacteria* (3-8%), *Deltabproteobacteria* (4-8%) *Bacilli* (4-11%), *Actinobacteria* (6-16%), *Clostridia* (56%), Planctomycetia (2-3%), *Bacteroidia* (2-5%), and *Methanomicrobia* (7-14%) (**Fig. 9**). Among these, *Gammaproteobacteria* emerged as the most dominant class, accounting for 1822% of the microbial composition in the Dzwerani digester (DD) and 12-20% in the agricultural digester (AD).

The abundance of main order observed in our work representative at the order level identified 21 bacterial orders: *Burkholderiales* (1-7%), *Corynebacteriales* (0.01-2%), *Bacillales* (3-9%), *Eubacteriales* (4%), *Flavobacteriales* (1-8%), and *Bacteroidales* (2-3%), Methanomicrobiales (0.6-2%), Desulfobacteriales (0.9-2%), *Desulfuromonadales* (0.7-2%), *Pseudomonadales* (2-9%), *Pirellulales* (1-2%), *Lactobacillales* (0.5-3%), *Micrococcales* (3-10%), *Cytophagales* (12%), *Methanobacteriales* (3-6%), *Xanthomonadales* (0.4—2%), *Rhodocyclales* (1%), *Sphingomonadales* (0.4-2%), *Hyphomicrobiales* (2-4%), *Propionibacteriales* (1%), *Rhodobacteriales* (1%) (**Fig. 10**). Notably, *Eubacteriales* was absent in the microbial community of the Dzwerani digester, consistent with its low abundance in other samples. This absence may be attributed to differences in conditions such as temperature and pH, as well as variability introduced during sample handling and sequencing.

Further taxonomic analysis identified 17 families, of which 14 were bacterial and 2 were archaeal. The archaeal families found across all samples included *Methanobacteriaceae* (3–6% relative abundance) and *Methanosarcinaceae* (1–2%). The bacterial families identified were diverse. These included *Xanthomonadaceae* (0.3–3%), *Methylococcaceae* (0.2–1%), and *Anaerolineaceae* (2–4%). Other families observed were *Micrococcaceae* (0.6–4%), *Planococcaceae* (0.2–2%), and *Bacillaceae* (1–5%). In addition, *Clostridiaceae* (0.9–1%), *Sphingobacteriaceae* (0.4–1%), and *Weeksellaceae* (0.2–3%) were present. The analysis also revealed *Moraxellaceae* (3–14%), *Syntrophaceae* (0.7–1%), and *Pseudomonadaceae* (2–9%). Finally, *Cellulomonadaceae* (1–2%) and *Flavobacteriaceae* (1–5%) were identified (**Fig.**

© University of Venda

11).*Methylococcaceae* showed 1% abundance only in sample AD2, while *Xanthomonadaceae* was similarly abundant (1%) in sample AD3. Even though both samples were collected from the same biodigester, the differences in microbial composition could be related to the time of sampling, with one sample possibly taken during the hydrolysis phase and the other during acetogenesis. Alternatively, differences may result from competition for substrates within the biodigester. *Moraxellaceae* was the most dominant family across all samples, followed by the two archaeal families, with *Pseudomonadaceae* being the fourth most abundant.

© University of Venda

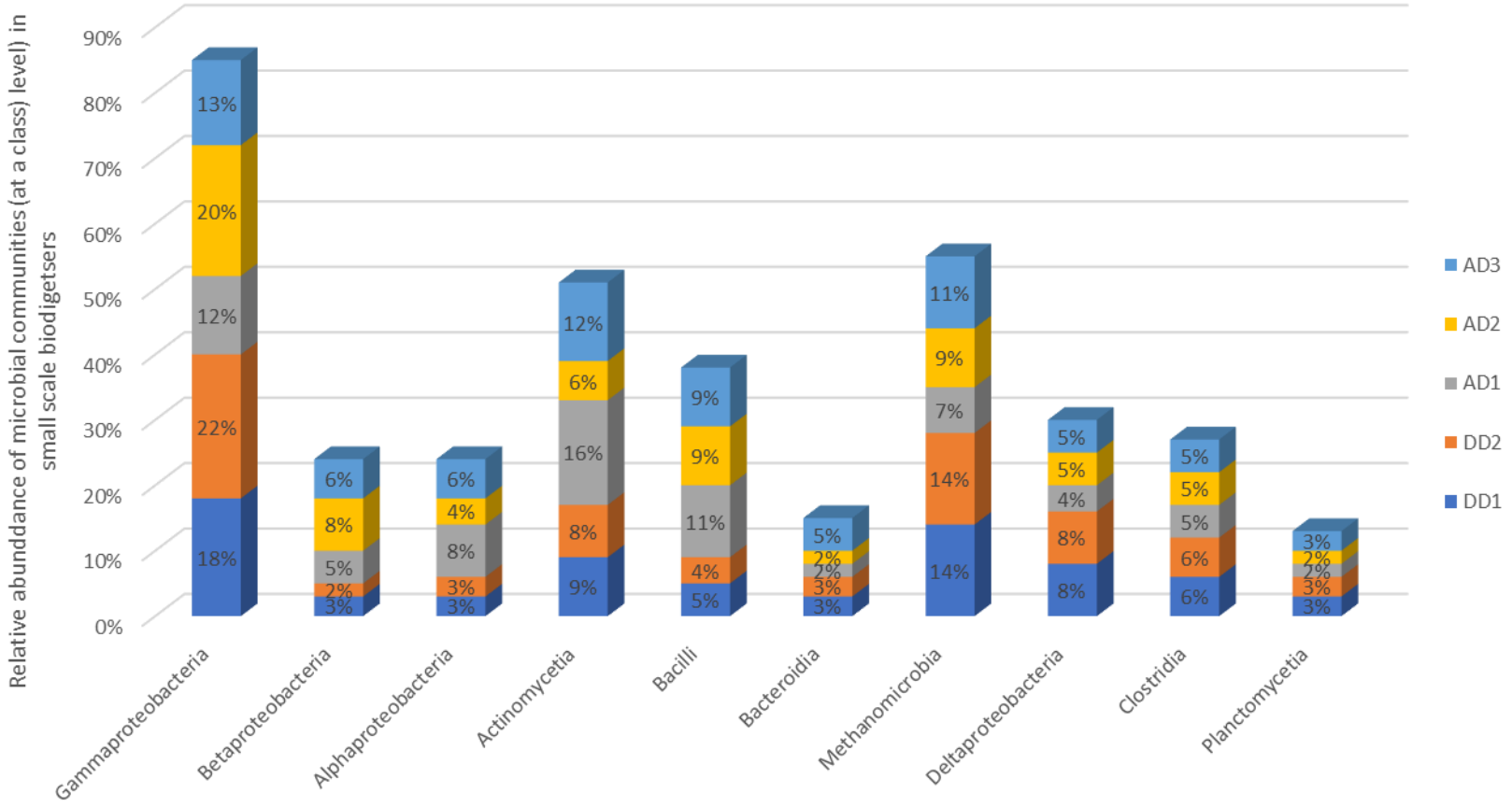


Figure 9: Relative abundance of microbes at the class level of taxonomic classification

© University of Venda

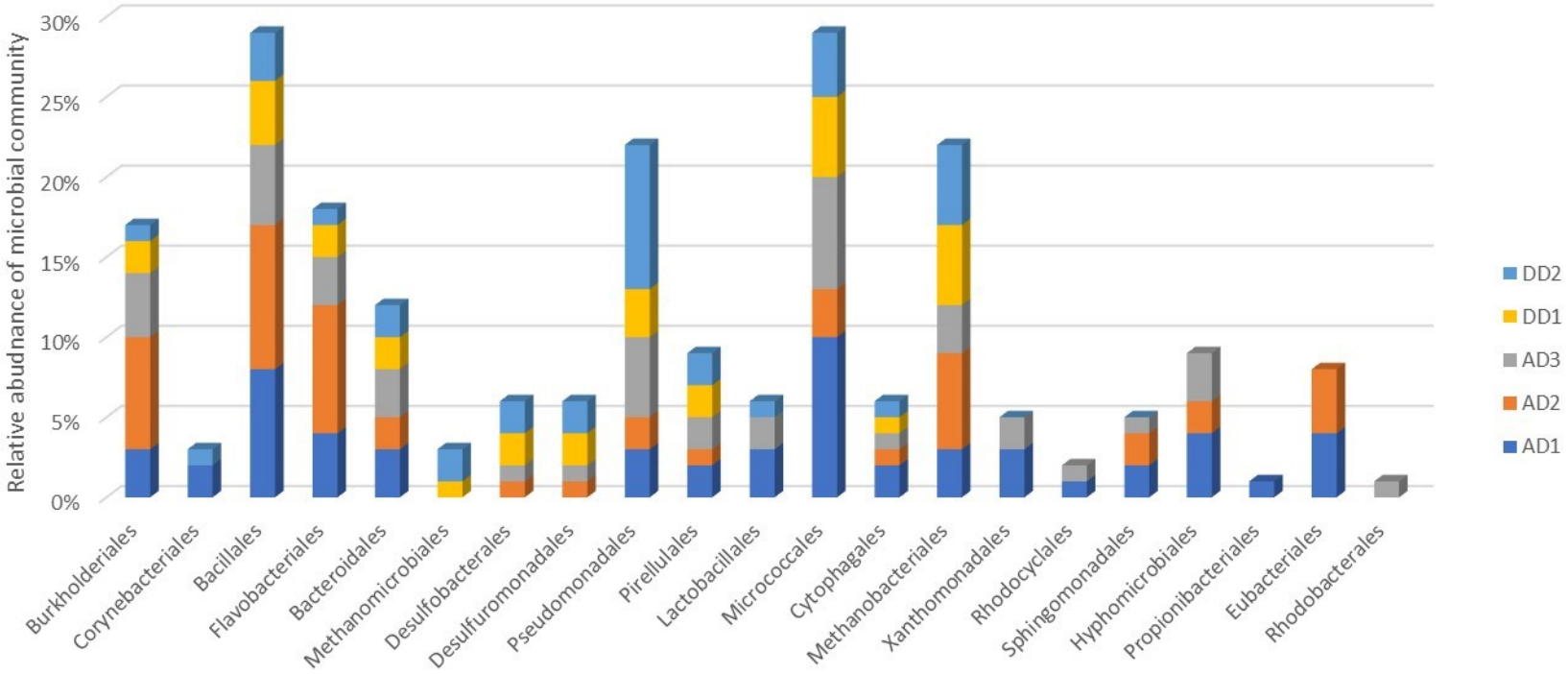


Figure 10: Relative abundance of microbes at the class level of taxonomic classification**.**

© University of Venda

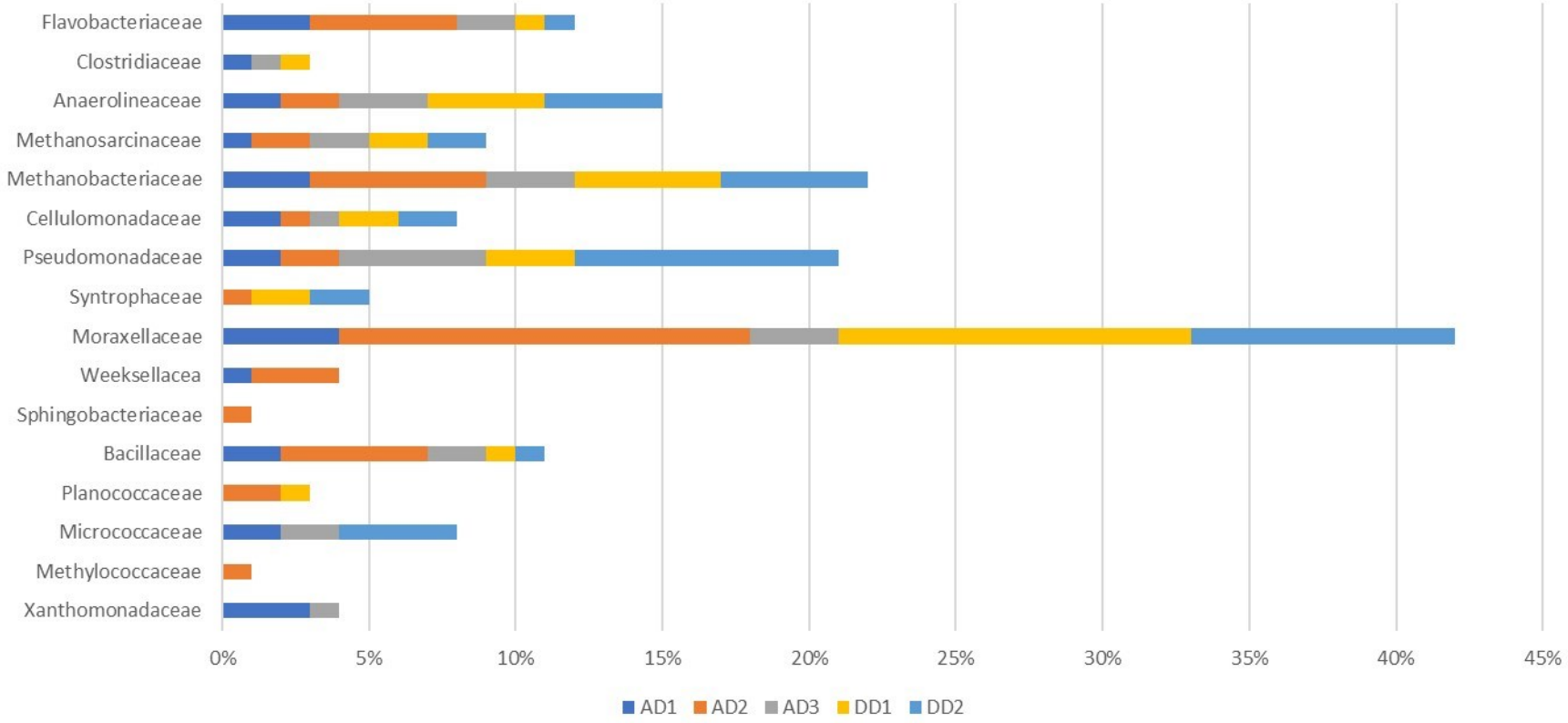


Figure 11: Relative abundance of bacterial and archaeal microbial (at family level) in Dzwerani digester (DD) and Agriculture digester (AD) samples.

© University of Venda

Table 3: Overview of relative abundance of microbial communities (excluding phyla, genera)

	Class	Order	Family
Number detected	10	21	17
Bacteria	*Gammaproteobacteria.*, *Betaproteobacteria.*, *Deltaproteobacteria.*, *Alphaproteobacteri.*, *Bacilli.*, *Actinomycetia* *Bacteroidia.*, *Clostridia.*, Planctomycetia	*Burkholderiales.*, *Corynebacteriales.*, *Bacillales.*, *Eubacteriales.*, *Flavobacteriales.* *Bacteroidales.*, *Desulfobacterales.*, *Desulfuromonadales.*, *Pseudomonadales.*, *Pirellulales.*, *Lactobacillales.*, *Micrococcales.*, *Cytophagales.*, *Xanthomonadales.*, *Rhodocyclales.*, *Sphingomonadales.*, *Hyphomicrobiales.*, *Propionibacteriales.*, *Eubacteriales.*, *Rhodobacterales*	*Xanthomonadaceae.*, *Methylococcaceae.*, *Anaerolineaceae.*, *Micrococcaceae.*,*Planococcaceae* *.,Bacillaceae.*, *Clostridiaceae.*, *Sphingobacteriaceae.*, *Weeksellaceae.*, *Moraxellaceae.*, *Syntrophaceae.*,*Pseudomonadaceae.*, *.,Cellulomonadaceae.*, *Flavobacteriaceae*
Archaea	*Methanomicrobia*	*Methanomicrobiales*	*Methanosarcinaceae.*, *Methanobacteriaceae*

4.6. Reference-independent whole genome assembly of filtered reads

The current study used MEGAHIT, a de-novo assembler tool, to reconstruct genomes from paired-end short Illumina-generated reads. In samples collected at Dzwerani, DD1, and DD2, 333,575 contigs were generated with an average N50 value of 1,032. The minimum contig length for all the generated contigs was ≥ 500. Sample AD2 had the highest number of contigs, and the lowest was 57,959 for the AD3 sample (**Table 4**). The N50 value represents the length of the shortest contig such that 50% of the total sequence is contained within contigs of that length or longer. In contrast, when combined, the L50 value denotes the number of contigs needed to cover at least 50% of the total assembly length. These are the standard metrics used to assess the quality of the assembly. Contigs generated for samples DD1, DD2, and AD2 were more contiguous than the other two samples (AD1 and AD3) because of the higher N50 (greater than 1000) and higher L50. This also suggests that the assembly was closer to the potential actual genome. Assembly for samples AD1 and AD3 were fragmented or had a high number of incompleteness or highly repetitive genomes, as the N50 metric was less than 1000.

Table 4: MetaQUAST (QUality ASsessment Tool) statistics for the sample assemblies generated using the de novo-based tool MEGAHIT.

	DD1	DD2	AD1	AD2	AD3
Generated contigs	66536	68282	61809	78992	57956
Largest contig length	135310	79846	94378	136925	68437
N50 value	1196	1143	872	1086	863
L50 value	12907	14456	18419	17791	17122
Mean GC	57.84	58.12	58.22	54.42	58.72

4.7. Taxonomic assignment of contigs

Due to the higher number of contigs generated for sample AD2, it was further subjected to the Contigs Annotation Tool (CAT) using the build-in CAT database (CAT_prepare20210107) in the galaxy for metagenomic to unveil the taxonomy of microbial communities. The analysis identified four superkingdoms: bacteria, archaea, viruses, and eukaryotes. Due to the observed similarity in identified taxa between samples collected from the Dzwerani and Univen

© University of Venda

Agriculture digesters using the read-based tool, further analysis was conducted on the AD2 sample, as it produced a higher number of contigs. As observed when utilizing a read-based tool, bacteria were the most abundant superkingdom with 3846666 hits, followed by archaea (48355). The least abundant were Viruses (289) and eukaryotes (71). According to the hierarchy taxonomy from the phylum level, a high number (115) of microbes were identified. There was a decrease in the class level. The number of microbes classified on the order level moving to the species level increased (**Table 3**). This study unveiled 2045 species inhabiting the small-scale biodigesters in rural areas. Due to the high number of taxa identified, only the relative abundance of each taxon was analyzed. On the phylum level, the first five abundant were bacterial phyla: *Proteobacteria* (333649), *Actinobacteria* (189797), *Firmicutes* (106320), *Bacteroidetes* (88630), *Chloroflexi* (60633). As observed with the Kaiju classifier, *Euryarchaeota* phyla, with 38249 hits, was the most abundant. Surprisingly, archaeal microbes belonging to *Candidatus Bathyarchaeota* phyla were also plentiful (648 hits) (**Fig. 12**). Similar results were not observed in the read-based analysis. *Proteobacteria* were represented by *Gamma*-(113183 number of hits), *Delta*- (31713 number of hits), *Beta*- (25326 number of hits), and *Alphaproteobacteria* (17480 number of hits) classes, all in the top 7 representatives of microbial communities in small scale biodigesters. The *Bacilli* class, known for its role in the phases of anaerobic digestion, was also observed in abundance in the present study. Two methanogenic classes, *Methanobacteria* and *Methanomicrobia*, with 17729 and 15725 hits, respectively, were identified (**Fig. 13**). This was also observed in read-based tool analysis. At the order level, the top 20 microbial communities by relative abundance were identified, with 17 being bacterial and 3 belonging to archaeal orders. Among the bacterial orders, *Pseudomonadaceae* were the most abundant, while *Methanobacteriaceae* were the dominant archaeal order (**Fig. 14**).

© University of Venda

Table 5: Overview of the number of microbes identified based on taxonomic hierarchy utilizing long read-based classifiers (CAT).

Superkingdoms	4
Phylum	115
Class	106
Order	188
Family	320
Genus	737
Species	2045

© University of Venda

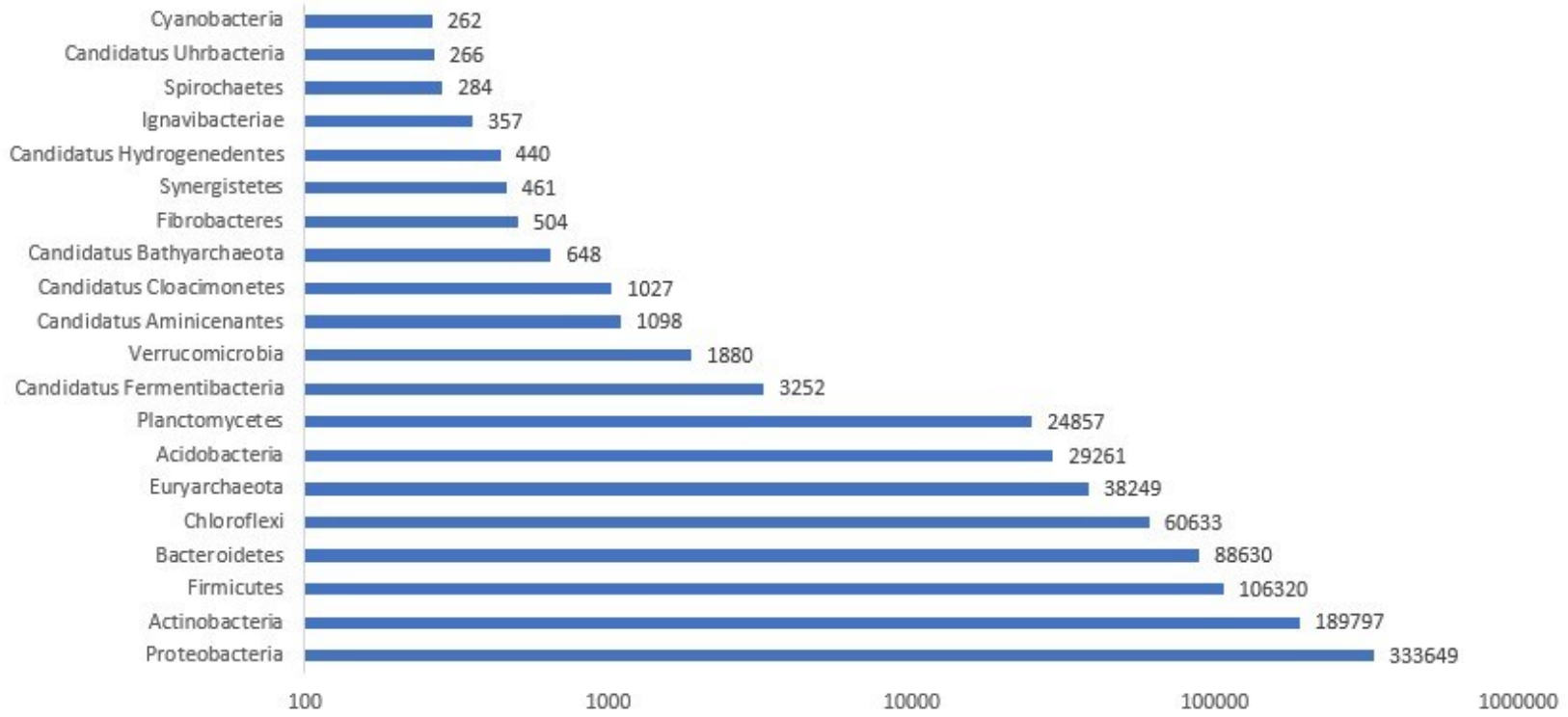


Figure 12: Microbial community composition at the phylum level based on CAT (Contig Annotation Tool) analysis. The clustered bar chart depicts the number of hits detected for each microbial phylum, indicating their relative abundance in the dataset. The x-axis represents the number of hits on a logarithmic scale, while the y-axis shows the different microbial phyla (top 20) identified. *Proteobacteria, Actinobacteria, Firmicutes*, and *Bacteroidetes* were the most abundant phyla as shown by their highest number of hits. Microbes belonging to *Cyanobacteria*, *Candidatus Uhrabacteria, and Spirochaetes* exhibit relatively lower hit counts, implying lower representation in biodigesters.

© University of Venda

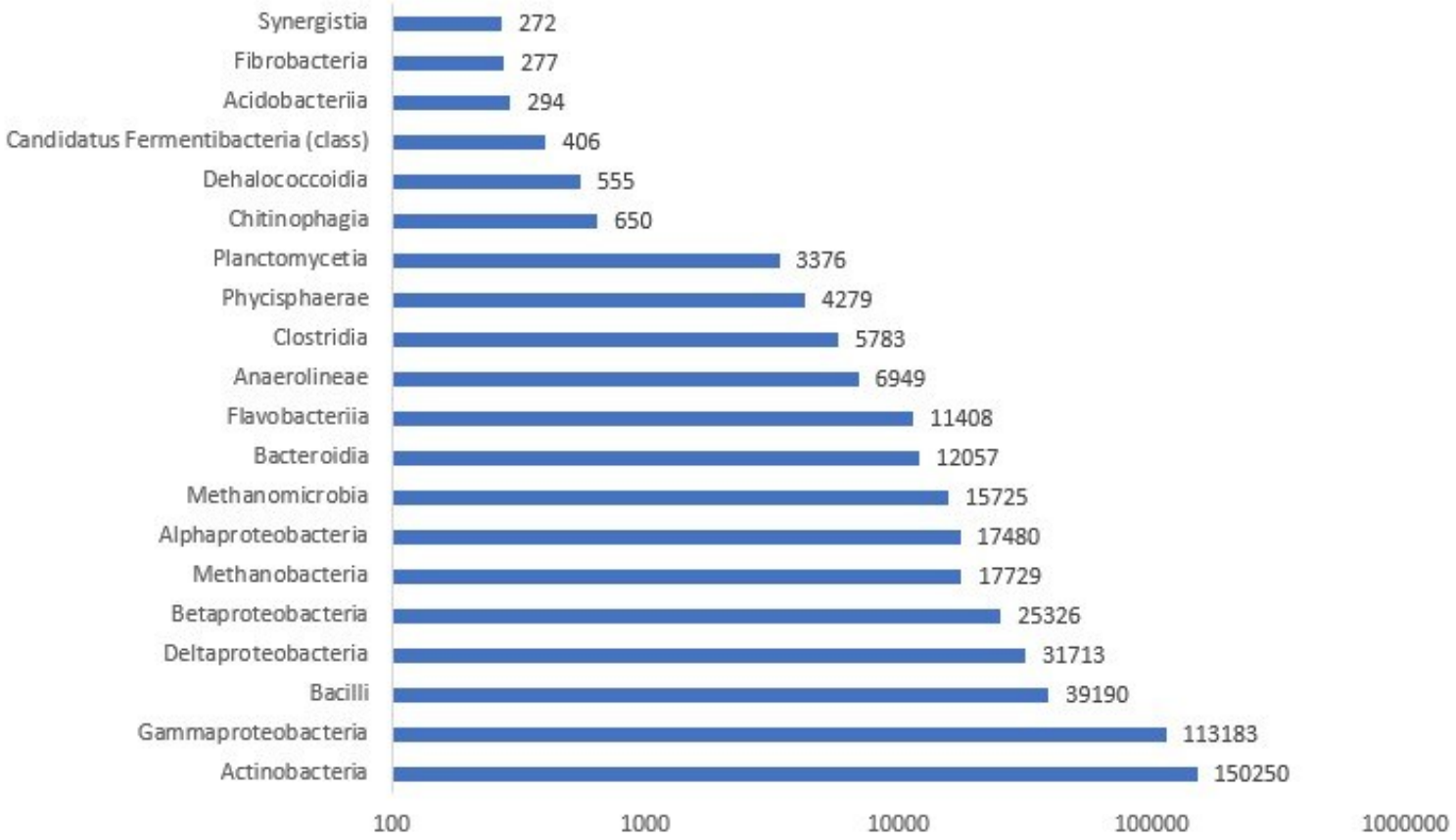


Figure 13: Microbial communities at the class level were identified using CAT, showing a decrease in abundance from the bottom to the top of this clustered bar graph. *Actinobacteria, Gammaproteobacteria, Bacilli*, and *Deltaproteobacteria* were the most abundant classes, while *Synergistia* was the least abundant. Additionally, two archaeal classes, *Methanobacteria* and *Methanomicrobia,* were detected.

© University of Venda

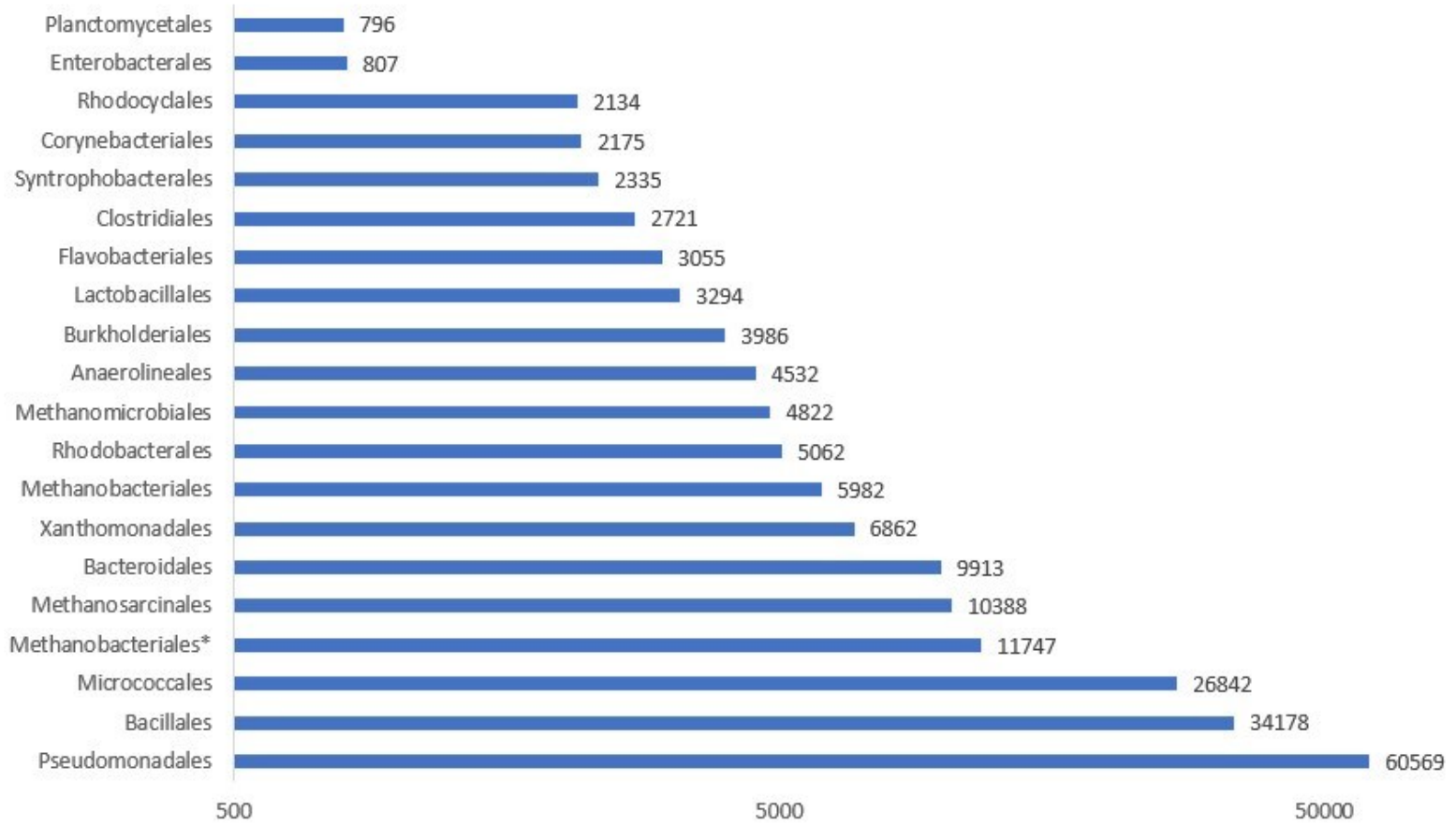


Figure 14: The top 20 microbial communities by relative abundance at the order level in small-scale biodigesters were identified, with 17 being bacterial and 3 belonging to archaeal orders. Among the bacterial orders, *Pseudomonadaceae* were the most abundant, while *Methanobacteriaceae* were the dominant archaeal order.

© University of Venda

4.8. Reconstruction of the metagenome-assembled genome (MAGs)

The genomes of individual species reconstructed from mixed cultures are known as bins and can be successfully reconstructed by binning the contigs obtained from the metagenomic assembly. The resulting contigs were assigned to relevant taxonomic groups based on the shared similarities of composition (GC content) and coverage depth using MetaBAT2. The quality of the bins was assessed using CheckM v1.0.18. Out of 333,575 total contigs generated with length > 500 bp for the samples, binning extracted 38 bins. A total of 8 Metagenome-Assembled Genomes (MAGs) of high quality, showing over 90% completeness with contamination levels at or below 5%, were extracted. Additionally, 10 MAGs were classified as medium quality, possessing at least 50% completeness and contamination levels no higher than 10%, following the Minimum Information about a Metagenome Assembled Genome (MIMAG) standards set by the Genomic Standards Consortium. The remaining were poor quality, < 50% completeness, and >10% contamination. Taxonomic assignment showed that none of the MAGs could be assigned to the species level. Twenty-eight (14 per each) MAGs were assigned to kingdom and phylum level, followed by 3 at order, 1 class, 2 Family, and 4 roots. Bacteria and Archaea were the only kingdoms that recovered, with bacteria being the most abundant. Two bacterial phyla (*Bacteroidetes* and *Actinobacteria*) were identified (**Fig.15**). Euryarchaeota was the only archaeal phyla-generated MAG of high quality in all samples.

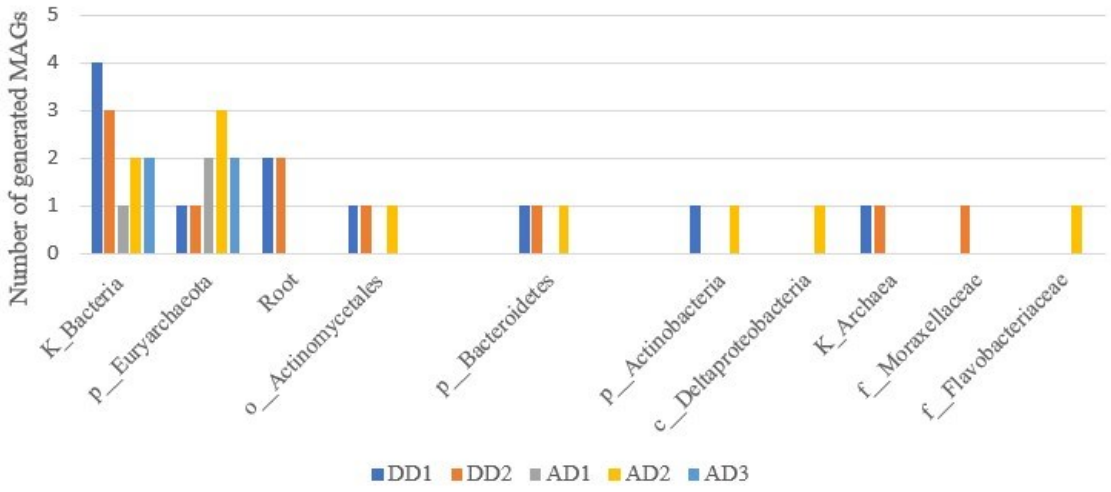


Figure 15: Number of identified taxa in each bin generated using MetaBAT2

4.9. Taxonomic classification of MAGs utilizing GTDB-tk

Taxonomic classification of dereplicated 10 medium-high quality MAGs was done using Genome Taxonomy Database Toolkit (GTDB-tk). As observed with read and contig-based tools, most MAGs were bacteria, and only a number belonged to archaea, with methanogens as the only identified species in the kingdom. Most of the unveiled species belonged to *Proteobacteria* (15), *Actinobacteria* (4), *Bacteroidetes* (29), *Firmicutes* (approximately 60 species in each MAG), and *Euryarchaeota* (36). Most identified microbes within the firmicutes phyla belonged to *the Clostridia* and *Bacilli* class, except for *Acidaminococcus fermentans*, which belonged to *Negativicutes*. Most of the species belonging to *Firmicutes* showed an evolutionary relationship to species belonging to *Actinobacteria* phyla, which can be traced back to their Last Common Ancestor. Taxonomic classification utilizing MAG-based tool has unveiled archaeal species belonging to the following 11 genera: *Methanobrevibacter, Methanoculleus, Methanosarcin, Methanobacterium, Methanococcus, Methanothermobacter, Methanosaeta, Methanosphaera, Methanospirillum, Methanocorpusculum, and Methanocaldococcus* (**Fig. 16**)

© University of Venda

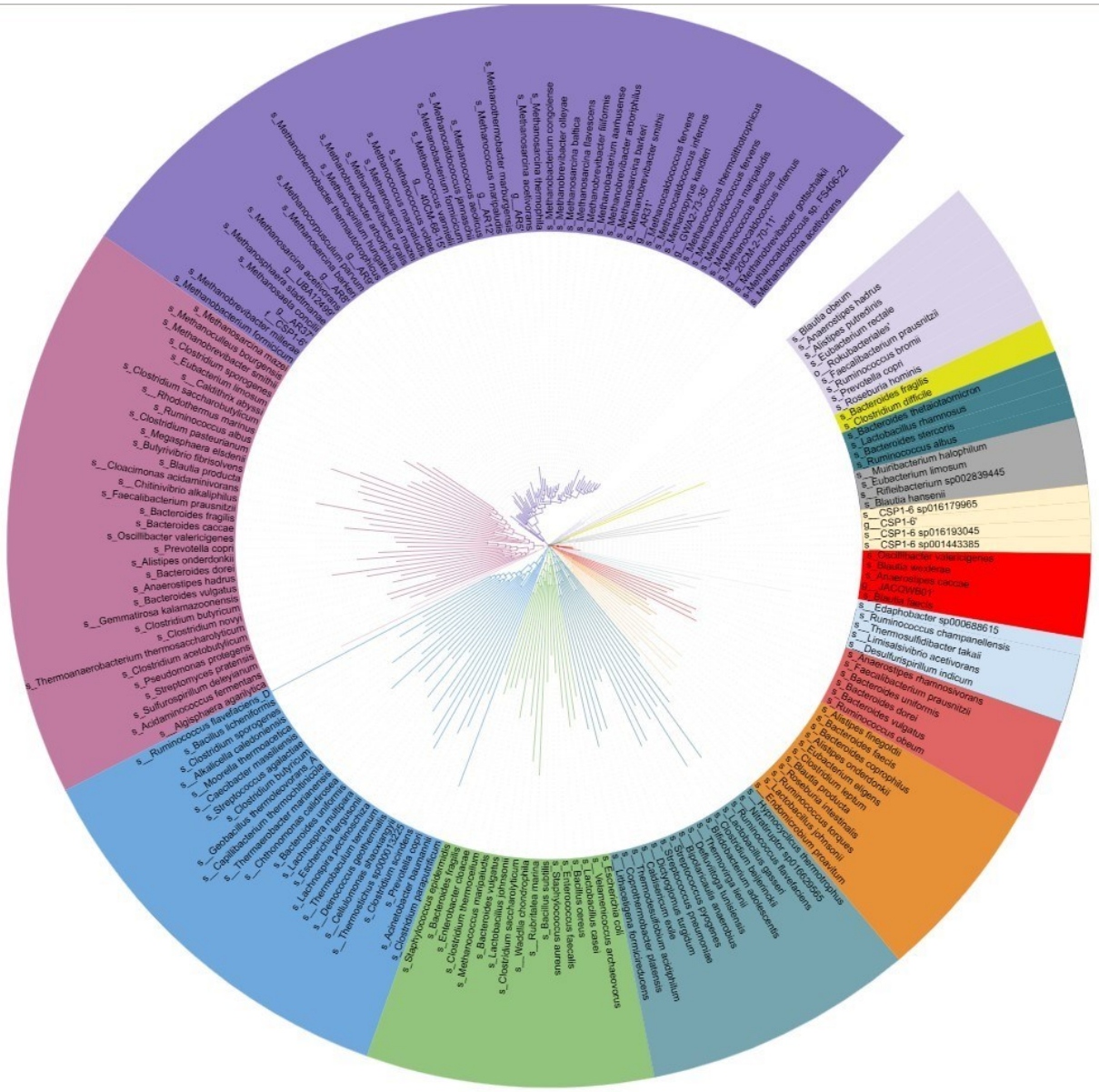


Figure 16: Phylogenetic tree showing the distribution of bacterial and archaeal species abundant in biodigesters. Taxonomic classification was achieved by GTDB-Tk on medium-high MAGs. The tree was annotated using GenBank and RefSeq databases. This visualization emphasizes microbial diversity and evolutionary paths across biodigester. Species belonging to methanogens are shown in dark purple. Most identified species belong to Firmicutes (within *Clostridia* and *Bacilli* class).

4.10. Functional characterization of microbial population in small-scale biodigesters

To explore the functional potential of microbial communities in small-scale biodigesters, a detailed analysis of generated contigs and medium-high (>50 % completeness and < 10 % contamination) quality bins was performed against Rapid Annotation using Subsystem Technology Tool Kit (RAST-tk) in RAST server. Focusing on *Euryarchaeota* phyla, there were genes responsible for the respiration and metabolism of organic compounds such as phosphate, carbohydrates, DNA, RNA, protein, nitrogen, potassium, and secondary metabolism. Genes responsible for respiration, virulence, and defence were also identified (**Fig. 17**). Only genes responsible for the metabolism of organic compounds that aim at energy production were of interest.

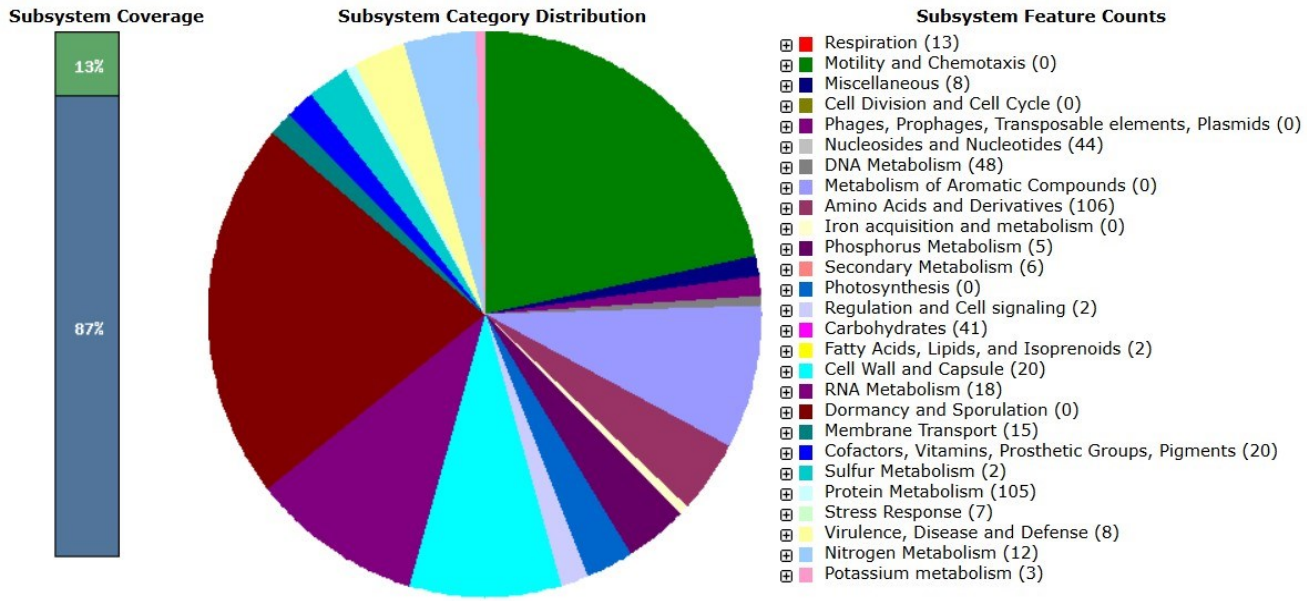


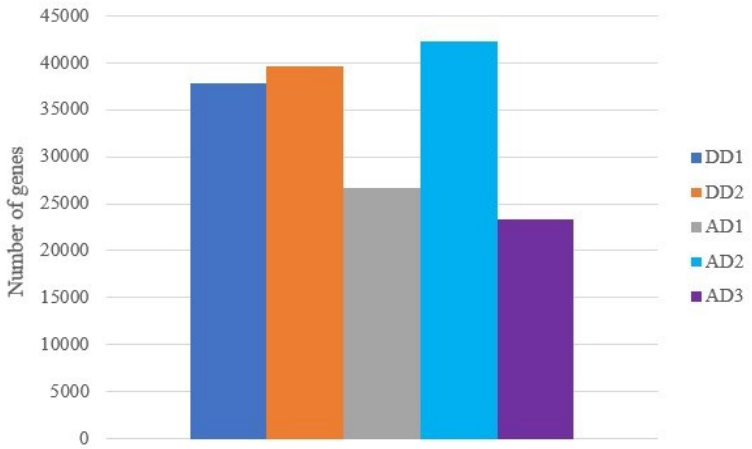
Figure 17: Functional capability of *Euryarchaeota* phyla. Genes involved in the metabolism of protein (105), amino acids and derivates (106), and carbohydrate metabolism (41) were identified

4.11. Overall functional potential across the samples for bacteria and archaea

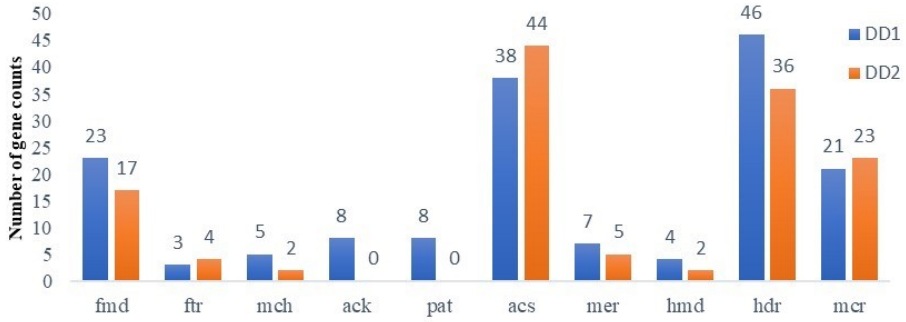
Functional annotation of contigs revealed a total gene count ranging from 18170 to 36082 in all the samples. Samples collected at Dzwerani had a high number of genes, 35326 for the DD1 sample and 36082 for DD2. The sample with less gene count was AD2 with 15451 (**Fig. 18 A**). Enzymes coding for genes responsible for carbohydrates, amino acids, fatty acids, lipids, and isoprenoid metabolism were identified. This includes Acetate kinase, Phosphate acetyltransferase, Alcohol dehydrogenase, Aldehyde dehydrogenase, Butyrate kinase, Indolepyruvate oxidoreductase subunit IorA, Phosphate acetyltransferase known to play a role in the three first phases of anaerobic digestion for energy production. Enzyme coding for genes that play a role in methanogenesis was identified as formylmethanofuran dehydrogenase (fmd), N(5), N(10)-methenyltetrahydromethanopterin cyclohydrolase (much), F420-dependent methylenetetrahydromethanopterin or reductase (mer), Methyl CoM reductase (MCR) CoBCoM heterodisulfide reductase (hdr) (**Fig.18 B**, **Fig.18 C**). The latter has been dominated in sample AD2 (59 counts) and DD1 (46 counts). The Mcr gene is responsible for the final release of methane.

© University of Venda

A.



B.



C.

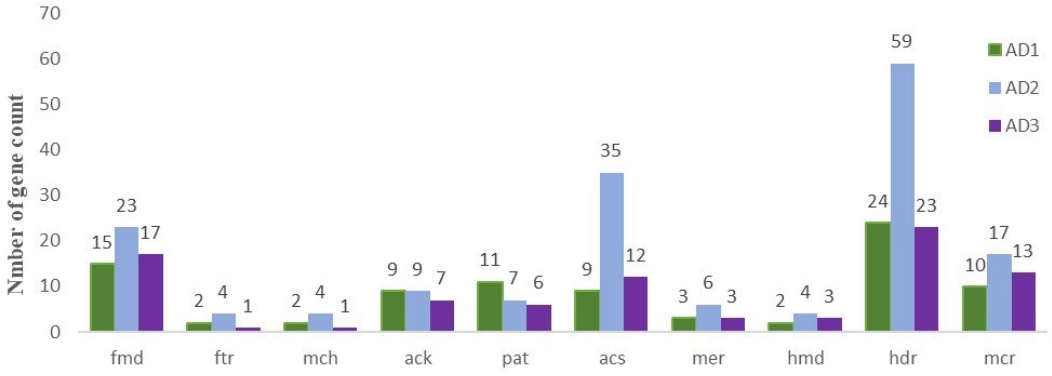


Figure 18: Total gene count identified using RAST-Tk and Relative abundance of hits annotated to the different methanogenesis-related genes (A) Agriculture samples **(B)** Acetate kinase (ack), Methyl CoM reductase (mcr), phosphate acetyltransferase (pta), acetyl-CoA synthetase acs, formylmethanofuran dehydrogenase (fmd), formylmethanofuran-H4MPT formyltransferase (ftr), N(5), N(10)-methenyltetrahydromethanopterin cyclohydrolase (mch), F420-dependent methylenetetrahydromethanopterin or reductase (mer), Dehydrogenase (hmd), CoB-CoM heterodisulfide reductase (hdr).

© University of Venda

CHAPTER 5: DISCUSSION AND CONCLUSION

Advancements in genomics, coupled with the increasing availability of bioinformatic tools, have facilitated the discovery of novel microbial communities, spanning from the superkingdom to species level, that were unculturable outside their natural environments. Previous genomic studies on microbial communities within biodigesters have employed 16S rRNA gene-based and whole genome shotgun metagenomics approaches to explore microbial diversity and function; however, many of these investigations have relied exclusively on contigs, reads or MAG-based approaches. Such singular methodologies may result in incomplete identification of certain microorganisms since some databases used may need more information on specific microbes. This study employed an integrative approach, combining both reads and genome-based methods to comprehensively classify microbial communities and their functional potential in small-scale biodigesters.

An in-depth analysis of whole genome shotgun (WGS) metagenomic data revealed the presence of 115 microbial phyla, 107 classes, 189 orders, 322 families, 738 genera, and 2046 species in Univen agriculture digester. The two digesters investigated in this study differed by only one or two groups of microorganisms at each taxonomic level even though cow dung was used as a substrate for biodigesters. These differences are likely due to subtle variations in the chemical and physical composition of the substrate, including fiber and nutrient content, as well as differences in digester conditions such as temperature, pH, and mixing. Due to the low presence of most common microbes, only a subset could be linked to crucial roles in the four phases of anaerobic digestion. Therefore, only the core taxa identified based on relative abundance, were included in further analysis. Specifically, the top 20 taxa (comprising 18 bacterial and 2 archaeal classes) that consistently exhibited high abundance across all samples, as determined by read- and contig-based classifiers, were considered for detailed analysis.

The taxonomic analysis using both methods revealed that bacteria were the most dominant superkingdom, followed by archaea. This result aligns with previous studies on microbial communities in biodigesters (Bedoya *et al*., 2019; Gaio *et al*., 2023; Vendruscolo *et al*., 2020) and was expected, given that the first three phases of anaerobic digestion (hydrolysis, acidogenesis, and acetogenesis) are primarily mediated by bacteria, while archaea carry out methanogenesis (Ghosh *et al*., 2020; Kumar et al., 2021; Harirchi *et al*., 2022; Subramanian and Suresh, 2024).

© University of Venda

The dominant bacterial and archaeal communities identified in cow dung-fed biodigesters in this study, when compared to similar research focusing on microbial communities in biodigesters, revealed that *Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, and to a lesser extent, *Chloroflexi* were the most abundant bacterial phyla. Among archaea, *Euryarchaeota* and *Candidatus Bathyarchaeota* were the most prevalent phyla. The bacterial findings differ slightly from those reported by García-Lozano *et al*., who observed *Firmicutes*, *Proteobacteria*, *Latescibacteria*, and *Thermotogae* as the dominant phyla using 16S rRNA sequencing (García-Lozano *et al*., 2019). This study, *Latescibacteria* and *Thermotogae* were present at low abundances (<1%). Despite both studies examining microbial communities in digestate derived from biodigesters operating at ambient temperatures, the observed differences in phyla composition may be attributed to variations in temperature during sampling, as *Thermotogae* are known to thrive under higher temperature conditions. A study conducted by Senés-Guerrero *et al*. also reported a slightly different bacterial community composition compared to the present study, with *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* identified as the most abundant bacterial phyla (Senés-Guerrero *et al*., 2019).

Additionally, contig-based analysis (CAT) identified viruses, predominantly those from the *Uroviricota* phylum, as the third most abundant microbial group, with *Eukaryota*, mainly fungi, being the least represented. Although detected in less number, the identified virus phyla in this study are bacteriophages (viruses that infect bacteria). They are known to control the population of *Proteobacteria*, specifically those belonging to *the Gammaproteobacteria* class. In the present study, we observed *Proteobacteria*, particularly the class *Gammaproteobacteria*, as the second abundant bacterial class when utilizing CAT; this could tell us that the presence of viruses in the biodigesters has less impact on the bacterial communities, possibly the amount of the presence of fungi in biodigesters suggested to be associated with their potential role in the hydrolysis phase, where they assist cellulolytic bacteria by leveraging their high protease activity, which enables them to adhere to the substrate (cow dung) and penetrate cell walls (Kazda *et al*., 2014).

Phylum *Proteobacteria*, particularly the class *Gammaproteobacteria*, was the most abundant in our study and is known for its capability of hydrolyzing and fermenting organic substrates in both the initial and second phases of anaerobic digestion, respectively (Arelli *et al*., 2023; Mang *et al*., 2022). In the descending order of abundance, *Xanthomonadaceae*, *Acinetobacter*, *Pseudomonas*, and *Methylococcaceae* were the identified genera within the class. It is important to note that most of the species, such as *Pseudomonas flexibilis*, *Pseudomonas*

© University of Venda

stutzeri, and *Pseudomonas linyingensis* within the genus of *Pseudomonas* are known as facultative anaerobic for their ability to survive in both aerobic and anaerobic conditions. Their presence in biodigesters where there are temperature and pH fluctuations and competition between anaerobic microbes for substrate degradation could be linked to the presence of survival enzymes such as Bacterioferritin-associated ferredoxin (Bfd), ferredoxin oxidoreductase that was identified in this study. These genes aid in various redox reactions that allow microbes to contribute to the efficient degradation of substrates for energy production in anaerobic biodigesters. The Bfd enzyme was identified as part of the defense and virulence subsystem, indicating its role in supporting survival by aiding in biofilm formation through the encoded gene. Both methods revealed bacterial phylum: *Proteobacteria* (41-64%), *Firmicutes* (1-19%), and *Bacteroidetes* (8-22%) in descending order. *Euryarchaeota* is the most abundant phylum in the archaeal kingdom; this aligns with other studies *(*El Houari *et al*., 2020; Bonugli-Santos *et al*., 2022).

In cow dung-fed biodigesters, methanogenesis represents the final stage of anaerobic digestion, where methanogenic archaea convert carbon intermediates generated from the other phases into methane, serving as a renewable energy source (Karakashev *et al*., 2005). However, the overall anaerobic digestion process involves multiple phases, such as hydrolysis, acidogenesis, and acetogenesis. These phases are crucial for breaking down complex organic substrates into simpler compounds (Li *et al*., 2021). The functional potential of genes involved in these initial stages, such as *Acetate kinase* (ack), *Phosphate acetyltransferase* (pta), *Alcohol dehydrogenase* (adh), *Aldehyde dehydrogenase* (aldh), *Butyrate kinase* (buk), and *Indolepyruvate oxidoreductase subunit I or A* (IorA), complements the methanogenesis pathway by preparing and providing substrates for methanogenic archaea (Liu *et al*., 2018). All these genes were identified in this study.

Alcohol dehydrogenase and *Aldehyde dehydrogenase* are involved in the metabolism of alcohols and aldehydes, compounds typically produced during the acidogenesis phase of anaerobic digestion (Kato *et al*., 2014). These enzymes facilitate the conversion of alcohols into aldehydes and further into acids, contributing to the pool of volatile fatty acids (VFAs), such as acetate and propionate, which serve as intermediates for methanogenesis. The presence of these genes underscores the importance of acidogenic bacteria in cow dung-fed biodigesters and their role in transforming complex organic matter into substrates for methane production (Mishra *et al*., 2018). Butyrate kinase catalyzes the formation of butyrate, another significant VFA in anaerobic digestion. Butyrate, like acetate, is a critical intermediate in methanogenesis,

© University of Venda

particularly in syntrophic relationships where fermentative bacteria produce butyrate, and methanogens consume it (Schink & Stams, 2013). The detection of *buk* genes suggests active butyrate formation and highlights the interplay between different microbial populations in converting fatty acids into methane precursors (Rotaru *et al*., 2014).

Similarly, Acetate kinase and Phosphate acetyltransferase are critical enzymes in converting acetyl-CoA to acetate, one of the primary substrates for methanogenesis. Their presence in the microbial community reflects the capacity for the third phase, acetogenesis, where volatile fatty acids (VFAs) like acetate are produced from the breakdown of organic matter (Siegert *et al*., 2014). In this context, the abundance of *ack* and *pta* genes indicates an active acetogenesis pathway that supplies acetate as a substrate for methane production via acetoclastic methanogens (Tao *et al*., 2019).

Indolepyruvate oxidoreductase subunit *IorA* plays a role in the oxidative decarboxylation of keto acids to acyl-CoA, another important pathway in acetogenesis (Wu *et al*., 2020). Its detection in cow dung-fed biodigesters points to the involvement of more complex metabolic processes that break down aromatic amino acids and other organic compounds, ultimately feeding into the methanogenesis pathway (Yan *et al*., 2020).

The genes involved in hydrolysis, acidogenesis, and acetogenesis interact with methanogenesis-related genes such as *fmd*, *mch*, *mer*, *mcr*, and *hdr*. The earlier stages of digestion, represented by enzymes like Acetate kinase, Phosphate acetyltransferase, Alcohol dehydrogenase, and Butyrate kinase, produce vital intermediates such as acetate, butyrate, and alcohols. These intermediates are then converted into methane by methanogenic archaea, completing the anaerobic digestion (Conrad, 2020). Therefore, understanding the functional potential of these genes in all stages of digestion is critical for optimizing methane production and improving overall biodigester performance.

This study also identified genes responsible for the metabolism of carbohydrates (Pyruvate: ferredoxin oxidoreductase, PFOR) and amino acids in association with *Euryarchaeota* highquality MAG that were subjected to RAST server. This could suggest that some archaea also play a role in hydrolysis and acidogenesis. A study conducted by Maus *et al*. 2018, which constructed the MAGs belonging to Phylum *Bathyarchaeota*, predicted the microbes belonging to these phyla to be involved in hydrolysis and acidogenesis based on the identification of genes encoding carbohydrate-active enzymes, which implies that these microbes have the potential to utilize different carbohydrates and amino acids metabolism. In our study, the Contig-based

© University of Venda

tool (CAT) revealed phylum *Candidatus Bathyarchaeota* to be the second most abundant archaeal phyla.

Looking at the abundance of bacteria and archaea identified in this study compared to other superkingdoms, we can conclude that they are the leading role players in the cow dung-fed biodigesters, and some microbes involved in hydrolysis also play a role in acidogenesis. Microbes involved in acetogenesis may be specifically focused on that phase. The three phases produce intermediates used by methanogens in the final stage of anaerobic digestion to produce methane.

The primary limitation of this study is the sample size, which was constrained by budget limitations, analyzing only five samples may have resulted in an underrepresentation of the microbial diversity in small-scale biodigesters, given the complexity of their structure.

We recommend that future studies increase the sample size, as some microbes not identified in this study may be discovered; subsequently, cultivating and monitoring these identified microbes in a laboratory setting would help determine optimal conditions for maximizing gas production, ultimately enhancing environmental bioaugmentation.

© University of Venda

CHAPTER 6: REFERENCES

Achudhan, A.B., Kannan, P., Gupta, A. and Saleena, L.M., 2023. A Review of web-based metagenomics platforms for analysing next-generation sequence data. *Biochemical Genetics*, pp.1-12.

Akter, S., Rahman, M.S., Ali, H., Minch, B., Mehzabin, K., Siddique, M.M., Galib, S.M., Yesmin, F., Azmuda, N., Adnan, N. and Hasan, N.A., 2023. Exploring Microbial diversity and functional potential along the Bay of Bengal coastline in Bangladesh: insights from amplicon sequencing and shotgun metagenomics. *bioRxiv*, pp.2023-04.

Almeida, O.G. and De Martinis, E.C., 2019. Bioinformatics tools to assess metagenomic data for applied microbiology. *Applied microbiology and biotechnology*, *103*(1), pp.69-82.

Arelli, V., Mamindlapelli, N.K. and Anupoju, G.R., 2023. Influence of solids concentration on microbial diversity and methane yield in the anaerobic digestion of rice husk. *Bioresource Technology Reports*, *22*, p.101455.

Ballesteros, N., Páez, L., Luna, N., Reina, A., Urrea, V., Sánchez, C., Ramírez, A., Ramirez, J.D. and Muñoz, M., 2023. Characterization of microbial communities in seven wetlands with different anthropogenic burden using Next Generation Sequencing in Bogotá, Colombia. *Scientific Reports*, *13*(1), p.16973.

Bedoya, K., Coltell, O., Cabarcas, F. and Alzate, J.F., 2019. Metagenomic assessment of the microbial community and methanogenic pathways in biosolids from a municipal wastewater treatment plant in Medellín, Colombia. *Science of the total environment*, *648*, pp.572-581.

Belone, M.C.L., Brosens, D., Kokko, M. and Sarlin, E., 2024. Effects of mesophilic and thermophilic anaerobic digestion of sewage sludge on different polymers: Perspectives on the potential of the treatment to degrade microplastics. *Science of the Total Environment*, *907*, pp.168014.

Bonugli-Santos, R.C., Marteres, T.J., Luiz, F.N., Somer, J.G., Mari, Â.G. and Passarini, M.R.Z., 2022. Prolonged acetogenic phase and biological succession during anaerobic digestion using swine manure. *Folia Microbiologica*, *67*(5), pp.733-745.

Brumfield, K.D., Huq, A., Colwell, R.R., Olds, J.L. and Leddy, M.B., 2020. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One*, *15*(2), p.e0228899.

© University of Venda

Cabezas, A., de Araujo, J.C., Callejas, C., Galès, A., Hamelin, J., Marone, A., Sousa, D.Z., Trably, E. and Etchebehere, C., 2015. How to use molecular biology tools for the study of the anaerobic digestion process?. *Reviews in Environmental Science and Bio/Technology*, *14*, pp.555-593.

Cambuy, D.D., Coutinho, F.H. and Dutilh, B.E., 2016. Contig annotation tool CAT robustly classifies assembled metagenomic contigs and long sequences. *BioRxiv*, p.072868.

Campanaro, S., Treu, L., Kougias, P.G., De Francisci, D., Valle, G. and 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*, pp.1-17.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. and Parks, D.H., 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.

Chinwendu, A.O., Catherine, A.O.B. and Okon, E.U., 2020. Metagenomics in-sight into the microbial structural diversity of anaerobic digester utilising fruit waste as substrate. *Journal of microbiology, biotechnology and food sciences*, *9*(5), pp.965-969.

Conrad, R. (2020). Importance of hydrogenotrophic, aceticlastic, and methylotrophic methanogenesis for methane production in acidic, methanogenic environments. *Frontiers in Microbiology*, 11, 471.

Dabban, I.A., Ahmad, M., Enejiyon, S.O., Hauwau, A.N., Gani, M., Oyewole, O.A. and Adetunji, C.O., 2024. Isolation Techniques Used for Molecular Characterization of Beneficial Microorganisms: Cultural, Biochemical and Molecular Characterization. *Handbook of Agricultural Biotechnology*, pp.491-545.

Deng, C., Zhao, R., Qiu, Z., Li, B., Zhang, T., Guo, F., Mu, R., Wu, Y., Qiao, X., Zhang, L. and Cheng, J.J., 2022. Genome-centric metagenomics provides new insights into the microbial community and metabolic potential of landfill leachate microbiota. *Science of The Total Environment*, *816*, p.151635.

Deppenmeier, U. (2002). The unique biochemistry of methanogenesis. *Progress in Nucleic Acid Research and Molecular Biology*, 71, pp. 223–283.

Diagi, E.A., Akinyemi, M.L., Emetere, M.E., Ogunrinola, I.E., and Ndubuisi, A.O., 2019, September. Comparative analysis of biogas produced from cow dung and poultry droppings.

© University of Venda

In *IOP Conference Series: Earth and Environmental Science* (Vol. 331, No. 1, p. 012064). IOP Publishing.

Duan, X., Chen, Y., Feng, L. and Zhou, Q., 2021. Metagenomic analysis reveals nonylphenolshaped acidification and methanogenesis during sludge anaerobic digestion. *Water Research*, *196*, p.117004.

El Houari, A., Ranchou-Peyruse, M., Ranchou-Peyruse, A., Bennisse, R., Bouterfas, R., Goni Urriza, M.S., Qatibi, A.I. and Guyoneaud, R., 2020. Microbial Communities and SulfateReducing Microorganisms Abundance and Diversity in Municipal Anaerobic Sewage Sludge Digesters from a Wastewater Treatment Plant (Marrakech, Morocco). *Processes*, *8*(10), p.1284.

Escobar-Zepeda, A., Godoy-Lozano, E.E., Raggi, L., Segovia, L., Merino, E., Gutiérrez-Rios, R.M., Juarez, K., Licea-Navarro, A.F., Pardo-Lopez, L. and Sanchez-Flores, A., 2018. Analysis of sequencing strategies and tools for taxonomic annotation: defining standards for progressive metagenomics. *Scientific reports*, *8*(1), p.12034.

Ferry, J.G. (2011). Fundamentals of methanogenic pathways that are key to the biomethanation of complex biomass. *Current Opinion in Biotechnology*, 22(3), pp. 351–357.

Flaviani, F., Schroeder, D.C., Lebret, K., Balestreri, C., Highfield, A.C., Schroeder, J.L., Thorpe, S.E., Moore, K., Pasckiewicz, K., Pfaff, M.C. and Rybicki, E.P., 2018. Distinct oceanic microbiomes from viruses to protists located near the antarctic circumpolar current. *Frontiers in Microbiology*, *9*, p.1474.

Forry, S.P., Servetas, S.L., Dootz, J.N., Hunter, M.E., Kralj, J.G., Filliben, J.J. and Jackson, S.A., 2025. A sensitivity analysis of methodological variables associated with microbiome measurements. *Microbiology Spectrum*, pp.e00696-24.

Fu, S., Lian, S., Angelidaki, I. and Guo, R., 2023. Micro-aeration: an attractive strategy to facilitate anaerobic digestion. *Trends in Biotechnology*, *41*(5), pp.714-726.

Fukuda, K., Ogawa, M., Taniguchi, H. and Saito, M., 2016. Molecular approaches to studying microbial communities: targeting the 16S ribosomal RNA gene. *Journal of UOEH*, *38*(3), pp.223-232.

© University of Venda

Gaio, J., Lora, N.L., Iltchenco, J., Magrini, F.E. and Paesi, S., 2023. Seasonal characterization of the prokaryotic microbiota of full-scale anaerobic UASB reactors treating domestic sewage in southern Brazil. *Bioprocess and Biosystems Engineering*, *46*(1), pp.69-87.

Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., Gao, N., Shen, W. and Schnabl, B., 2021. An introduction to next generation sequencing bioinformatic analysis in gut microbiome studies. *Biomolecules*, *11*(4), p.530.

García-Lozano, M., Hernández-De Lira, I.O., Huber, D.H. and Balagurusamy, N., 2019. Spatial variations of bacterial communities of an anaerobic lagoon-type biodigester fed with dairy manure. *Processes*, *7*(7), p.408.

Ghosh, P., Kumar, M., Kapoor, R., Kumar, S.S., Singh, L., Vijay, V., Vijay, V.K., Kumar, V. and Thakur, I.S., 2020. Enhanced biogas production from municipal solid waste via co-digestion with sewage sludge and metabolic pathway analysis. Bioresource Technology, 296, p.122275.

Harirchi, S., Wainaina, S., Sar, T., Nojoumi, S.A., Parchami, M., Parchami, M., Varjani, S., Khanal, S.K., Wong, J., Awasthi, M.K. and Taherzadeh, M.J., 2022. Microbiological insights into anaerobic digestion for biogas, hydrogen or volatile fatty acids (VFAs): a review. *Bioengineered*, *13*(3), pp.6521-6557.

Ijoma, G.N., Nkuna, R., Mutungwazi, A., Rashama, C. and Matambo, T.S., 2021. Applying PICRUSt and 16S rRNA functional characterisation to predicting co-digestion strategies of various animal manures for biogas production. *Scientific Reports*, *11*(1), p.19913.

Jain, K., Pandita, P., Mathuria, A., Mehak, Das, D., Saini, A. and Mani, I., 2024. Emerging Tools for Generating Genomics Data. In *Advances in Genomics: Methods and Applications* (pp. 1-39). Singapore: Springer Nature Singapore.

Janesch, E., Pereira, J., Neubauer, P. and Junne, S., 2021. Phase separation in anaerobic digestion: a potential for easier process combination?. *Frontiers in Chemical Engineering*, *3*, p.711971.

Jünemann, S., Kleinbölting, N., Jaenicke, S., Henke, C., Hassa, J., Nelkner, J., Stolze, Y., Albaum, S.P., Schlüter, A., Goesmann, A. and Sczyrba, A., 2017. Bioinformatics for NGSbased metagenomics and the application to biogas research. *Journal of biotechnology*, *261*, pp.10-23.

© University of Venda

Kainthola, J., Kalamdhad, A.S. and Goud, V.V., 2019. A review on enhanced biogas production from anaerobic digestion of lignocellulosic biomass by different enhancement techniques. *Process Biochemistry*, *84*, pp.81-90.

Kang, D.D., Froula, J., Egan, R. and Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, p.e1165.

Karakashev, D., Batstone, D. J., & Angelidaki, I. (2005). Influence of environmental factors on methanogenesis in anaerobic digesters. *Critical Reviews in Biotechnology*, 25(2), pp.173-199.

Kato, S., Kosaka, T., & Watanabe, K. (2014). Methanogenesis facilitated by electric syntrophy via (semi)conductive iron-oxide minerals. *Environmental Microbiology*, 16(12), pp. 3686-3697

Kazda, M., Langer, S. and Bengelsdorf, F.R., 2014. Fungi open new possibilities for anaerobic fermentation of organic residues. *Energy, Sustainability and Society*, *4*, pp.1-9.

Kim, N.K., Lee, S.H., Kim, Y. and Park, H.D., 2022. Current understanding and perspectives in anaerobic digestion based on genome-resolved metagenomic approaches. *Bioresource technology*, *344*, p.126350.

Kimisto, A.K., Muia, A.W., Ong'ondo, G.O. and Ndung'u, K.C., 2023. Molecular characterization of microorganisms with industrial potential for methane production in sludge from Kangemi sewage treatment plant, Nyeri county–Kenya. *Heliyon*, *9*(5).

Kumar, V., Nabaterega, R., Khoei, S. and Eskicioglu, C., 2021. Insight into interactions between syntrophic bacteria and archaea in anaerobic digestion amended with conductive materials. *Renewable and Sustainable Energy Reviews*, *144*, p.110965.

Laiq Ur Rehman, M., Iqbal, A., Chang, C.C., Li, W. and Ju, M., 2019. Anaerobic digestion. *Water Environment Research*, *91*(10), pp.1253-1271.

Lee, J.Y., Sim, Y.B., Jung, J.H., Pandey, A.K., Kyung, D. and Kim, S.H., 2024. Greenhouse gas emissions and net energy production of dark fermentation from food waste followed by anaerobic digestion. *Energy*, *312*, p.133559.

Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., Yamashita, H. and Lam, T.W., 2016. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, *102*, pp.3-11.

© University of Venda

Li, Y., Jing, Z., Pan, J., Luo, G., Feng, L., Jiang, H., Zhou, H., Xu, Q., Lu, Y. and Liu, H., 2022. Multi-omics joint analysis of the effect of temperature on microbial communities, metabolism, and genetics in full-scale biogas reactors with food waste. *Renewable and Sustainable Energy Reviews*, *160*, p.112261.

Li, Y., Zhang, R., He, Y., Zhang, C., Liu, X., & Chen, C. (2021). Anaerobic digestion of organic wastes: A review of microbial community dynamics and management strategies. *Frontiers in Microbiology*, 12, p.634837.

Lindner, B.G., Gerhardt, K., Feistel, D.J., Rodriguez-R, L.M., Hatt, J.K. and Konstantinidis, K.T., 2024. A user's guide to the bioinformatic analysis of shotgun metagenomic sequence data for bacterial pathogen detection. *International journal of food microbiology*, *410*, p.110488.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M., 2012. Comparison of next-generation sequencing systems. *BioMed research international*, *2012*(1), p.251364.

Liu, S., Rodriguez, J.S., Munteanu, V., Ronkowski, C., Sharma, N.K., Alser, M., Andreace, F., Blekhman, R., Błaszczyk, D., Chikhi, R. and Crandall, K.A., 2025. Analysis of metagenomic data. *Nature Reviews Methods Primers*, *5*(1), p.5.

Liu, T., Zhang, J., Wang, Y., Fang, S., & Yuan, X. (2018). Functional potential and diversity of methanogenic archaea in full-scale anaerobic digesters treating swine wastewater. *Frontiers in Microbiology*, 9, pp. 1236.

Mahmudul, H.M., Akbar, D., Rasul, M.G., Narayanan, R. and Mofijur, M., 2022. Estimation of the sustainable production of gaseous biofuels, generation of electricity, and reduction of greenhouse gas emissions using food waste in anaerobic digesters. *Fuel*, *310*, p.122346.

Mallawaarachchi, V., Wickramarachchi, A., Xue, H., Papudeshi, B., Grigson, S.R., Bouras, G., Prahl, R.E., Kaphle, A., Verich, A., Talamantes-Becerra, B. and Dinsdale, E.A., 2024. Solving genomic puzzles: computational methods for metagenomic binning. *Briefings in Bioinformatics*, *25*(5), p.bbae372.

Mang, S.M., Trotta, V., Scopa, A. and Camele, I., 2022. Metagenomic analysis of bacterial community structure and dynamics of a digestate and a more stabilized digestate-derived compost from agricultural waste. *Processes*, *10*(2), p.379.

© University of Venda

Mao, C., Feng, Y., Wang, X. and Ren, G., 2015. Review on research achievements of biogas from anaerobic digestion. *Renewable and sustainable energy reviews*, *45*, pp.540-555.

Mathieu, A., Leclercq, M., Sanabria, M., Perin, O. and Droit, A., 2022. Machine learning and deep learning applications in metagenomic taxonomy and functional annotation. *Frontiers in Microbiology*, *13*, p.811495.

Maus, I., Rumming, M., Bergmann, I., Heeg, K., Pohl, M., Nettmann, E., Jaenicke, S., Blom, J., Pühler, A., Schlüter, A. and Sczyrba, A., 2018. Characterization of Bathyarchaeota genomes assembled from metagenomes of biofilms residing in mesophilic and thermophilic biogas reactors. *Biotechnology for biofuels*, *11*, pp.1-13.

Mazur-Marzec, H., Andersson, A.F., Błaszczyk, A., Dąbek, P., Górecka, E., Grabski, M., Jankowska, K., Jurczak-Kurek, A., Kaczorowska, A.K., Kaczorowski, T. and Karlson, B., 2024. Biodiversity of microorganisms in the Baltic Sea: the power of novel methods in the identification of marine microbes. *FEMS Microbiology Reviews*, *48*(5), p.fuae024.

McGregor, B.A., Razmjou, E., Hooshyar, H., Seeger, D.R., Golovko, S.A., Golovko, M.Y., Singer, S.M., Hur, J. and Solaymani-Mohammadi, S., 2023. A shotgun metagenomic analysis of the fecal microbiome in humans infected with Giardia duodenalis. *Parasites & Vectors*, *16*(1), p.239.

Meegoda, J.N., Li, B., Patel, K. and Wang, L.B., 2018. A review of the processes, parameters, and optimization of anaerobic digestion. *International journal of environmental research and public health*, *15*(10), p.2224.

Mellado, M. and Vera, J., 2021. Microorganisms that participate in biochemical cycles in wetlands. *Canadian Journal of Microbiology*, *67*(11), pp.771-788.

Menzel, P., Ng, K.L. and Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature communications*, *7*(1), p.11257.

Menzel, T., Neubauer, P. and Junne, S., 2020. Role of microbial hydrolysis in anaerobic digestion. *Energies*, *13*(21), p.5555.

Mishra, S., Singh, R., & Sarma, A. K. (2018). Anaerobic digestion of cow dung and kitchen waste for biogas production: A review. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 40(16), pp. 1916-1926.

© University of Venda

Mukhuba, M., Roopnarain, A., Moeletsi, M.E. and Adeleke, R., 2020. Metagenomic insights into the microbial community and biogas production pattern during anaerobic digestion of cow dung and mixed food waste. *Journal of Chemical Technology & Biotechnology*, *95*(1), pp.151-162.

Muñoz-Ramírez, Z.Y., González-Escobedo, R., Avila-Quezada, G.D., Ramírez-Sánchez, O., Higareda-Alvear, V.M., Zapata-Chávez, E., Borrego-Loya, A. and Muñoz-Castellanos, L.N., 2024. Exploring Microbial Rhizosphere Communities in Asymptomatic and Symptomatic Apple Trees Using Amplicon Sequencing and Shotgun Metagenomics. *Agronomy*, *14*(2), p.357.

Navgire, G.S., Goel, N., Sawhney, G., Sharma, M., Kaushik, P., Mohanta, Y.K., Mohanta, T.K. and Al-Harrasi, A., 2022. Analysis and Interpretation of metagenomics data: An approach. *Biological Procedures Online*, *24*(1), p.18.

Niya, B., Yaakoubi, K., Beraich, F.Z., Arouch, M. and Kadmiri, I.M., 2024. Current status and future developments of assessing microbiome composition and dynamics in anaerobic digestion systems using metagenomic approaches. *Heliyon*.

Noraini, M., Sanusi, S., Elham, O.S.J., Sukor, M.Z. and Halim, K., 2017. Factors affecting production of biogas from organic solid waste via anaerobic digestion process: A review. *Solid state science and technology*, *25*(1), pp.29-39.

Olm, M.R., Brown, C.T., Brooks, B. and Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through dereplication. *The ISME journal*, *11*(12), pp.2864-2868.

Olson, N.D., Treangen, T.J., Hill, C.M., Cepeda-Espinoza, V., Ghurye, J., Koren, S. and Pop, M., 2019. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in bioinformatics*, *20*(4), pp.1140-1150.

Ondov, B.D., Bergman, N.H. and Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. *BMC bioinformatics*, *12*, pp.1-10.

Pasalari, H., Gholami, M., Rezaee, A., Esrafili, A. and Farzadkia, M., 2021. Perspectives on microbial community in anaerobic digestion with emphasis on environmental parameters: a systematic review. *Chemosphere*, *270*, p.128618.

© University of Venda

Pérez-Llano, Y., Batista-García, R.A., del Rayo Sánchez-Carbente, M., Folch-Mallol, J.L., Cuadros-Orellana, S., Nicolás-Morales, A.D. and Castro-Alarcón, N., 2021. Metagenomic Tools for Taxonomic and Functional Annotation. In *Metagenomics and Microbial Ecology* (pp. 21-44). CRC Press.

Pillai, B.B.K., Meghvansi, M.K., Sudha, M.C. and Sreenivasulu, M., 2022. Microbial Community Dynamics in Anaerobic Digester Treating Human Waste: A Review. *Anaerobic Biodigesters for Human Waste Treatment*, pp.95-111.

Pramanik, S.K., Suja, F.B., Zain, S.M. and Pramanik, B.K., 2019. The anaerobic digestion process of biogas production from food waste: Prospects and constraints. *Bioresource Technology Reports*, *8*, p.100310.

Pyzik, A., Ciezkowska, M., Krawczyk, P.S., Sobczak, A., Drewniak, L., Dziembowski, A. and Lipinski, L., 2018. Comparative analysis of deep sequenced methanogenic communities: identification of microorganisms responsible for methane production. *Microbial cell factories*, *17*, pp.1-16.

Qurbani, K., Hussein, S. and Ahmed, S.K., 2024. Molecular Biotechniques for the Isolation and Identification of Microorganisms.

Rahman, M.S., Hoque, M.N., Puspo, J.A., Islam, M.R., Das, N., Siddique, M.A., Hossain, M.A. and Sultana, M., 2021. Microbiome signature and diversity regulates the level of energy production under anaerobic condition. *Scientific reports*, *11*(1), p.19777.

Rotaru, A. E., Shrestha, P. M., Liu, F., Shrestha, M., Shrestha, D., & Lovley, D. R. (2014). Direct interspecies electron transfer between Geobacter and Methanosarcina species in syntrophic cocultures. *Science*, 342(6163), p. 334-337.

Saetta, D., Bullard, T., Smith, A., Yeh, D., Fischer, J., Dixit, A., Spern, C., Khodadad, C. and Roberson, L., 2023, July. Survey of Microbial Community in Bioreactors Used for Bioregenerative Water Purification. 2023 International Conference on Environmental Systems.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G. and Malonia, S.K., 2023. Next-generation sequencing technology: current trends and advancements. *Biology*, *12*(7), p.997.

Schink, B., & Stams, A. J. M. (2013). Syntrophism among prokaryotes. In E. Rosenberg (Ed.), *The Prokaryotes* (pp. 471-493). Springer

© University of Venda

Scholz, M.B., Lo, C.C. and Chain, P.S., 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current opinion in biotechnology*, *23*(1), pp.9-15.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), pp.2068-2069.

Senés-Guerrero, C., Colón-Contreras, F.A., Reynoso-Lobo, J.F., Tinoco-Pérez, B., SillerCepeda, J.H. and Pacheco, A., 2019. Biogas-producing microbial composition of an anaerobic digester and associated bovine residues. *MicrobiologyOpen*, *8*(9), p.e00854.

Shabbirahmed, A.M., Somu, P., Yang, H.H., Hiruthyaswamy, S.P., Karua, C.S. and Yadav, A.K., 2024. Challenges and strategies for waste food anaerobic digestion: insights and future directions. *Environment, Development and Sustainability*, pp.1-34.

Sharma, P., Bano, A., Singh, S.P. and Tong, Y.W., 2024. Microbial Communities Driving Pollution Degradation in Contaminated Environments: Metagenomic Insights. In *Microbial Diversity in the Genomic Era* (pp. 777-789). Academic Press.

Sharma, P., Bano, A., Singh, S.P., Srivastava, S.K., Singh, S.P., Iqbal, H.M. and Varjani, S., 2023. Different stages of microbial community during the anaerobic digestion of food waste. *Journal of Food Science and Technology*, *60*(8), pp.2079-2091.

Siegert, M., Li, X. F., Yates, M. D., & Logan, B. E. (2014). The presence of acetate kinase and phosphate acetyltransferase genes in acetate- and ethanol-consuming Geobacter sulfurreducens strains. *Applied and Environmental Microbiology*, 80(18), p. 5908-5916.

Singh, A. and Singh, A., 2022. Microbial degradation and value addition to food and agriculture waste. *Current Microbiology*, *79*(4), p.119.

St-Pierre, B. and Wright, A.D.G., 2014. Comparative metagenomic analysis of bacterial populations in three full-scale mesophilic anaerobic manure digesters. *Applied microbiology and biotechnology*, *98*, pp.2709-2717.

Subramanian, K. and Suresh, K., 2024. Metabolic and Biochemical Pathways for Anaerobic Digestion. In *Anaerobes and Waste Conversion Technologies* (pp. 3-25). Singapore: Springer Nature Singapore.

Sudarikov, K., Tyakht, A. and Alexeev, D., 2017. Methods for the metagenomic data visualization and analysis. *Current issues in molecular biology*, *24*(1), pp.37-58.

© University of Venda

Tao, Y., Gao, D., Liu, Y., & Zhang, Y. (2019). Functional insights into microbial community and methanogenic pathways in household-scale anaerobic digesters fed with cow manure. *Journal of Cleaner Production*, 228, p. 1187-1194.

Thomas, T., Gilbert, J. and Meyer, F., 2012. Metagenomics-a guide from sampling to data analysis. *Microbial informatics and experimentation*, *2*, pp.1-12.

Uche, A.M., Emmanuel, O.T., Paul, O.U., Olawale, A., Frank, K.B., Rita, O.O. and Martin, O.S., 2020. Design and construction of fixed dome digester for biogas production using cow dung and water hyacinth. *African Journal of Environmental Science and Technology*, *14*(1), pp.15-25.

Upadhyay, A., Kovalev, A.A., Zhuravleva, E.A., Kovalev, D.A., Litti, Y.V., Masakapalli, S.K., Pareek, N. and Vivekanand, V., 2023. A Review of Basic Bioinformatic Techniques for Microbial Community Analysis in an Anaerobic Digester. *Fermentation*, *9*(1), p.62.

Uyaguari-Diaz, M.I., Chan, M., Chaban, B.L., Croxen, M.A., Finke, J.F., Hill, J.E., Peabody, M.A., Van Rossum, T., Suttle, C.A., Brinkman, F.S. and Isaac-Renton, J., 2016. A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples. *Microbiome*, *4*, pp.1-19.

Van der Walt, A.J., Van Goethem, M.W., Ramond, J.B., Makhalanyane, T.P., Reva, O. and Cowan, D.A., 2017. Assembling metagenomes, one community at a time. *BMC genomics*, *18*, pp.1-13.

Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J. and Tyson, G.W., 2014. Linking microbial community structure, interactions, and function in anaerobic digesters using new molecular techniques. *Current opinion in biotechnology*, *27*, pp.55-64.

Vendruscolo, E.C.G., Mesa, D., Rissi, D.V., Meyer, B.H., de Oliveira Pedrosa, F., de Souza, E.M. and Cruz, L.M., 2020. Microbial communities network analysis of anaerobic reactors fed with bovine and swine slurry. *Science of the Total Environment*, *742*, p.140314.

Wu, J., Zhang, J., & Liu, L. (2020). Molecular characterization and functional analysis of indolepyruvate oxidoreductase gene from anaerobic rumen microbes. *Frontiers in Microbiology*, 11, p. 1527

Yan, Y., Zhao, C., Wang, H., & Li, X. (2020). Functionality of indolepyruvate oxidoreductase in amino acid metabolism in methanogens. *Journal of Bacteriology*, 202(6), e00624-19.

© University of Venda

Yoshitake, K., Kimura, G., Sakami, T., Watanabe, T., Taniuchi, Y., Kakehi, S., Kuwata, A., Yamaguchi, H., Kataoka, T., Kawachi, M. and Ikeo, K., 2021. Development of a time-series shotgun metagenomics database for monitoring microbial communities at the Pacific coast of Japan. *Scientific reports*, *11*(1), p.12222.

Zain, M.M. and Mohamed, A.R., 2018. An overview on conversion technologies to produce value added products from CH4 and CO2 as major biogas constituents. *Renewable and Sustainable Energy Reviews*, *98*, pp.56-63.

Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H. and Yi, W., 2021. Advances in metagenomics and its application in environmental microorganisms. *Frontiers in microbiology*, *12*, p.766364.

Zhang, L., Loh, K.C., Lim, J.W. and Zhang, J., 2019. Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters: A review. *Renewable and Sustainable Energy Reviews*, *100*, pp.110-12

Zhou, Y., Liu, M. and Yang, J., 2022. Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: methods, applications, challenges, and opportunities. *Microbiological research*, *260*, p.127023.

© University of Venda