Metagenomics survey unravel diversity of biogas' microbiome with potential to enhance its' productivity in Kenya

Muturi S.M.[1,4*], Muthui L.W.[2], Njogu P.M.[3], Onguso J.M.[4], Wachira F.N.[5], Opiyo S.O.[6,7], and Pelle R.[2]

[1]Department of Biological Sciences, University of Eldoret, [2]Biosciences eastern and central Africa - International Livestock Research Institute (BecA-ILRI) Hub, [3]Institute for Energy and Environmental Technology, Jomo Kenyatta University of Agriculture and Technology, [4]Institute for Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, [5]South Eastern Kenya University, [6]OARDC, Molecular and Cellular Imaging Center-Columbus, Ohio state University, USA, and [6]The University of Sacread Heart, Gulu, Uganda.

**Corresponding author, email:samuelmuturi37@gmail.com*

ABSTRACT

The obstacle to optimal utilization of biogas technology is poor understanding of biogas' microbiome diversities over a wide geographical coverage. We performed random shotgun sequencing on twelve environmental samples. Randomized complete block design was utilized to assign the twelve treatments to four blocks, within eastern and central regions of Kenya. We obtained 42 million paired-end reads that were annotated against sixteen reference databases using two ENVO ontologies, prior to β-diversities studies. We identified 37 phyla, 65 classes and 132 orders. *Bacteria* richness dominated and composed 28 phyla, 42 classes and 92 orders, conveying substrate's versatility in the treatments. Though, *fungi* and *Archaea* composed 5 phyla, the *fungi* were richer; suggesting importance of hydrolysis and fermentation in biogas production. High β-diversity within the taxa was largely linked to communities' metabolic capabilities. *Clostridiales* and *Bacteroidales*, the most prevalent guilds, metabolized organic macromolecules. *Cytophagales*, *Alteromonadales*, *Flavobacteriales*, *Fusobacteriales*, *Deferribacterales*, *Elusimicrobiales*, *Chlamydiales*, *Synergistales* to mention but few, also catabolized macromolecules into smaller substrates to conserve energy. Furthermore, *δ-Proteobacteria*, *Gloeobacteria* and *Clostridia* affiliates syntrophically regulated P_{H2} and reduced metal to provide reducing equivalents. *Methanomicrobiales* and other *Methanomicrobia* species were the most prevalence *Archaea*, converting formate, $CO_{2(g)}$, acetate and methylated substrates into $CH_{4(g)}$. *Thermococci*, *Thermoplasmata* and *Thermoprotei* were among the sulfur and other metal reducing *Archaea* that contributed to redox balancing and other metabolism within treatments. Eukaryotes, mainly fungi were the least abundant guild, conveying largely *Ascomycota* and *Basidiomycota* species. *Chytridiomycetes*, *Blastocladiomycetes* and *Mortierellomycetes* were among the rare species, suggesting their metabolic and substrates limitations. Generally, we observed that environmental and treatment perturbations influenced communities' abundance, β-diversity and reactor performance largely through stochastic effect. Understanding diversity of biogas' microbiomes over wide environmental conditions and its productivity provides insights into better management and other effective strategies that ameliorate existing biochemical issues.

Keywords: *Survey, Metagenomics, microbiome, diversity, biogas production*

Background

The optimal utilization of micro-organisms in developing new products and improving the existing bioprocesses is at the basis of numerous strategies [1]. One of the widely applied such technologies, utilizing micro-organisms is anaerobic digestion (AD) of wastes materials. The process generates biogas and digestate from wastes, the former being an energy carrier

molecule [2]. Over 30,000 industrial installations globally, have been revealed to utilize AD to convert organic wastes into permanent power production sources, amounting to10,000 MW [3]. Moreover, stable AD processes emerges as one of the efficient environmental model for organic wastes valorization, permitting nutrient recovery from the digestate [4], suggesting another economic benefits of AD systems [5], compared to other methods of wastes managements [6]. In addition, African countries, including Kenya, have proven to harbor huge amount of wastes [5], with high biogas generating potentiality. Despite the identified wastes, the biogas benefits and the efforts played by the Biogas partnership programmes [7], the AD systems in Africa are still fewer and the existing ones continue to be abantedly. The programme has installed approximately 16000 biogas plants in Africa and the installations includes 16,419 in Kenya, 13584 in Ethiopia, and 13,037, 7518, and 6504 in Tanzania, Burkina Faso and Uganda repectively [5, 7]. Though Kenya leads in Africa, the adoption of the technology is still poor, and does not commensulate the country's energy needs. Secondly, the popularity of the technology in the country, compared to its' counterparts, justified Kenya as the appropriate pilot site among other African countries. Similar guidelines were followed to select the studied Kenyan regions. The cause of poor adoption of this technology in Africa largely is due to poor knowledge on the biogas generating microbiomes, the populations that spin the AD process. These processes comprise thermodynamic unfavorable reactions that contribute to reactors' instabilities and poor performance. The microbial populations and the reactions involved are highy sensitive to environmental and reactor operation conditions; the traits that prompt development of careful and sustainable management strategies, for effective biogas production, otherwise, complete breakdown of the whole system occurs. In the events of complete failure, the humble farmers either abandon the systems or undergo a costly restoration process. All these reduce the economic and environmental benefits of the AD systems, and constrain the green economy concepts.

Although, previous reports in other continents, indicates dominance of Bacterial and Archaeal species to tolerate reactor perturbations, limitations of traditional culture-dependent and independent methods e.g 16S rRNA gene [8-9], still limit the knowledge on global biogas' microbiomes diversities within these systems. The techniques provide low resolution [10, 11, 12], not enough to unravel high level of microbial diversity in any ecosystems. All these limit sufficient microbial composition and diversity information availability; further constraining biogas systems' improvement and better management. The 16S rRNA gene does not reveal full extent of microbial diversity in an ecosystem [13-17]. The current diversity estimates are extrapolations of empirical scaling laws [13], and other theoretical models of biodiversity [14]. The weakness limit the possibility of scrutinizing and applying the biodiversity scaling laws, macroecological theories and other ecological theories [12-14] in the context of biogas production; that directly constrains the available microbial diversity knowledge; which affect severely, the developed mitigition strategies that targets biochemical reaction failures. Despite covering a small fraction of microbial diversities in a datasets, marker gene based studies, annotate the detected species against their biased amplicon sequence databases. Consequently, conflicts arise on the links between microbial populations and reactor performance. One of such conflicts emerged from Wittebolle *et al*. [18] and Konopka *et al* [19]. Wittebolle *et al*. considered higher microbial diversity as a reservoir of redundant metabolic pathways that possess desirable traits that respond to reactor pertubations [18] and other environmental conditions. In contrast, Konopka *et al*. argued that less diverse communities confer system stability by expressing complementary pathways [19] that avoid direct competition over the available resources. These kinds of findings prompted us to ask whether microbial diversity in biogas reactor matter or it is the existence of the core

microbiomes that determine the AD stability and maximum biogas production. Answers to these questions provoked us to further characterize the system's micro-organisms using next generation sequencing platforms also termed "massively parallel sequencing" over a wide geographic area. Previously, the technology has enabled massive sequencing of environmental genomes with high quality and accurancy [20]. However, application of the technologies in the biogas systems is still at its infant stage, whereas the AD stability and biogas production still continue to be abantedly, due not much microbial composition information available. Furthermore, eukaryotes, mainly fungi are normally neglected by the majority of the studies, yet the species are very crucial in metabolism of large macromolecules in the AD systems.

In this paper, Illumina's sequencing-by-synthesis technology was used to generate sequence datasets from environmental samples that were assigned to twelve treatments. Randomized complete block design was the key experimental design employed to distribute the twelve treatments to our four blocks. The generated sequences were *de novo* assembled and annotated against 16 references biological databases. Two levels of annotation: nucleotide and protein level annotation and two ENVO ontologies: large lake and terrestrial biomes were utilized. Specifically, the experiments were designed to increase detection sensitivity of biogas' microbiomes over wide environmental conditions and test whether the composition, richness and abundance within and among the treatments were significantly different; extending the same to different taxonomic groups. Our result suggests high α and β-diversities among the treatment, with few exhibiting low β-diversities, which convey unpredictable microbial changes. We postulated that environmental and other operation conditions enhanced ecological drift simultineously weakening the microbiome's niche selection ability that introduced stochastic effects into the systems. The outcome of all these was a variation of microbial composition and biogas production among the treatments. The findings of this study contribute to formation of testable hypotheses for future application of metacommunity theory on the links between microbial diversity and environmental conditions. Also the study contributes to the understanding of biogas' microbiome for better improvement of biogas production.

Materials and Methods

Ethics, experimental design, reactor descriptions, sampling and geographic co-ordinates

The sampled treatment sites (reactors) were installed in the premises of privately owned land (Table 1) and permission to sample was granted by the owners. Randomized complete block design was utilized as previously described [21] to assign the twelve cowdung digesting treatments to the four blocks, which operated at mesophilic conditions. Organic loading rate and the hydraulic retention time were largely influenced by the the treatment size and the farmers' routine activities. Sampling was conducted following the previously described procedure of Boaro *et al*. [22]. Geographic co-ordinate (Table 1) and agroecological zones (Table 2) were generated by GPS (etrex, Summit^{HC}, Garmin) and ArcGIS® software (ver10.1) [23] respectively. A pre-tested questionnaire was utilized according to Oksenberg *et al*. [24] protocol. Small scale farming and zero grazing of Freshian breeds were the main activities among the sampled farmers. However, the owner of reactor 3 and 11 (Meru and Kiambu blocks), were found to keep Jersey breeds. The selected blocks alternate warm and cool climatic conditions.

Genomic DNA extraction, Qualification, Quantification and Sequencing

The double stranded deoxyribonucleic acids (dsDNA) was extracted from the duplicate sludge samples using ZR soil microbe DNA kit (Zymo Research, CA., USA), according to manufacturers instructions. The quantity and quality of the extracted dsDNA were conducted as previously described by Campanaro *et al*. [25]. The dsDNA were normalized and library

preparation performed using Nextera™ DNA sample prep kit (Illumina, San Diego, CA, USA), according to manufacturers instructions. Library normalization was conducted using Illumina TruSeq DNA PCR-free sample preparation kit buffer (Illumina, USA). To prepare for cluster generation and sequencing, equal volumes of normalized libraries were pooled, diluted in hybridization buffer and denatured, according to Nextera™ protocol. *De novo* deep sequencing was conducted by an Illumina MiSeq system as previously described by Caporaso *et al*. [26] but adopting a v3 kit chemistry with 300 cycles (2×150) as previously described [25].

Bioinformatics and Statistical analysis

All paired-end reads in FASTQ format were pre-processed with FastQC software (ver0.11.5) following the procedure described by Zhou *et al*. [27]. Trimmomatic software was utilized to trim and filter the adaptor according to [28]. *De novo* assembly of the paired-end contigs were peformed by MetaSpade software (ver3.10) as described previously [29]. To re-trim low-quality regions of the scaffolds, we utilized SolexaQA as described by Cox *et al*. [30] while DRISEE, metagenomic quality assessment software was used to remove duplicated scaffolds according to Keegan *et al*. [31]. We screened near-exact scaffolds that matched human, mouse and cow genomes following Langmead *et al*. procedure that used a sort read aligner, the Bowtie [32]. M5nr [33], a non-redundant tool, FragGeneScan [34] and sBLAT [35] tools were utilized to compute datasets against 16 reference biological databases as previously described by Wilke *et al.* [33]. Two ENVO ontologies: large lake and terrestrial biomes were utilized according to Field *et al*. [36] and Glass *et al*. [37]. Alignment similarity cutoff were set at single read abundance; ≥60% identity, 15 amino acid residues and maximum e-value of $1e^{-5}$. The annotated datasets were archived in CloVR [38]. Prior to statistical analysis, the rarefaction curves were reconstructed, datasets was log2(x+1) transformed and normalized using an R package, DESeq software [39]. PCoA based on euclidean model was set at $P \leq 0.05$, and conducted against the twelve treatment samples. Results were visualized by MG-RAST pipeline (ver4.01) [40] and Krona radial space filling display [41]. Optimal scripts and restful API were utilized in our analysis.

Results

i). Overview of metagenome, annotation and taxonomic assignment

We used Illumina based shotgun metagenomics to identify and compare microbial communities within and among the treatments. In total we analysed sludge samples that were assigned to the twelve steady-state operating treatments systems. The procedure yielded a total of 44 million raw reads and out of these, 2 million reads were excluded from downstream analysis due to poor quality. We revealed 24,189,921 reads of 35-151bps from the remaining 42 million reads, which were further filtered (based on the %G:C content bias and duplicates) according to Gomez-Alvarez *et al*. [42]. The procedure revealed approximately 23,373,513 total reads that conveyed a range of 1,261833 to 3,986675 reads per the treatment (SFig. 1a). Thereafter, we trimmed and assembled the contigs using *de novo* approach and the obtained scaffolds (3,819,927 scaffolds, STable 1) were annotated against the sixteen databases and the two ENVO ontologies. Taxonomic assignments conveyed by SEED database were the highest (180, 000 scaffolds), while the assignment of NOG, LSU, RDP and COG databases comprised the least annotations (<8000 scaffolds). Among the database, IMG, Genbank and TrEMBL databases conveyed almost equal scaffold annotations (144,000 scaffolds). Other databases annotated the reads as indicated in Fig. 1. Between the two ontologies, large lake biome revealed the highest annotation compared to the terrestrial biomes with a precision statistical threshold of 95% confidence. Furthermore, 39.75% to 46.70% and 53.07% to 60.11% genes were conveyed to known and unknown proteins.

Table 1: Geographic co-ordinate of the sampled reactors in eastern and central regions of Kenya

Reactor Number	Sample id	Blocks	Treatment	Northings	Eastings	Altitude	Location
EA12/11/09/004	S_1	Kiambu	L-C	S(00°15.477')	E (37°38.828')	1468 M	Ngenda
EA12/10/07/003	S_2	Kiambu	S-C	S(00.21347°)	E(37.61867°)	1674 M	Karimoni
EA12/10/04/001	S_3	Meru	S-E	S(00°08.180')	E(37°39.477')	1443 M	Kaithe
CE20/12/01/020	S_4	Kirinyaga	S-C	S(00°32'45.97")	E(37°26'40.52")	1364 M	Murinduko
EA12/10/08/004	S_5	Meru	L-E	N(00°01.997')	E(37°35.353')	1958 M	Kithaku
EA12/10/09/002	S_6	Meru	M-E	N(00°04.046')	E(37°38458')	1650 M	Ntima
EA12/11/12/002	S_7	Meru	L-E	N(00.02946°)	E(37.10077°)	1987 M	Ontulili
EA14/11/05/001	S_8	Embu	L-E	S(00°033.269')	E(37°28.247')	1306 M	Municipality
EA14/11/09/001	S_9	Embu	M-E	S(00°27.824')	E (37°32.537')	1417 M	Gaturi South
CE12/10/06/012	S_10	Kiambu	L-C	S(00.24874°)	E(37.62438°)	1589 M	Kianyau
CE12/10/07/013	S_11	Kiambu	S-C	S(00°16'17.57")	E(37°38'40.16")	1577 M	Dururumo
CE14/11/11/001	S_12	Kirinyaga	M-C	S(00.48637°),	E (37.44057°)	1539 M	Ngandori

Legend: **CE**: Central regions, **EA**: Eastern regions, **L-C**-Large size, Central regions, M-C: Medium size, Central region, **S-C**: Small size, Central regions, **L-E:** Large size, Eastern regions, M-E: Medium size, Eastern regions **S-E**: Small size, Eastern regions.

Table 2: Environmental and Phenotypic features of the sampled reactors in eastern and central regions of Kenya

		S-T	AEZ		Ann.	Mean	Livestock kept			Family size			Reactor Performance				
R-Id	Alt.			Ann. Av. R-f	Temp ^{0}C		CO	PO	PI	M	F	T	IG	ACG	RV	FACoG	GGV
S_1	1468 M	Lithosol	UM 4	800- 900	20.7-19.9		4	-	-	5	4	9	4.8	4.0	16M3	14M3	2
S_2	1674 M	Ferralsol	UM 2	1100-1400	19.5-18.4		3	-	-	2	1	3	2.4	2.0	8M3	7M3	1
S_3	1443 M	Nitosol	UM 3	1400-2200	20.6-19.2		11	-	-	3	4	7	2.4	2.0	8M3	5M3	3
S_4	1364 M	Andosol	UM 2	1220-1500	20.1-19.0		6	-	-	4	2	6	1.8	1.5	6M3	4.5 M3	1.5
S_5	1958 M	Phaeozem	L H 1	1700-2600	17.4-14.9		3	-	-	3	1	3	4.8	4.0	16M3	12M3	4
S_6	1650 M	Nitosol	UM1	1650-2400	19.2-17.6		2	30	-	2	2	4	3.0	2.5	10M3	6M3	4
S_7	1987 M	Phaeozem	LH 1	1700-2600	17.4-14.9		5	-	-	5	2	7	4.8	4.0	16M3	10M3	6
S_8	1306 M	Nitosol	UM 3	1000-1250	20.7-19.6		5	-	6	5	4	9	4.8	4.0	16M3	11M3	5
S_9	1417 M	Andosol	UM 2	1200-1500	20.1-18.9		2	-	-	3	2	5	3.0	2.5	10M3	8M3	2
S_10	1589 M	Acrisols	UM 2	1100-1400	19.5-18.4		3	-	-	2	2	4	4.8	4.0	16M3	13M3	3
S_11	1577 M	Cambisol	UM 3	800-1200	19.9-19.5		14	-	-	5	4	9	2.4	2.0	8M3	6M3	2
S_12	1539 M	Ferralsol	UM1	1400-1700	19.3-17.5		3	-	-	4	3	7	3.0	2.5	10M3	8M3	2

Key: **S-T**: Soil type; **AEZ**: Agro-ecological zones; **UM 3**: Marginal Coffee zone; **UM 1**: Coffee-Tea Zone; **UM 2**: Main coffee zone; **LH 1**: Tea-Dairy Zone; **R-F**: Rainfall; **Temp:** Temperature; **CO**: Cows; **PO**: Poultry; **PI**: Pigs; **M**: Male; **F**: Female; **T**: Total; **IG**: Ideal gas; **ACG**: Actual calculated gas; **RV:** Reactor Volume; **FACoG**: Farmers Actual Collected gas; D**GG:** Difference between the RV and FACoG; **Alt**: Altitude,

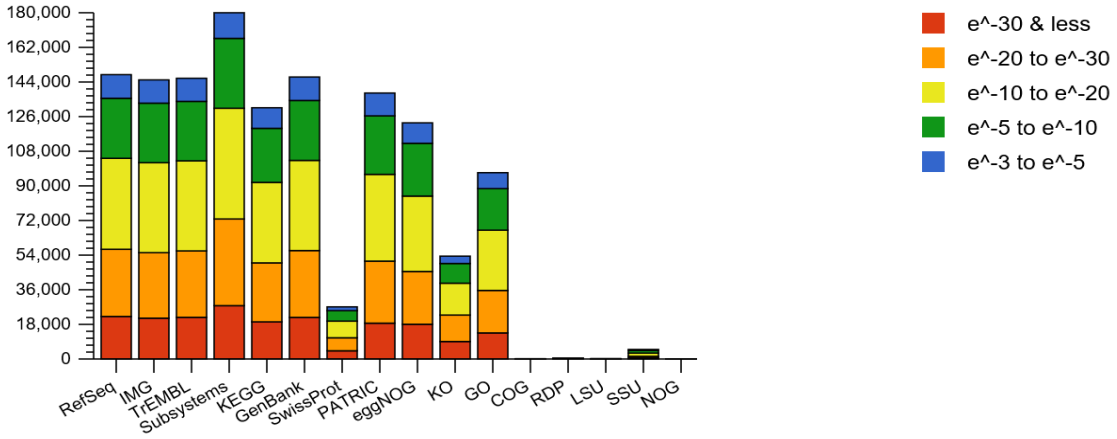

Fig. 1: Stacked bar charts showing the source of hits distributions from the sixteen databases: These include nucleotide and protein databases, protein databases with functional hierarchy information, and ribosomal RNA databases. The bars representing annotated reads are colored by e-value range. Different databases have different numbers of hits but can also have different types of annotation data.

Taxonomically four broad but relatively simple distributions of microorganisms were revealed, with the species of three domains being responsible for methanogenesis. In summary, 2, 617,791 reads hits were generated and 213681 hits were assigned to *Archaea*, 2379293 hits to *Bacteria* and 21886 hits *Eukaryotes* (SFig. 1a). Thus the revealed abundance hits ratio of *Bacteria* to *Archaea* to *Eukaryotes* in the treatments was 109:9:1. The observed high abundances of Bacterial species were linked to their metabolic capabalities. Unlike other cellular organisms, *Eukaryotes* mainly fungi were the least abundances (< 1% of the total hits) among the treatments. Richness, another microbial community parameter, was consistent among the studied treatments. In total 37 phyla, 73 classes and 132 orders were conveyed, although high community β-diversity among the majority of the treatments was the key observation, except for the communities detected in reactor 1, 3 and 6 that were found to cluster partially on the upper right quadrant of the plot, Fig. 2), at a $P \leq 0.05$. All the identified phylotypes were further annotated and β-diversity determined in details.

The *Bacteria* richness, abundances and β-diversities,

Further, we evaluated the abundance and composition variation of 92 Bacterial domain orders that were assigned to 41 classes and 21 phyla (SFig. 2a, Table 3). Generally we observed significantly higher Bacterial abundances (89-93% of the total hits), compared to other domains. Similarly, significantly high β-diversities of the *Bacteria* communities were also observed at the phylum level. However, the communities of reactor 2 and 10 (within the same block) and reactor 4 and 9 (on different blocks) were found to reveal low β-diversity (SFig. 2b). In addition, we further revealed low divesity variabilities between communities of reactor 3 and 6 and those of reactor 8 and 11 when compared at lower (class and order) taxonomic ranks (SFig. 3a and 3b). Hence, only four out of twelve treatments had high β-diversities, which could indicate genetic plasticity due to species evolution.

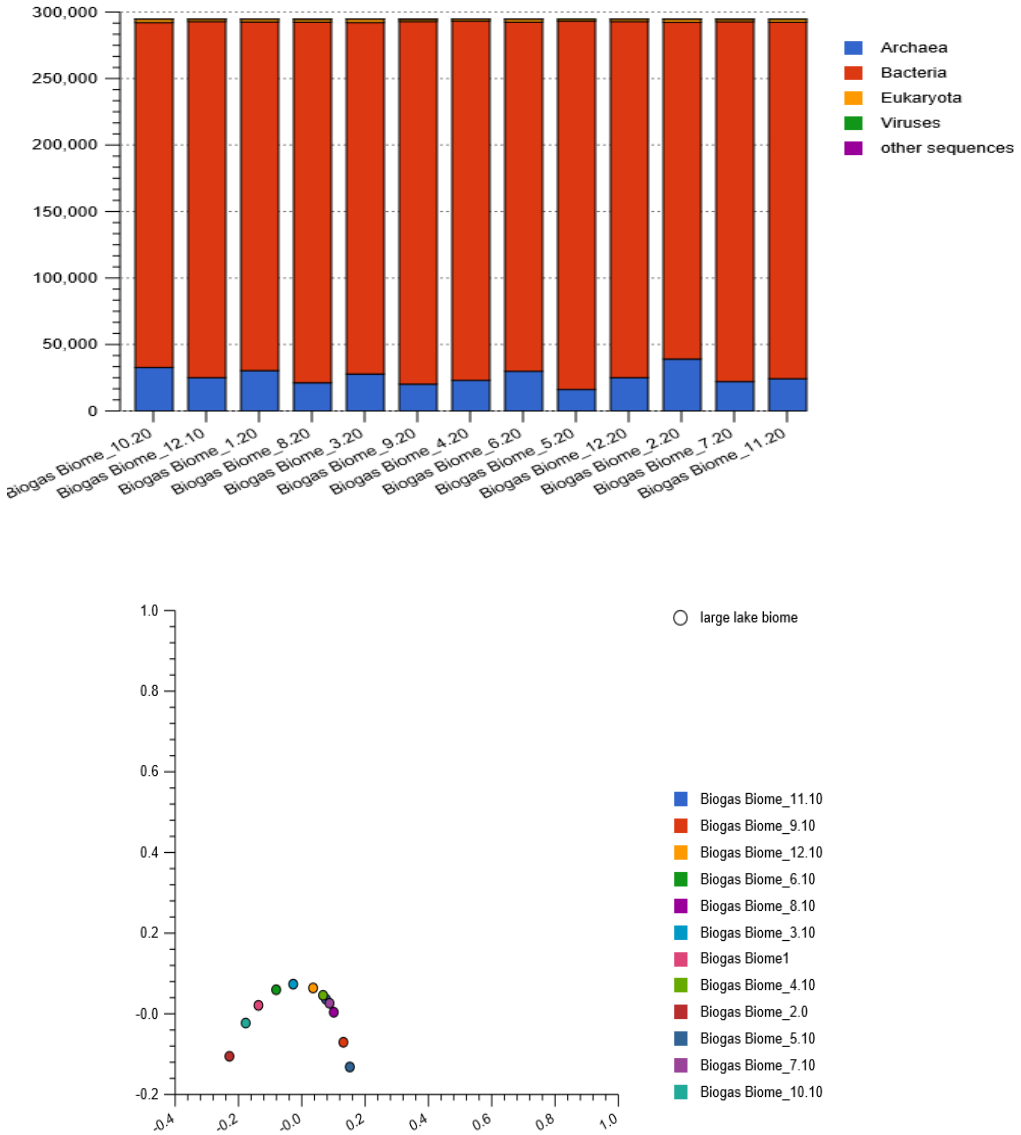

Fig. 2: The stacked barchart (a) showing the three main domains, proportion of relative abundances and their PCoA plot (b) based on the Euclidean model. The PCoA plot revealed dissimilarities among the studied reactors. However, only three reactors partially clustered in the upper right quadrant of the plot. Though the Viruses nucleotide contributed to the curve they were not considered in the downstream analysis.

Among the species, the affiliates of *Firmicutes* were the most abundant in seven out of twelve treatments and comprised 15-34% of total hits, Table 3). However, their counterparts, the *Proteobacteria* species dominated in the other five treatments, composing 19-32% of total hits, (Fig. 3 and Table 3). Contradicting the findings, were the *Proteobacteria* richness that superseded the richness of *Firmicutes* species. The communities of *Proteobacteria* were classified into 38 orders, which were affiliated to 7 classes (SFig. 4a). We linked their versatility to their ability of circumventing the oxidative stress in the systems. Significantly high similarity of microbial composition were revealed among the treatment 1, 3 and 6, and treatment 4, 7 and 12, all located on the upper and lower right quadrant of the plot, (SFig. 4b and 5b; at the class and order level). However, low β-diversities were observed for the identified *Proteobacteria* communities of reactor 4, 7, 8 and 9, (SFig. 5) at the order level. Other treatments were found to have high β-diversities of *Proteobacteria* species.

Table 3: Detection of microbiome composition and abundances in the twelve reactor within eastern and central regions of Kenya

Phylum	S_11.10	S_9.10	S_12.10	S_6.10	S_8.10	S_3.10	S_1	S_4.10	S_2.0	S_5.10	S_7.10	S_10.10
Firmicutes	27%	34%	26%	21%	25%	22%	22%	31%	15%	30%	29%	15%
Proteobacteria	26%	19%	20%	27%	24%	27%	25%	19%	31%	25%	21%	32%
Bacteroidetes	15%	18%	22%	11%	18%	11%	15%	20%	10%	16%	18%	10%
Actinobacteria	4%	3%	5%	5%	4%	4%	4%	4%	7%	5%	7%	6%
Chloroflexi	3%	2%	3%	6%	4%	6%	4%	3%	7%	2%	3%	7%
Spirochaetes	2%	3%	1%	3%	2%	3%	4%	2%	2%	1%	1%	1%
Verrucomicrobia	2%	2%	1%	2%	1%	2%	2%	1%	2%	0.9%	1%	2%
Cyanobacteria	2%	1%	2%	2%	2%	2%	2%	2%	3%	1%	2%	3%
Planctomycetes	2%	1%	2%	3%	2%	3%	3%	2%	3%	2%	2%	3%
Chlorobi	1%	1%	1%	1%	2%	1%	1%	1%	1%	0.9%	1%	1%
Acidobacteria	1%	1%	1%	2%	1%	2%	1%	0.9%	2%	0.8%	1%	2%
Thermotogae	1%	0.9%	1%	1%	1%	1%	1%	0.9%	1%	0.8%	1%	1%
unclassified reads	0.6%	0.6%	1%	1%	2%	1%	0.9%	1%	0.6%	1%	1%	0.7%
Synergistetes	0.8%	0.6%	1%	0.7%	0.7%	0.6%	0.6%	0.8%	0.8%	0.6%	0.8%	0.9%
Lentisphaerae	0.8%	0.5%	0.3%	0.4%	0.3%	0.4%	0.4%	0.3%	0.3%	0.2%	0.4%	0.2%
Fusobacteria	0.6%	0.8%	0.6%	0.5%	0.6%	0.5%	0.5%	0.7%	0.4%	0.7%	0.7%	0.4%
D. Thermus	0.5%	0.4%	0.6%	0.9%	0.6%	0.8%	0.7%	0.6%	1%	0.5%	0.6%	1%
Aquificae	0.5%	0.4%	0.4%	0.5%	0.5%	0.5%	0.5%	0.4%	0.4%	0.3%	0.4%	0.5%
Tenericutes	0.3%	0.6%	0.3%	0.2%	0.4%	0.3%	0.2%	0.3%	0.2%	1%	0.5%	0.3%
Deferribacteres	0.3%	0.3%	0.3%	0.4%	0.4%	0.4%	0.3%	0.3%	0.3%	0.2%	0.3%	0.3%
Nitrospirae	0.3%	0.2%	0.3%	0.4%	0.4%	0.4%	0.4%	0.3%	0.4%	0.2%	0.3%	0.4%
Chlamydiae	0.3%	0.2%	0.1%	0.2%	0.2%	0.1%	0.2%	0.1%	0.2%%	0.08%	0.2%	0.1%
Fibrobacteres	0.2%	0.5%	0.1%	0.2%	0.1%	0.2%	0.2%	0.2%	0.08%	0.2%	0.2%	0.08%
Dictyoglomi	0.2%	0.2%	0.2%	0.4%	0.3%	0.4%	0.3%	0.3%	0.3%	0.2%	0.2%	0.3%
Elusimicrobia	0.2%	0.2%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%
Chrysiogenetes	0.08%	0.06%	0.07%	0.07%	0.07%	0.09%	0.06%	0.07%	0.09%	0.05%	0.07%	0.09%
Gemmatimonadete	0.06%	0.06%	0.07%	0.1%	0.1%	0.1%	0.1%	0.07%	0.2%	0.6%	0.08%	0.2%
C. Poribacteria	0.03%	0.03%	0.03%	0.05%	0.03%	0.05%	0.03%	0.03%	0.05%	0.02%	0.03%	0.04%

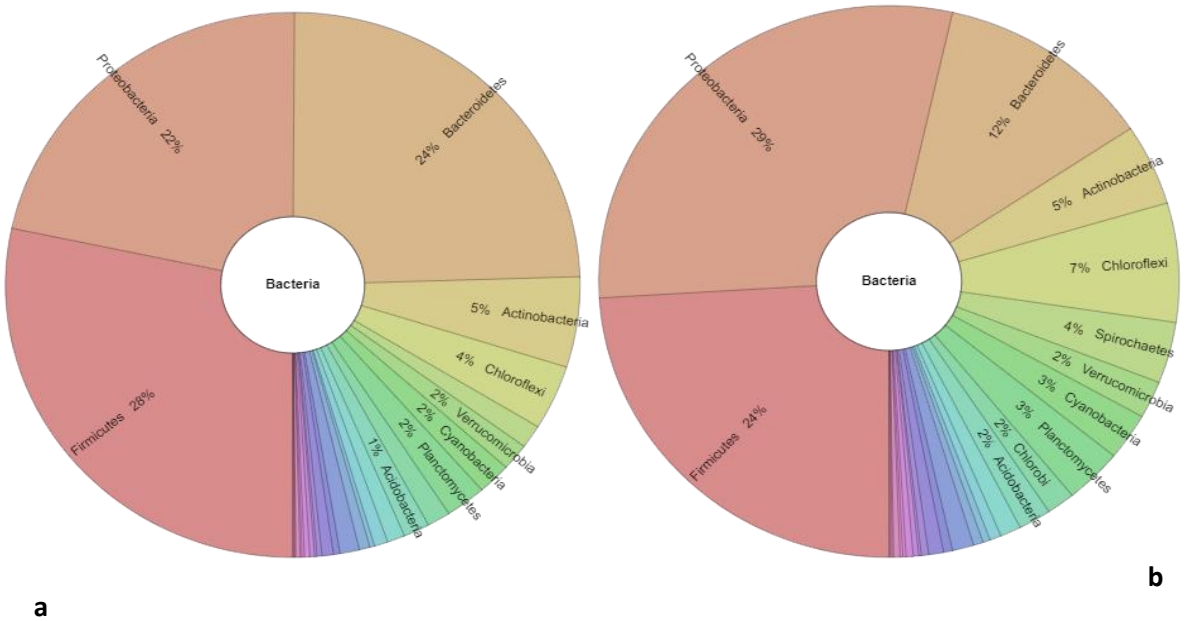

Fig. 3: The Krona radial space filling display showing the relative abundances of the Bacterial reads in reactor 12 and 3, and their annotation: a) *Firmicutes* reads dominated in reactor 12, b: *Proteobacteria* reads dominated in reactor 3.

Specifically, we revealed *δ-Proteobacteria* species as the most abundant guild, conveying 9-14% of the total microbial communities. Globally the species were the second most abundant after the *Clostridial* species that composed 11-26% of the total communities. The *δ-Proteobacteria* communities were classified into 7 orders (SFig. 6a). Among them were *Syntrophobacteriales* that composed 2-4% of the total identified communities, followed by *Desulfuromonadales* and *Myxococcales* species that composed 2-4% and 0.9-4% of the total microbial compositions respectively. *Desulfobacterales* and *Desulfovibronales* communities were found to compose 1-2% of the total species abundances. Among the low abundances *δ-Proteobacterial* species were the affiliates of *Desulforculales* and *Bdellovibrionales* that comprised < 1% of the identified total communities. The non-classified reads derived from the class *δ-Proteobacteria*, also composed < 1% of the total communities. However, the PCoA plots indicated high sample to sample variation of the *δ-Proteobacteria* communities, although the composition of reactor 4 and 7 were found to be similar (SFig. 6b).

The communities' affilliated to the class *γ-Proteobacteria* conveyed 4-11% of the total relative abundances. The affiliates were the fifth most abundant group in the entire microbiome and were classified into 14 orders (SFig.7a). Among the identified orders were *Pseudomonadales*, that dominated the class, composing <1 to 6% of the total communities. The abundances of these phylotypes were followed by *Alteromonadales* and *Enterobacterales* species. The affillitates of *Chromatiales*, *Vibrionales*, *Methylococcales*, *Oceanspirillales* and *Pasteurelales* respectively were also among the detected species. Other identified orders were *Xanthomonadales*, *Thiotrichales*, *Legionellales*, *Aeromonadales*, *Acidithiobacillales* and *Cardiobacterales*, all conveying <1% of the total relative's abundances. Similarly, we also identified unclassified *γ-Proteobacteria* reads. In addition, high β-diversity ($P \leq 0.05$) of the communities in the six treatments was revealed. In contrast, the other six communities identified in reactor 1, 3, 4, 7, 8 and 9 were found to clustered on the upper right quadrant of the PCoA plot (SFig.7b), indicating genetic composition similarity within the population.

Low abundances (2-4% of the identified microbial communities) of the *β-Proteobacteria*

communities were revealed. The species were classified into 7 orders (SFig. 8a). *Burkholderiales* affiliates conveyed 1-2% of the microbial populations while the affiliates of *Rhodocyclales*, *Neisseriales*, and *Nitrosomonadales* species were found to compose <1% of the total population respectively, across the treatments. Other identified *β-Proteobacteria* species were assigned to *Methylophilales*, *Gallionellales* and *Hydrogenophilales* orders and they composed much lower abundance of <1% of the total identified communities among the twelve treatments. High sample to sample variation was also observed for the *β-Proteobacteria* species (SFig. 8b).

Like *δ* and *β-Proteobacteria*, the affilates of *α-Proteobacteria* communities were revealed to compose 3-8% of the total microbial commpositions, and were classified into 7 orders (SFig. 9a). In this case the *Rhizobiales* dominated, comprising 1-2% of the total microbial populations. The relative abundances of *Rhodobacterales*, *Rhodospirillales*, and *Sphingomonadales* followed, representing <1% of the total population. Other subsequent *α-Proteobacteria* species identified were the affiliates of *Caulobacterales*, *Rickettsiales* and *Parvularculales* orders. We also found significantly high dissimilarities among the *α-Proteobacteria* within the majority of the studied treatments. Only the communities of reactor 3 and 11 were revealed to express low β-diversities (SFig. 9b), the trait that was attributed to similar treatment conditions (Table 1 and 2).

The affiliates of ε and *Z-Proteobacteria* were among the least abundant *Proteobacteria* phylotypes and were found to comprise ≤1% of the total identified populations. The *ε-Proteobacteria* communities were classified into *Campylobacteriales* and *Nautiliales* species (SFig. 10a), both conveying <1% of the total populations. However, we found similarities between the communities of reactor 4 and 8 (clustered) while the communities of reactor 5 and 9 (clustered partially) revealed low β-diversity (SFig. 10b). All the identified *Z-Proteobacteria* were the *Mariprofundales* affillites, and comprised <1% of the total microbial abundances in the studied treatments.

The identified *Firmicutes* communities were the second most diverse guild and connveyed 15-34% of the total microbial compositions. The communities were classified into 8 orders all grouped into 4 *Firmicutes* classes (SFig.11a and 12a). The *Firmicutes* in the three treatments; reactor 2, 10 and 12 were revealed to clustered in the lower left quadrant of the plot, indicating significant similarity of the genetic composition among the treatments. However, their community species in reactor 1 and 7 were found to cluster partially on the upper right quadrant of the plot (significant threshold; $P \leq 0.05$; SFig.11b), indicating low β-diversity. At the order level, the identified *Firmicutes* of reactor 2 and 10; 3 and 6; 7 and 12; and reactor 1 and 9, were also revealed to exhibit low β-diversities (SFig.12b). Other local communities were revealed to express high β-diversities at class and order levels.

The *Clostridia*, one of the most abundant *Firmicutes* class, composed 8-22% of the identified total communities and were classified into 4 orders (SFig. 13a). *Clostridiales* species dominanted the entire microbiome communities and constituted 8-22% of the total communities. *Thermoanaerobiales* species were the second most abundance within the class and constituted 2-3% of the total communities' abundances. Other identified low abundances *Clostridia* species were the affiliates of *Halanaerobiales* and *Natranaerobiales*, all represented <1% of the total microbial populations, albeit the *Halanaerobiales* communities in reactor 11 that conveyed 2% of the total populations. Among the twelve treatments, three treatments (i.e reactor 7, 11 and 12) clustered on the lower left quadrant of the plot, (SFig.13b), indicating their communities had high similarities. Contrary, the communities of reactor 5 and 9 were found to cluster partially on the upper left quandrant of the PCoA plot,

an indication of significantly low β-diversity. The other seven treatments expressed high β-diversites, which implied metabolic and genetic plasticity among the identified *Clostridia* species. The class *Bacilli* represented 4-9% of the total communities, and were the fourth most abundant among the entire microbiome populations. The *Bacilli* were classified into *Bacillales* and *Lactobacillales* (SFig.14a), which conveyed 3-4%; and 1-4% of the total identified communities respectively. Among the communities of the treatments only those of reactor 5 and 8 that were found to exhibit significant dissimilarities, while other communities expressed pairwise occurrence which indicated a low β-diversity of the respective communities (SFig.14b). Other detected low abundance *Firmicutes* were *Erysipelotrichales* (*Erysipelotrichi*) and *Selenomonadales* (*Negativicutes*) and they all composed <1% of the total identified communities. Nonetheless, the *Erysipelotrichales* species in reactor 5 and 9 were found to compose 1% of the total microbial abundance, implying the conditions were more favourable compared to the conditins in the other treatments.

*Bacteroidete*s, the third most abundant phylotype composed 10-22% of the total microbial population and were classified into 4 classes and orders (SFig.15a and SFig.16a). Among them, were the *Bacteroidia* that conveyed 5-12% of the microbial communities and all belonged to the *Bacteroidales*. Other members were *Flavobacteria* that belonged to *Flavobacterales* order, and composed 2-3% of the total populations. *Cytophagia* communities were all assigned to *Cytophagales* species and represented 1-2% of the total communities. Lastly, the identified *Sphingobacteria* communities were all affiliated to *Sphingobacteriales*, conveying 2-3% of the total microbial abundance. We observed partial clustering of the *Bacteroidete*s communities identified in reactor 2 and 10 and those in reactor 8 and 12, which indicated low β-diversity between the respective communities (SFig. 15b and 16b). However, the communities of reactor 4 and 9 were revealed to cluster on the lower left quandrant of the plot, (SFig 15b and 16b), implying significant similarities between the two treatments.

The identified *Actinobacteria* species were the affiliates of 6 orders (SFig. 17a) and conveyed 3-7% of identified reactor communities. The *Actinomycetales*' were among the identified affiliates and were revealed to compose 2-6% of the total communities. The *Coriobacteriales* and *Bifidobacteriales* affiliates respectively followed, and were found to comprised <1% of the microbial compositions. Other much lower abundant *Actinobacteria* species were assigned to order *Rubrobacteriales*, *Solirubrobacteriales* and *Acidimicrobiales* respectively. Out of twelve treatments, the communities of the seven treatments (reactor 1, 3 and 12; reactor 5 and 11 and reactor 7 and 8; SFig.17b) were found to exhibit significantly ($P \leq 0.05$) low β-diversities. Only three out of twelve treatments were found to convey high β-diversities of *Actinobacteria* species.

The *Chloroflexi* were categorized into 4 classes (SFig.18a) which were further classified into 5 orders (SFig.19a). The phylotypes composed 2-7% of the microbial communities as indicated (Table 2). The communities were classified into *Chloroflexi* (1-4% of the communities) and *Dehalococcoidetes* (≤1-3% of the communities) as the most abundant phylotypes. Although all the *Dehalococcoidetes* reads were unclassified at a lower level, we identified *Chloroflexales* and *Herpetosiphonales* as the affiliates of *Chloroflexi* class (SFig. 20). The two orders composed 1-2% and <1% of the total microbial composition respectively. Other identified classes of the phylum were *Thermomicrobia* and *Ktedonobacteria* and they conveyed <1% of the total communities. The *Thermomicrobia* communities were classified into *Sphaerobacterales* and *Thermomicrobiales* orders, (SFig. 21) and all represented <1% of the total communities. Nevertheless, all the identified *Ktedonobacteria* community were classified as *Ktedonoacteriales* species (also representing

<1% of the communities). We had identified two treatments (reactor 1 and 7) that revealed low β-diversities of their local communities (SFig. 18b). However, four treatments (reactor 4, 5, 10 and 12), out of the remaining ten were revealed to cluster on the lower right quandrant of the PCoA plot (SFig. 18b and 19b). Similarly, the communities of reactor 1 and 5 and those of reactor 3 and 7 were found to cluster on the upper right and lower right quandrant of the PCoA plot, when the affiliates of *Chloroflexi* class were considered (SFig. 20b). The *Chloroflexi* class communities of reactor 1 and 10 were found to cluster partially with the cluster of reactor 3 and 7. We also identified partial clustering of reactor 8, 11 and 12 communities which indicated low β-diversities (SFig. 20b) while the rest of the treatments were revealed to have high β-diversities. In contrast, the *Thermomicrobia*'s communities among the treatments were significantly dissimilar as indicated (SFig. 21b).

The *Cyanobacteria* guild composed 1-3% abundance of the total microbial compositions and majority, composing 1-2% were unclassified. Only <1% of the total population were assigned to *Gloeobacteria* (SFig. 22a). Unlike other phylotypes, only the *Cyanobacteria* communities of reactor 3, 7 and 10 were found to exhibit significantly high β-diversity ($P \leq 0.05$), (SFig. 22b) while the communities in the other treatments were found to express partial similarities between or among themselves. The *Gloeobacteria* communities were all classified as *Gloeobacteriales* order. Other identified communities (unclassified reads) were the affiliates of *Chroococcales*, *Nostocales*, *Oscillatoriales* and *Prochlorales* orders (SFig. 23a) respectively. Majority of the local *Cyanobacteria* communities in their respective treatments were found to reveal low β-diversities while the other communities were found to cluster within the plot, indicating significant similarity. An exception was the identified communities of the reactor 4, 6, 9 and 12 that had shown high β-diversities (SFig. 23b). Similar to the four identified phylotypes derived from unclassified reads (SFig. 24a), which were also found to reveal high community dissimilarities among the twelve treatments (SFig. 24b).

The *Acidobacteria* abundances ranged between 1-2% of the total microbial abundances. Majority of these populations were the affiliates of *Acidobacteria* mainly *Acidobacteriales*, comprising <1-2% of the total microbial communities. The remaining species were affiliated to *Solibacteres*, particularly *Solibacterales* order and conveyed <1% of the total microbial communities, (SFig. 25a and 26a). The communities of seven out of twelve treatments were revealed to express low β-diversities. These communities were the habitant of reactor 1 and 3, and reactor 2, 8 and 12 respectively. Keenly we observed clustering of the *Acidobacteria* communities of reactor 2 and 8, on the upper left quadrant of the plot (at order level; SFig. 25b and 26b), which indicated the two local communities were significantly similar compared to their counterparts identified in reactor 12. *Deinococcus-Thermus*, another phylum, had *Deinococcales* and *Thermales* species (SFig. 27a), the two belonging to the class *Deinococci*. The relative abundances of the two orders were ≤1% of the total microbial populations. We observed low β-divesities between the *Deinococci* species of reactor 11 and 12 while the diversity of its communites in reactor 3 and 6 and those of reactor 7 and 9 were found to be significantly similar (SFig. 27b). All the other local *Deinococcus-Thermus* communities expressed high β-diversities.

Furthermore, *Verrucomicrobia* communities were classified into 3 classes that composed 1-2% of the total composition in the treatments. Among the identified classes were *Opitutae*, *Verrucomicrobiae* and *Spartobacteria* (SFig. 28a), respectively. The affiliates of *Verrucomicrobiales* were the most abundance followed by the *Puniceicoccales* and *methylacidiphilales* species all representing <1% of the total communities (SFig. 29b). Notably, we found that all the *Opitutae* and *Spartobacteria* reads were unclassified at a lower level. Only the *Verrucomicrobia* communities of reactor 3 and 7 were found to cluster

partially on the the lower right quadrant of the plot, (at the class level, SFig. 28b). In contrast, several local communities including the ones identified in reactor 1 and 10, and reactor 3, 6, 7 and 11 were found to cluster partially (low β-divesities) within the plots (at order level; SFig. 29b). However, their communities in reactor 5 and 8 were found to cluster significantly on the upper right quadrant, (SFig. 29b), implying high similarities between the *Verrucomicrobia* species in the two treatments. Other treatments were found to reveal high β-diversities. Among the identified *Tenericute*'s, were *Acholeplasmatales*, *Mycoplasmatales* and *Entomoplasmatales*, all composing <1% of the total microbial communities and belonged to the class *Mollicutes* (SFig. 30a). Interesting, the member species were absolutely absent in reactor 5, contrary to other phyla species probably due to natural selection pressure. Nevertheless, we revealed significantly high similarities of the identified *Tenericute*'s species habiting the reactor 4 and 9 and reactor 6 and 8 (SFig. 30b). The local communities of the other treatments were significantly found to be dissimilar.

The identified *Planctomycetacia* affiliates were revealed to convey 1-3% of the total microbial communities and were all assigned to *Plantomycetacetales* (*Planctomycetacia*) order. Similarly, *Chlorobia* affiliates were found to compose 1-2% of the total communities and were all assigned to *Chlorobiales* order. *Spirochaetes* also conveyed <1-4% of the total microbial composition, and also were all affiliated to *Spirochaetales* order. *Thermotogae* were found to compose ≤1% of the total microbial composition, all classified as *Thermotogales*. Other less abundant communities were *Synergistales* (*Synergistates*, *Synergistia*), *Lentisphaerales* (*Lentisphaerae*, *Lentisphaerae*), *Fusobacterales* (*Fusobacteria*, *Fusobacteria*), and *Aquificales* (*Aquificae*, *Aquificae*), all representing <1% of the total communities. *Deferribacteriales*, *Chlamydiales*, *Chrysiogenales*, *Nitrospirales*, *Fibrobacteriales*, *Elusimicrobiales*, *Dictyoglomales*, *Gemmatimonadales* were another identified guilds and their abundances were much lesser compared to other communities. The *Candidatus poribacteria* affiliates, comprising <1% of the total communities were not accurately classified at the lower taxonomic ranks (Table 3).

The *Archaea* richness, abundances and β-diversities

Similarly we analyzed the identified *Archaea* domain and we found their species composed 6-12% of the total microbial populations. Out of the identified communities, we obtained 5 phyla (SFig. 31a), 9 classes (SFig. 32a) and 16 orders (SFig. 33a). Among the communities, those identified in reactor 4 and 12 were found to cluster on the lower left quadrant of the plot, significantly revealing high similarity. We also revealed low β-diversities of the identified *Archaeal* communities, of reactor 3 and 6 and those detected in reactor 7 and 9 (Phylum level; SFig. 31b). However, at a lower level, only those identified in reactor 3 and 7 were significantly found to express low β-diversity (Class and order; SFig. 32b and 33b). All other treatments were revealed to contain highly dissimilar communities at different taxonomic ranks.

Majority of *Archaea* species were classified as *Euryarchaeota* and composed 5-11% of the total microbial communities. The *Euryarchaeota* communities were classified into 8 classes (SFig. 34a), which were further categorized into 10 orders (SFig. 35a). Interestingly, we found significant dissimilarities (high β-diversity) of the *Euryarchaeotes* among the treatments at the two taxonomic levels (Class and order; SFig. 34b and 35b). The class *Methanomicrobia* was among the most abundant identified *Euryarchaeota* species and composed 4-9% of the identified total microbial populations. Further we revealed majority of the class as the affiliates of *Methanomicrobiales* order (2-4% of the identified total communities), followed by those assigned to *Methanosarcinales* and *Methanocellales* (SFig. 36a), both composing 1-3% and ≤1% of the identified total microbial communities. The local

Methanomicrobia species were highly dissimilar among the treatments albeit the ones identified in reactor 1 and 7 that were found to cluster. The communities of the two treatments revealed significant similarities (significant threshold, $P \leq 0.05$; SFig. 36b). Other identified methanogenic communities were assigned to *Methanobacterales*, the class *Methanobacteria*, *Methanococcales* belonging to *Methanococci* class and *Methanopyrales*, the affiliates of *Methanopyri* phylotype. All these species were revealed to compose <1% of the detected microbial abundances.

Other non-methanogenic *Euryarchaeotes* were the affiliates of *Thermococcales* (*Thermococci*), *Archaeglobales* (*Archaeglobus*), *Halobacterales* (*Halobacteria*), and *Thermoplasmatales* (*Thermoplasmata*) species. Among the phylotypes, *Thermococcales'* affiliates dominated the group while the *Thermoplasmatales* compsed the least relative abundances. *Crenarchaeota* and *Thaurmarchaeota* phyla were among the other identified non-methanogenic species. Between the two phyla species, *Crenarchaeotes* were the most diverse and were classified into 4 orders (SFig. 37a) compared to its counterpart. Among the identified *Crenarchaeota* orders were *Desulfurococcales*, *Thermoproteales*, *Sulfolobales* and *Acidilobales* (in order of abundances); all affiliated to *Thermoprotei* class, composing <1% of the identified microbial communities. Furhermore, we revealed half of the studied treatments (reactor 6 and 9, reactor 7 and 10 and reactor 2 and 11) had *Crenarchaeota* communities that conferred significantly low β-diversities, while the rest of the communities, in the other six treatments exhibited significantly high β-diversities (significant threshold, $P \leq 0.05$; SFig. 37b). *Thaurmarchaeota*, another non-methanogenic phylum was further classified into *Nitrosopumilales* and *Cenarchaeales* orders (SFig. 38a) and were revealed to exhibit high β-diversities (SFig. 38b). Other identified phyla were *Korarchaeota* and *Nanoarchaeota*, although their classifications at lower taxonomic ranks were inaccurate and their reads were treated as unclassified.

The fungi richness, abundances and β-diversities

Nevertheless, we also identified another hydrolytic guild, the fungal species. These fungal communities were classified into 5 phyla (SFig. 39a), 14 Class (SFig. 40a) and 23 orders (SFig. 41a). However, their relative abundances were low and composed <1% of the total microbial populations among the treatments. Significant high β-diversities of the identified local communities were the main observation at the phyla and class levels (SFig. 39b; SFig. 40b). However, at a lower taxa level, low β-diversities were the eminent feature particularly for the communities identified in reactor 2 and 10, reactor 3 and 6, reactor 4 and 9, and reactor 8 and 11 (SFig. 41b). Like in higher taxonomic levels, high β-diversities were revealed among the other studied treatments' communities.

The affiliates of *Ascomycota* were the most abundant guild (<1 of the total communities) within the *Eukaryotes* and were classified into 7 *Ascomycota* classes (SFig. 42a); that included *Eurotiomycetes*, *Saccharomycetes* and *Sordariomycetes* respectively. Other identified classes were *Schizosaccharomycetes*, *Dothideomycetes*, *Leotiomycetes* and *Pezizomycetes* respectively and were all revealed to comprise <1% of the total microbial communities in the respective treatments. We also further classified the identified *Ascomycota* species into 11 orders (SFig. 43a). Interestingly we found significant low β-diversity between reactor 7 and 8 fungal habitats. Other identified *Ascomycota* communities were found to reveal high β-diversities at a lower taxonomic level (Class and order level; SFig. 42b and 43b). Among the classes, *Sordariomycetes* were the richest within the phylum and their species were classified into 4 orders. The *Hypocreales*, *Sordariales*, *Magnaporthales* and *Phyllachorales*, respectively were among the identified *Sordariomycetes* orders (SFig. 44a). The *Sordariomycetes*' communities were found to

exhibit significantly high β-diversities (Statistical threshold, $P \leq 0.05$) (SFig.44b) among the studied treatments. *Eurotiomycetes*' another class of *Ascomycota* were classified into *Eurotiales*, and *Onygenales* orders (SFig.45a); despite their higher reads relative abundances compared to *Sordariomycetes* phylotypes. The two *Eurotiomycetes* orders were also shown to compose <1% of the identified total microbial populations. Like other fungal phylotypes, only the *Eurotiomycetes*' communities of reactor 1 and 3 were found to cluster on the plot, which indicated significantly similar community compositions. Other reactor communities were revealed significantly to comprised dissimilar *Eurotiomycetes* populations (significant threshold, $P \leq 0.05$; SFig. 45b) that expressed high *Eurotiomycetes* β-diversities. Other identified fungal species were the affiliates of *Saccharomycetales*, *Schizosaccharomycetales*, *Pleosporales*, *Helotiales* and *Pezizales* phylotypes, all composing <1% of the total microbial communities.

Among the identified fungi, *Basidiomycota* species were the second most abundance (<1% of the total communties); and were classified into 4 classes and 5 orders (SFig. 46a and 47a, respectively). Among the classes, *Agaricomycetes* dominated the treatments, followed by *Tremellomycetes*, *Ustilaginomycetes*, and *Exobasidiomycetes* species (SFig. 46a) respectively. The affiliates of the class *Agaricomycetes* were further classified as *Agaricules* and *Polyporales* species (SFig. 48a). Other *Basidiomycota* species were affiliated to *Tremellales* (*Tremellomycetes*), *Ustilaginales* (*Ustilaginomycetes*) and *Malasseziales* (*Exobasidiomycetes*) orders respectively. Similarly we also found two treatments' (reactor 11 and 12) containing *Agaricomycete*'s communities that revealed low β-diversities (SFig. 48b) while the other local *Basidiomycota* communities indicated significant high β-diversities (SFig. 46b and 47b).

The *Chytridiomycotes* were among the identified rare species, classified as *Chytridiomycetes* (more abundant) and *Monoblepharidomycetes* (SFig. 49). Among the *Chytridiomycetes* were *Spizellomycetales* (reactor 9 and 10), *Rhizophydiales* and *Cladochytriales*, the latter two dectected only in reactor 9 (SFig. 50). All the affiliates of *Monoblepharidomycetes* were classified as *Monoblepharidales* species and were also detected in reactor 9 only. Other identified rare phylotypes were *Blastocladiales*, (*Blastocladiomycetes*, *Blastocladiomycota*, detected in reactor 9), *Entomophthorales*, *Mortierellales*, and *Mucorales*, all derived from unclassified fungal reads, were also identified in reactor 9, reactor 10 and reactor 11 (SFig. 51) respectively. *Microsporidia* species were not accurately classisfied at a lower taxonomic levels and were regarded as unclassisfied though they composed <1% of the total microbial populations.

Sampling depth and β-diversity of the utilized shotgun metagenomic datasets

Sampling depth and α-diversity

Since environmental and technical variables can outweigh the biological variations when handling metagenomic datasets, prior to analysis, we asked ourselves, whether we had attained the maximum microbial richness and if our sampling was a representive of this maximum. To answer the questions, we conducted a rarefaction analysis through statistical resampling and plotted the total number of distinct species annotations agaist the total number of sequences sampled (Fig. 4). As a result, a steep slope, that leveled off towards an asymptote was obtained for each sample and later combined (Fig. 4). The findings indicated a reasonable number of individual species were sampled and more intensive sampling of the same treatments would yield probably few additional species, if not none. Further the α-diversity indices revealed more than 942 genera across the treatments, comfirming our maximum sampling effort was attained for further diversity studies.

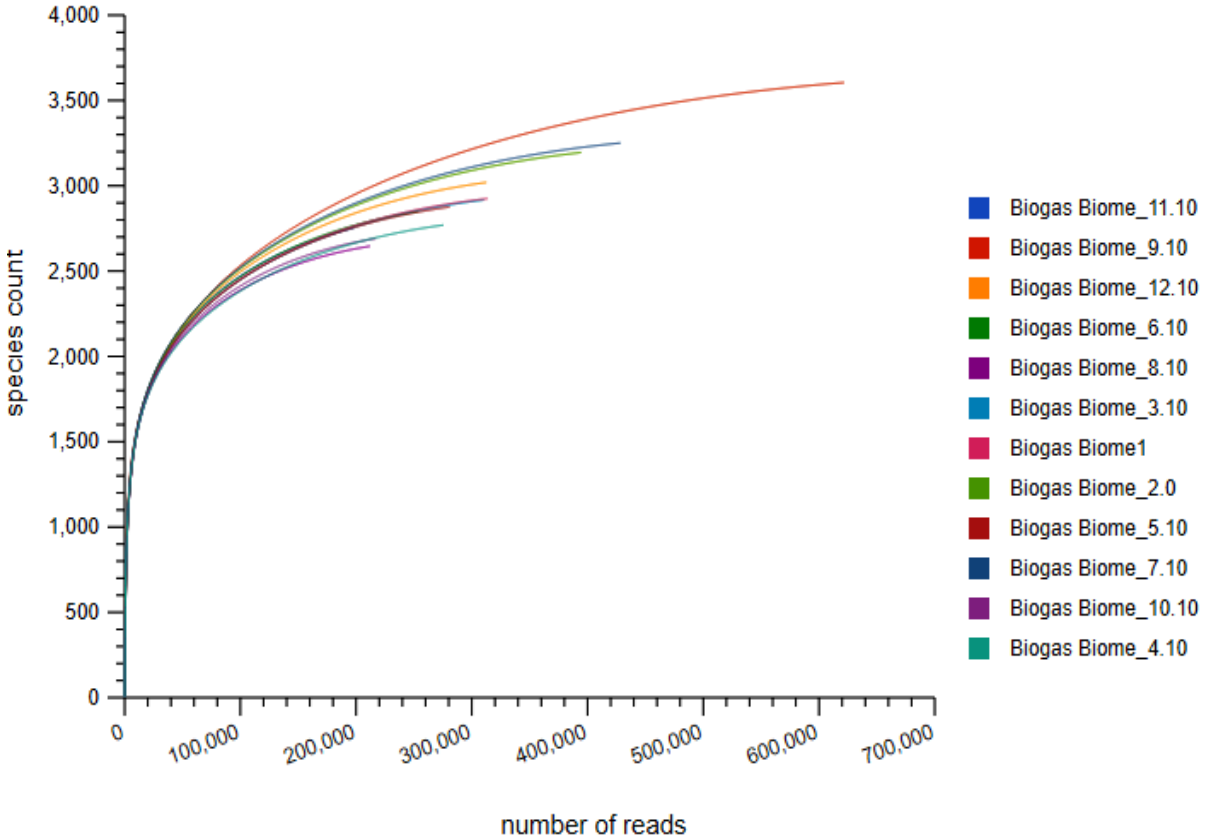

Fig. 4: Rarefaction plot showing a curve of annotated species richness. The curve is a plot of the total number of distinct species annotations as a function of the number of sequences sampled.

β-diversities of the identified communities

To understand the influence of environmental variations and other variables on the local microbial communities, we conducted a comparative community diversity studies on the twelve treatments. PCoA model was used due to its ability to reduce false discoveries; increased congruence between samples; and projected much distance between the sample distance matrices [43]. Prior to statistical analysis, we transformed our datasets to ensure our Euclidean model [distance (x, y) $=\{\sum_i(x_i-y_i)^2\}^{½}$] was not affected by different scales originating from our datasets. Furthermore, the PCoA plots were reconstructed at different taxa levels to minimize the complexity of the datasets. As a result, we revealed high β-diversity among the local communities in the majority of the treatments (at the domain level). The only observed exceptions were the local communities of reactor 4 and 11 that clustered on the lower left quadrant of the plot (SFig. 52a), revealing significant similarity between the organisms in the two treatments. Similarly, at a lower taxa level (phylum, class and order), only the communities of reactor 2 and 10 were found to be similar, and were located in either lower right or upper left quadrants of the plot (SFig. 52b and 52c). In contrast, we revealed low β-diversity between the communities of reactor 4 and 9 at lower taxa levels (phylum, class and order) and were found to cluster partially on the lower left and upper right quagrants (SFig. 52b and 52c) respecitvely. Other treatments that revealed low β-diversity between their communities were reactor 8 and 11 and reactor 3 and 6. The two categories of the communities were found to cluster partially in the lower right and lower left quadrants of the plot (Fig. 5).

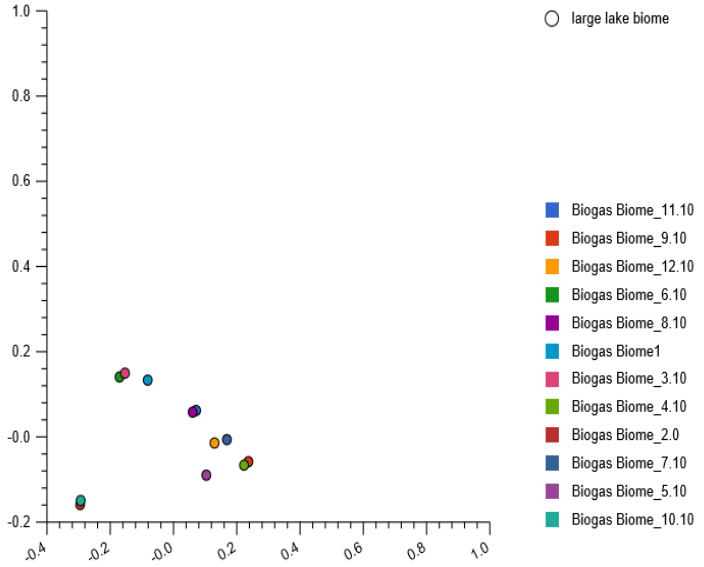

Fig. 5: The PCoA analysis revealing β-diversity of the twelve reactors at the order level: The reads compositions in reactor 2 and 10 were partially similar, located in the lower left quadrant, those in reactor 8 and 11 located in the upper right quadrant, reactor 4 and 9 in the lower right quadrant, while those in reactor 3 and 6 were located in the lower left quadrant of the plot.

Discussion

The statistics of metagenome annotation and taxonomic assignment overview:

The aim of our study was to use next generation sequencing technologies and myriads of bioinformatics tools to investigate the biogas' microbiome diversities on the environmental samples obtained from a wide geographic coverage. We combined marker genes and other genes on the assembled scaffolds to generate maximum taxonomic assignments. Among the sixteen utilized databases, SEED subsystems, yielded the highest reads hits annotations, the findings that corroborated with the previous reports [44]. Unlike other databases, SEED integrated several evidences that provided annotation benefits over the other 15 utilized databases. PATRIC database had less annotation compared to the IMG database, due to its high speciality of storing only pathosysstems information. The annotation of low protein sequences by the SWISS-PROT database was due to its manually curation and annotation systems. Contrasting the findings were the annotations of TrEMBL and Refseq databases that both utilized automated schemes, which superseded the annotation of SWISS-PROT database. However, in spite of its low resolution, majorly due to lack of regular updates, the outputs of SWISS-PROT are more accurate compared to the automated databases. Furthermore, orthologous groups databases such as COG, eggNOG etc were utilized to distigush orthologs from paralogs. In this case, we coupled gene classification schemes to the taxonomic assignements in order to increase detection resolution. Among these five databases (Fig. 1), eggNOG had the highest hits annotation due to its high phylogenetic resolution, automated annotation scheme, and covers more genes and genomes than its counterpart databases. In contrast, COG and NOG databases were among the databases that had the lowest hits annotations because of their manual curation and annotation schemes and fewer regular updates due to manual labor requirements. KEGG database was used to reconstructs metabolic pathways of the identified taxonomic assignements. Other utilized nucleotide databases were Genbank that ensured uniform and comprehensive annotation against the worldwide nucleotide sequences, Ribosomal Database Project (RDP) database that covered

small rRNA gene subunit (SSU) of *Archaea* and *Bacteria* domains, LSU and SSU databases that archived large and small rRNA subunit of *Archaea*, *Bacteria* and *Eukarya* sequences respectively. The latter two; provided a complete species detection advantage over the RDP database. The two databases archived nucleotide sequences together with their postulated secondary structures. The SSU database stored more community sequences than its counterpart; the LSU database and it used prokaryotic and eukaryotic models for secondary structure alignments. The extra feature of the SSU database explained why the database, had the highest annotation resolution compared to other amplicon sequences databases. However, its worth-noting that the three rRNA databases archived complete or near-complete rRNA sequences, although partial fragments of 70% of the estimated chain length of the combined molecule were also included. These characteristics of the amplicon sequence databases limited the annotation of our Miseq generated short reads. The ENVO ontologies analysis, which revealed high resolution of large lake biomes, suggested that majority of the cultured and sequenced species were derived from large water masses.

Our taxonomic annotations approach generated more detection, compared to the other annotations conducted elsewhere [45]. While other studies revealed 0.042% to 0.132% rRNA genes, we revealed 280 (0.00%) to 1,032 (0.00%) rRNA genes. These findings implied that the proportion of the rRNA genes to the total shotgun generated genes were negligible. The observations were highly attributed to our sampling and sequencing approaches as previously explained by Knight *et al*. [46]. Furthermore, while most of the studies used 19-35% of the total generated sequences, [45], our approach annotated and used 39.75-46.70% of our total generated sequences. All these findings indicate more than 53% of the biogas generating populations is still unknown. Briefly, our annotations were assigned to *Bacteria*, *Archaea*, *Eukaryotes* and *Viruses*, which corroborated the findings of other studies [47]. Among the domains, *Bacterial* species abundances dominated, conveying 87-93% of the total microbial populations, followed by *Archaea* and *Eukaryotes*, while *Viruses* comprised the least abundance population. The observed high *Bacterial* abundances, suggested their crucial metabolic roles in biomass conversion and other reactions within the systems [47]. As expected, we found low *Eukaryotes* (fungi) abundances which indicated the presence of low recalcitract biomasses concentration. Generally, the fungal species were involved in several biochemical reactions of biogas production including hydrolysis, fermentation and syntrophic reactions. The identified *Archaea* affililates were majorly consumers of smaller substrates that were generated by the *Bacterial* and fungal species. The *Archaea* species were able use different methanogenic routes to converts the substrates into $CH_{4(g)}$. Nevertheless, the main role of the identified low abundant group of the virus was unclear, though the species could be involved in microbial cell degradation in the AD systems. However, their detection was largely attributed to the viral pathogenicity on both *Bacteria* and *Archaea* species. High β-diversity among the local community (at the domain level) was an indication of environmental influences at different ecological location.

The Bacteria structure

More than 90 representative orders of *Bacteria* domain were revealed in our datasets. The first step of the AD process involved hydrolytic reactions that converted large macromolecules into smaller substrates [48]. Several *Bacterial* communities were capable of hydrolysis, even on high lignocellulosic plant biomasses. The identified species that participated in these reactions were annotated to *Firmicutes*, *Bacteroidetes*, *Chloroflexi* to mention but a few and their species harbored cellulosomes, that secrete extracellular enzymes, to degrade the substrate. Previous reports indicated the member species belonging to *Bacteroidales*, *Bifidobacteriales*, *Clostridiales Lactobacillales*, *Thermotogales* and many other identified *Bacterial* species catalyzed the first step of AD [48]. The step is a rate

limiting reactions, largely regulated by the substrate composition. The nature of the chemical composition of the substrate, generally depended on the environmental conditions (Table 1 and 2), and influenced microbial species' abundances and β-diversity differently. All these affected reactor stability and performance. As expected the class *Clostridia*, that comprised *Clostridiales*, *Thermoanaerobacteriale*s and *Halanaerobiale*s orders were the most dominant among the identified phylotypes. The guild converted large macromolecules into organic acids, alcohols, $CO_{2(g)}$ and $H_{2(g)}$ [49-55]. In this case the species used substrate level phosphorylation mechanism to conserve the energy. Acetogenic and homoacetogenic species belonging to the same guild such as *Moorella thermoacetica* and *Thermoanaerobacter kivui* were able to use reductive acetyl-CoA pathways to reduce $CO_{2(g)}$ into $CO_{(g)}$, concommitantly generating reducing equivalents [56, 57] and vice versa. As a consequence, the acetogens and homoacetogens used proton or sodium dependent mechanism to conserve their energy [57]. Above all, the *Clostridia* guild were able to catalyze other biochemical reactions such as reductive dechlorination of various compounds that led to production of sulfide from either thiosulfate or sulfite. The reactions were the consumers of pyruvate and lactate as the carbon source. In this case the member species were able to use [NiFe]-hydrogenases to generate $H_{2(g)}$ [55]. We also identified *Halanaerobiales* species that tolerated high salt concentrations [58], in our systems. Generally the treatments comprised high salts concentrations and the *Halanaerobiales* regulated the toxic cations from the ionization of the salts. All these *Clostridia*'s roles suggested their metabolic versatility, and the features were used to explain the observed low richness of *Firmicutes* compared to *Proteobacteria* species. Similarly, the observed high β-diversities among the treatments were linked to the versatile *Clostridia* roles. However, the observed low β-diversities (SFig.13b) of reactor 7, 11 and 12 communities were attributed to similar treatment perturbations and environmental conditions (Table 1 and 2).

The *Bacillales* and *Lactobacillales* species, all belonged to the class *Bacilli* and were able to use mixed fermentation routes to metabolize carbohydrates and other biological macromolecules [59, 60]. Formate dehydrogenase complex activities were the main mechanisms of energy conservation [60]. Generally, *Bacilli* co-existed with *Clostridia* species in lactate and acetate metabolism [61], a mechanism that was termed as lactate cross-feeding and indicated their symbiotic relationship in the AD systems. High β-diversities among the observed treatments clusters (SFig. 14b) were solely attributed to substrate composition and environmental conditions (Table 1 and 2). However, the Low β-diversities observed between or among the treatments (SFig.14b) that formed the clusters were largely attributed to reactor operation similarities. All these, implied that the organisms were in the same succession stage at the time of sampling. *Erysipelotrichi* another identified *Firmicutes* class, utilied syntrophic acetogenesis route to metabolize lipids into acetate, $H_{2(g)}$ and $CO_{2(g)}$ [45]. Its species, together with other *Firmicutes* species like *Smithella propionica* used non-randomizing pathway [62] to metabolize propionate into acetate. Among other identified *Firmicutes*, were *Syntrophomonas wolfei* (*Clostridiales*, *Clostridia*) that used β-oxidation pathway to oxidize C_4 to C_7 fatty acids in order to produce either acetate and $H_{2(g)}$ or propionate, acetate and $H_{2(g)}$ [63, 64]. Notably, the syntrophic species were consumers of 30% of the generated electrons, which bypassed the rate limiting steps of volatile fatty acids accumulation. All these, contributed to reactor stabilities. In general, we associated the observed low *Firmicutes* abundances and diversities (richness, SFig.11a and 12a) with genetic versitilities that encoded assortment of proteins; which futher metabolized a variety of biomasses into biogas. The observed high β-diversities (SFig.11b and 12b), among their local communities confirmed our suggestion and we further felt that the reactor sizes and

environmental conditions (Table 2), were the key contributors to their genetic evolution (Table 1 and 2).

In contrast, the identified *Synergistete*'s species were found to ferment amino acids into lactate, acetate, $H_{2(g)}$ and $CO_{2(g)}$ [65, 66] and expressed no ability to metabolize carbohydrates and fatty acids. One such identified example was *Dethiosulfovibrio peptidovorans*. The affiliates of *Anaerobaculum* genus belonged to *Synergistete*'s, and were revealed to catabolize peptides into acetate and short-chain fatty acids while their close relatives, *Aminobacterium*, were capable of fermenting amino acids into acetate, propionate and $H_{2(g)}$ [65, 66]. Other identified amino acid metabolizers were the affiliates of *Fusobacteria* [67], the species that catabolized the substrates into acetate, propionate, and butyrates, concomitantly releasing $H_{2(g)}$, $CO_{2(g)}$ and $NH_{3(g)}$. The member species of the two phyla used substrate-level phosphorylation to conserve their energy [68] from amino acids metabolism. The findings contradict the metabolism of the amino acid by the *Clostridia* species, where the energy conservation mechanisms were unclear and were postulated to occur through chemiosmosis [69]. However, the two phyla were found to generate the energy through syntrophic association with methanogens; otherwise the generated protons increased the systems' alkalinity that led to inhibited steady-state of the biochemical reactions. Based on their low abundances of the two phyla and richness among the treatments, we postulated low amino acid affinities among their identified species, which suggested another selection pressure in the AD systems.

The affiliates of *Bacteroidales*, *Flavobacterales*, *Cytophagales* and *Sphigobacteriales*, all belonging to *Bacteroidetes*, majorly fermented carbohydrates and proteins into mixed products (acetate, lactate, ethanol, and volatile organic acids), and concomitantly produced $H_{2(g)}$ [70]. Moreover, their species were also capable to metabolize aliphatic, aromatic and chloronated compounds [71] and were co-existing with methanogens to increase energy extraction from indigestible plant materials. The member species were highly varied among the treatments and we suggested that the cattle breeds and the operating conditions were the key drivers of the observed β-diversities. Interestingly, we found *Cytophagale*'s relative abundances were consistent among several treatment (2% of total reads), a trait that was attributed to their long periods of endurances in conditions of low nutrient limitation [72]. Although, the metabolic activity of the *Chlorobi* overlapped with that of *Bacteroidetes* [73], their species used reverse tri-carboxylic acid (rTCA) cycle to fix $CO_{2(g)}$ and $N_{2(g)}$ in anaerobic conditions. These reactions were utilized by the *Chlorobi* to oxidize reduced sulfur compounds [74, 75]. However, *Bacteroidetes* species had revealed higher selection advantage over *Chlorobi* due to the observed low *Chlorobi* abudunces and richness (i.e only single order, the *Chlorobale*).

The identified *Proteobacteria* guilds consisted metabolically versatile species and were revealed to comprise the highest number of species (richness) compared to other identified phyla in the treatments. The guilds were able to metabolize proteins [76] carbohydrates and lipids [77] and we associated their low β-diversity (SFig. 4b and 5b) with low substrate composition heterogeneity and differences in environmental conditions (Table 1 and 2). The *γ-Proteobacteria* was revealed as the most diverse class of the *Proteobacteria* and the class comprised 14 orders (SFig. 7). The *Enterobacteriales* such as *Escherichia coli*, were the facultative anaerobes and fermented carbohydrates into lactate, succinate, ethanol, acetate, $H_{2(g)}$ and $CO_{2(g)}$ and $H_{2(g)}$ [78]. Unlike other syntrophic species, the affiliates of *Enterobacteriales* were found to use methyl-citrate cycle [79] to metabolize propionate into pyruvate, a precursor for butanol, isopropanol and other mixed acid fermentation products. *Burkholderiales*, another *γ-Proteobacteria* phylotype, produced succinate from carbohydrates

concomitantly oxidizing $H_{2(g)}$, S^0 and Fe^0 into their respective oxidized forms to conserve energy [80, 81]. We revealed that environmental temperature (Table 2) was the main driver of the observed abundant and β-diversities of the *γ-Proteobacteria* communities (SFig. 7, Table 2).

In contrast, the affiliates of *Sutterella* genus, among other identified species of *β-Proteobacteria* were saccharolytic and nitrate reducers [81]. The species also consumed propionate, butyrate and acetate syntrophically [81, 77]. The agroecological zones and environmental temperatures (Table 1 and 2) were the main factors that impacted on the *β-Proteobacteria* diversities and abundances among the treatments (SFig. 8). The *Caulobacterales*, an order of *α-Proteobacteria* were among the identified low abundant species and were previously revealed to survive in low nutrient environments [82]. *Rhodospirillales* (*α-Proteobacteria*), such as *Rhodospirillum rubrum*, were found to participate in homoacetogenesis and converted $CO_{(g)}$ and formate into acetyl-CoA, $CO_{2(g)}$ and $H_{2(g)}$ [83]. The produced acetyl-CoA was further converted into acetate, a precursor for methanogenesis. Like members of *β-Proteobacteria*, the *α-Proteobacteria* species were also affected by the agroeclogical zones (Table 1 and 2) and we postulated that similar substrate composition were the main cause of the observed β-diversities and abundant variations (SFig. 9) aming the treatments. The *Campylobacterales* and *Nautiliales* species were the affiliates of *ε-Proteobacteria* and were previously found to metabolize proteins in the AD systems; however, they were also revealed to express an additional role of antibiotic resisistance in biofilms and granules [77].

Among the identified *δ-Proteobacteria* species, *Pseudomonadales*, the gemmules producing anaerobes [84] were the most abundant and were found to co-exist with the algae [84]. The reported syntrophic traits of the *Pseudomonadales* order were revealed to enable the species to withstand unfavourable environmental conditions, which offered the species the selection pressure advantage over other *δ-Proteobacteria* species. Generally, the *δ-Proteobacteria* species were syntrophic oxidizers [85-87] and converted volatile fatty acids into acetate, formate, $CO_{2(g)}$ and $H_{2(g)}$. The identified *Syntrophobacteriales* species e.g S*yntrophus aciditrophus* were able to use randomized methyl-malonyl-CoA pathway to oxidize butyrate, propionate and long-chain fatty acids into acetate, formate, $CO_{2(g)}$ and $H_{2(g)}$ syntropically [85, 86] while *Syntrophobacter wolinii* [87] another member species, used β-oxidation pathway to metabolize propionate into acetate. Other species belonging to *Syntrophaceticus* genus were reported to oxidize the produced acetate into $CO_{2(g)}$ and $H_{2(g)}$ [88]. Interestingly, for these species to effeciently oxidize the respective macromolecules, they were revealed to use an electron transport mechanism that created a positive redox potential by reversing electrons that were generated in syntrophic reactions. In contrast, the identified affiliates of *Desulfarculale*'s, *Desulfobacteriale*'s and *Desulfovibrionale*'s phylotypes were sulfate reducers, although in absence of sulfate (electron acceptor), the species were able to oxidize the organic compounds, (e.g lactate, acetate and ethanol), into $CO_{2(g)}$ and $H_{2(g)}$ syntrophically [89]. The *Desulfuromonadale*'s species were also revealed to use Fe^{3+} compounds to oxidize ethanol into $CO_{2(g)}$ and $H_{2(g)}$, but in strict co-operation with methanogens [77]. Lastly, *Myxococcales* and *Chlamydiae* species were also revealed as fermenters of carbohydrates, [90] and proteins [91], which were converted into organic acids. All these *δ-Proteobacterial* roles were contributing to the systems' stability by providing electrons, which acted as reducing equivalents in the AD treatment. We also postulated that substrate compositions were the main drivers of the observed *δ-Proteobacteria* abundances and β-diversities (SFig. 6b). The other low abundance species were assigned to *Mariprofundale*'s order, the affiliates of *Z-Proteobacteria* class. The affiliates co-existed with *Nitrosopumilales* to oxidize Fe^0 [92]

into Fe^{2+} or Fe^{3+} ions. Moreover, the two phylotype species were previously revealed to converts nitrogen substrates into $NO_3^-{}_{(aq)}$ in their syntrophy [93, 94]. In this study, through their syntrophy, we suggested completion of redox reactions in the treatments through electrons release. In summary, all affiliates of *Proteobacteria* were capable of reversing oxidative damage to methionine [77], a trait that explained why their species were more (richness) compared to other phyla species. Moreover, *Proteobacteria* species were able to express shorter diversification time-scales over other microbes that led to large population sizes, high growth rates, and strong selective regimes of the species. All these facilitated rapid *Proteobacteria* adaptation through mutation or recombination events.

We also identified *Solibacterales* and *Acidobacterales* species that were affiliated to *Acidobacteria* phyla. The member species fermented polysaccharides and monosacharides into acetate and propionate [95]. The species were also capable of fixing $CO_{2(g)}$ anaplerotically [96], ultimately replenishing carbon atoms in the systems. Furthermore, the affiliates were capable of catabolizing inorganic and organic nitrogen substrates as their N-sources [97]. The detection of these species were highly attributed to high amount of $NH_4^+{}_{(g)}$ in our AD systems. We postulated that reactor perturbations rather than environmental conditions were the main influencers of their genetic composition (Table 2 and SFig. 25b and 26b).

Generally, the detected *Mollicute*'s species (*Tenericutes* phylum) were facultative anaerobes and were all absent in reactor 5, probably due to the nature of the substrates fed to the cattle breeds. The species were fermenters of sugars, and they converted the substrates into organic acids such as lactic acid, saturated fatty acids and acetate [98, 99]. All the identified species were the affiliates of *Acholeplasmatales*, *Mycoplasmatales* and *Entomoplasmatales* and were previously shown to exhibit reduced respiratory systems, with incomplete TCA cycle, that lacked quinones and cytochromes. The metabolisms of Tenericutes were previously revealed to yield low amount of ATP concomitantly producing large quantities of acidic metabolic end products [98]. The produced acidic metabolites affected negatively the performance of other microbes in the systems; a phenotype that led to poor reactor performance. We suggested their occurrence to be driven by the agroecological zones and soil type, which directly affected the nature of substrate (Table 1 and 2 and SFig. 30).

Like other *Bacterial* species, the affiliates of *Actinobacteria* (SFig.17a), were able to metabolize a wide variety of substrates [100, 101, 102] to produce acetic acid, lactic acid, $H_{2(g)}$ and other products. Among the *Actinobacteria*, *Coriobacteriales* metabolized formate and $H_{2(g)}$, concomitantly reducing $NO_3^-{}_{(aq)}$ into $NH_{3(g)}$ [100] and *Slackia Heliotrinireducens* was one of its type species. The *Bifidobacteriale* was another order that metabolized oligosaccharides and releases lactic acid [101]. Other species belonging to *Actinomycetales* were previously revealed to metabolize syntrophically propionate into acetate [102]. The *Actinobacteria*'s role of $NH_{4(aq)}$ metabolism, were shown to protect methanogenic *Archaea* from other oxidizing agents. The previous reports revealed the species were more sensitive to nutrients perturbations [102] and due to this, reactor perturbations particularly the feeding regimes (SFig.17b) contributed to the observed species abundances and diversities among the treatments. The low *Plactomycete*'s abundances corroborated with the findings obtained elsewhere [103]. The member species were found to metabolize sulphate-containing polysaccharides [103], recalcitrant hydrocarbons and other organic compounds [104]. However, when in syntrophic association with *Chloroflexi* (SFig.18a, 19a, 20a and 21a), the species were able to oxidize butyrate [105] into smaller substrates.

Similarly the low abundances of *Thermotogae*'s communities were also revealed to corroborate with the findings of previous studies [106]. The carbohydrate metabolism roles of the guild were previously postulated based on the gene content of *Defluviitoga tunisiensis* [107]. The *Thermotogae*s were reported to convert macromolecules into acetate, ethanol, $CO_{2(g)}$, and $H_{2(g)}$. In most cases, the species utilized the Embden-Meyerhof parnas (EMP) pathway to generate pyruvate, which was either converted into acetyl-CoA, a precursor for acetate and ATP formation or reduced to ethanol, concommitantly oxidizing electron carriers. However, the $H_{2(g)}$ produced in these reactions were coupled to the oxidation of NADH and reduction of ferredoxin [107]. These reactions were found to favour proton production which reduced the P_{H2} in the AD systems. However, for these reactions to proceed favourably, *Thermotogale*'s were revealed to depend on *Clostridiales* species [108] which implied that *Thermotogale*'s abundances and richness were regulated by the *Clostridiales* species. Like other low abundant phylotypes, *Spirochaete*'s abundances were found to corroborate with the findings of Wang and his co-wokers [109]. The identified *Spirochaete*'s were homoacetogens and acetogens and were previously revealed to convert $H_{2(g)}$, $CO_{2(g)}$ and $CO_{(g)}$ into acetate [109, 110]. Its member species were also revealed to co-exist with methanogens to provide an alternative route, termed as the butyrate oxidation pathway for $H_{2(g)}$ consumption [110]. Similarly, the *Spirochaetes* species were also found to co-exist with the *Thermotogale*'s species to ferments tetrasaccharides and cellobiose into organic acids [153]. These syntrophies protected *Clostridiale*'s species from substrate inhibition mechanisms [108]. Based on these findings and those reported elsewhere, we suggested co-operation of numerous microorganisms in the AD systems that used assortment of mechanisms to metabolize the substrate. Due to this, we further postulated the holobiontic mechanism that regulated symbionts and other syntrophic species in the treatment through ecological selection, via co-evolution to minimize conflict between or among the symbiotic species.

The identified *Verrucomicrobia* communities were affiliated to *Methylacidiphiles*, *Puniceicoccales* and *Verrucomicrobiale* (SFig. 28a and 29a). The communities metabolized cellulose and other C_1-contaning compounds [111], while the affiliates of *Cyanobacteria* were classified as *Chroococcales*, *Oscillatoriales* and *Nostocales* species (SFig. 23a and 24a) and were mainly involved in $H_{2(g)}$ and $N_{2(g)}$ metabolism [112-114] in addition to carbohydrate metabolism. The conversion of N-containing substrate into $NH_{3(g)}$ was an endergonic reaction and required ATP. The reaction were also found to oxidize $H_{2(g)}$ into obligate $H^+_{(aq)}$ ions, circumventing high P_{H2} in the AD systems. Therefore, the recycling of $H_{2(g)}$ through oxyhydrogen route, also termed as Knallgas reaction by the *Cyanobacteria* species were revealed to provide ATP and protected strict anaerobes and other enzymes from oxidizing agents, by providing reducing equivalents and removing toxic $O_{2(g)}$ species [115, 116]. In our case, we attributed their abundance variation with $H_{2(g)}$ concentration variability among the treatments.

Another identified phylotype was the *Candidatus poribacterial* species, which fermented carbohydrate and proteins, including urea [117] into organic acids. The species were revealed to use Wood-Ljungdahl pathway to fix $CO_{2(g)}$ generated from carbohydrates and other macromolecules. However, the species were previously revealed to utilize oxidative deamination pathway [118] to ferment the amino acids, while randomized methyl-malonyl-CoA pathway was the possible route for propionate oxidation. All these reactions were found to produce $H_{2(g)}$ and we postulated commensal symbiontism between *C. poribacteria* and acetogenic species. Like other *Bacteria* species *Fibrobacter* members such as *Fibrobacter succinogenes* were involved in the metabolism of cellulose [119] and other macromolecules in AD systems.

Although it's not a common tradition to explain the existence of metal-reducing organisms in AD systems, is worthy noting that electron sink ability of any systems depends on these organisms. One of such phylotypes identified were the affiliates of *Gloeobacterales* (SFig. 22a and 23a). The member species, were previously reported to transfer electrons directly from methanogens to syntrophic *Bacteria* [120]. These kinds of reactions were viewd to enhance clean biogas production through inhibiting sulfate reducing bacteria. Other identified metal reducing species were the members of *Clostridiales* that were previously revealed to utilize Fe^{3+} Co^{3+} and Cr^{6+} as electron acceptor [120]. Other identified *Clostridiales*' species were also reported to use acrylate pathway to reduce arsenate and thiosulfate concomintantly fermenting organic molecules [121]. The affiliate of *Nitrospirales* (*Nitrospira* phylum), were found to reduce sulfate using $H_{2(g)}$ and the reaction were revealed to promote acetate oxidation [122, 123]. All these metal reducing reactions were contributing to the reactor stability through providing reducing equivalents in the systems.

The *Archaea* structure

The generated volatile organic acids, $CO_{2(g)}$ and $H_{2(g)}$ by acidogens, acetogens, and homoacetogens are further metabolized by *Archaea*, particularly the species that generate $CH_{4(g)}$ [124]. The domain constituted six (6) methanogenic and ten (10) non-methanogenic guilds, although the two groups were equally important in the AD systems. The methanogenic species were consumers of *Bacterial* products and converted them into $CH_{4(g)}$, the biogas. These species were the affiliates of *Euryarchaeota*, and they utilized hydrogenotrophic, acetoclastic, methylotrophic and methyl-reducing pathways to metabolize the substrates into $CH_{4(g)}$. Generally, the species of *Eurychaeota* indirectly stabilized the AD system by consuming $H_{2(g)}$ [125] and we suggested that the varied $H_{2(g)}$ concentration among the treatments, largely contributed to the observed *Euryarchaeota*'s abundance and β-diversity differences among the local communities. However, environmental (Table 1 and 2) and reactor perturbations have have also influenced the community abundant. Notably, the identified *Eurychaeota* species were also categorized into methanogenic and non-methanogenic guilds. Among the methanogens, *Methanosarcinales*, the affiliates of *Methanomicrobia* class, were the second most abundant, conveying 1-3% of the total communities. The affiliates of *Methanosarcinales* used three metabolic routes singly or in combination to convert methanol and methylamines, acetate or $CO_{2(g)}/H_{2(g)}$ and other potential substrates into $CH_{4(g)}$. The produced $H_{2(g)}$ from the reactions were consumed by the same species to maintain P_{H2} [126]. However, some species within the guild, lower the accumulated P_{H2} through reassimilating acetate, the process that transfer $H_{2(g)}$ to acetogens. In certain conditions, the species of *Methanosarcinales* (e.g *Methanosarcina acetivorans C2A* and *Methanosarcina barkeri*) and *Methanobacteriales* (*Methanothermobacter thermoautotrophicus*) expressed metabolic features, similar to acetogenic organisms. Previously, the species were revealed to use acetyl-CoA pathway to metabolize $CO_{(g)}$, into acetate and formate, under different $P_{CO(g)}$ [127]. These reactions were found to reveal the alternative route for ATP generation from $CO_{2(g)}$ reduction by the generated $H_{2(g)}$ in these species. Normally, the *Methanosarcinales* couple the methanogenesis to ion transport, also termed as chemiosmotic coupling or chemiosmosis, to establish an electrochemical gradient across the membrane [127] because the reaction ($CO_{2(g)}$ reduction route) is normally unfavourable. Generally the M*ethanosarcinales* express metabolic plasticity, although they were the second most dominance among the *Archaea* species.

Among the *Archaea*, *Methanomicrobiale*'s, another affiliates of *Methanomicrobia* class (SFig. 36a) were the most abundant within the class, conveying 2-4% of the identified total microbial populations. The affiliates used two methanogenesis routes:- hydrogenotrophic and methylotrophic routes, to convert formate, $H_{2(g)}/CO_{2(g)}$, 2-propanol/$CO2_{(g)}$ and secondary

alcohols [128] into $CH_{4(g)}$. The *Methanomicrobiales* consumed isopropanol to yield acetone and the reducing equivalent, NADPH [129]. However, the species were lacking $CO_{2(g)}$ reductive pathway, uncoupling methanogenesis from $CO_{2(g)}$ fixation, which implied the species used electron birfucation to conserve the energy. Due to these reactions, the *Methanomicrobiales* species were able to express high acetate requirement phenotype for carbon assimilation [130]. In costrast, the identified *Methanocellale*'s species, the affiliates of the same class, used solely acetoclastic route to convert $CO_{2(g)}$ into $CH_{4(g)}$ [131]. Majority of its species were ables to reduce $CO_{2(g)}$ with $H_{2(g)}$, except in rare cases where formate and secondary alcohols were used as an alternative electron donors. Similarly, all the identified *Methanobacteriales*' species, the affiliates of *Methanobacteria* class used $CO_{2(g)}$ reductive route to converts $CO_{2(g)}$ into $CH_{4(g)}$ with formate and $H_{2(g)}$ acting as electron donor. We were prompted to postulate that formate metabolism particularly in acetoclastic methanogenesis was triggered by the low P_{H2} and through electron bifurcation the species were able to conserve their energy. Other detected methanogenic species were the affiliates of *Methanopyrales*, belonging to *Methanopyri* class and *Methanococcales* species that were assigned to *Methanococc*i class (SFig. 34a and 35a). The members of the two classes were previously reported to metabolize $H_{2(g)}$ and $CO_{2(g)}$ into $CH_{4(g)}$ under thermophilic and saline conditions, [132, 133], indicating that our treatment had high amount of salts and the temperatures were high enough to support the species.

In contrast, in addition to other macromolecular metabolism, non-methanogenic group of *Archaea* reduced metal elements and their compounds [120, 121], to provide reducing equivalents in the AD systems. The affiliates of *Halobacteriales* (*Halobacteria*, *Euryarchaeota)* were among the detected non-methanogenic species. The member species used arginine deiminase pathway [134, 135] or other metabolic routes of amino acids to catabolize the amino acids substrates. The energy conservation in *Halobacteriales* occured when the intermediate metabolites; ornithine and carbamoylphosphate, were further converted into $CO_{2(g)}$ and $NH_{3(g)}$. The varied relative abundances among the treatments were attributed to the concentration of the alternative electron acceptors and salts concentrations. However, other deterministic factors such as competition and niche differentiation could have lead to the observed *Halobacteriales* abundant variations among the treatments.

Thermoplasmatales, the affiliates of *Thermoplasmata* class were another identified guild, which utilized sulfur for energy consevation [136]. However, their close relatives, *Methanomassiliicoccales*, formerly *Methanoplasmatales* were found to use methyl-reducing pathway to convert methanol into $CH_{4(g)}$. These relatives conserved energy through electron bifurcation [137] and we postulated the affiliates of *Thermoplasmatales* wer also involved in $CH_{4(g)}$ production only that few or none of the biochemical assays have been previously conducted to test their biogas roduction abilities. We also identified *Thermococcales* species, the class *Thermococci*. Previously, their affiliates were shown to reveal short generation time and the species withstood nutrient stress for a longer duration [138]. Like *Bacterial* species, and in the absence of sulfur, *Thermococcales*, were involved in both acidogenesis and acetogenesis reactions. Its' species metabolized proteins and carbohydrates into organic acids, $CO_{2(g)}$, and $H_{2(g)}$ [139, 140]. Other *Thermococcales*' affiliates were involved in homoacetogenic reactions where they oxidized $CO_{(g)}$ into $CO_{2(g)}$, and catabolized pyruvate into $H_{2(g)}$ and other products [141, 142]. Similarly, *Archaeoglobales*, the affiliates of *Archaeoglobus* class, used sulfate, as the reducing agent, to oxidize lactate, long fatty acids and acetate into $CO_{2(g)}$ [143], although some species e.g *Geoglobus ahangari* reduced Fe^{3+} compound with $H_{2(g)}$ [144]; if we emphasis on the role of microbes in metal reduction and reactor stability.

Sulfolobales and *Desulfurococcales*, all belongs to *Thermoprotei* and *Crenarchaeota* phylotypes, (SFig. 37a). The member species metabolized carbohydrates and amino acids [145, 146] into pyruvate and $CO_{2(g)}$. The organisms further fermented pyruvate into lactate and $CO_{2(g)}$. The generated $CO_{2(g)}$ either entered into the $CO_{2(g)}$ reductive pathway, the route that was catalyzed by acetogens/homoacetogens to produce $CH_{4(g)}$. or the precursor was fixed through 3-hydroxypropionate/4-hydroxybutyrate or dicarboxylate/4-hydroxybutyrate cycles by the identified *Thermoprotei* species to generate the energy [147]. Among other *Thermoprotei* affiliates were *Thermoproteales* that used sulfur and its compounds as an electron acceptors, to ferment carbohydrates, organic acids and alcohols [148]. In these reactions, the member species consumed $CO_{2(g)}$ and $H_{2(g)}$. At a lesser growth rates the species fermented and oxidized amino acids and propionate respectively [149]. Other affiliates like *T. neutrophilus* assimilated acetate and succinate in the presence of $CO_{2(g)}$ and $H_{2(g)}$ [150]. Again in this case, the species used $H_{2(g)}$ to reduce the sulfur. Unlike most of the other species, *Thermoproteales* used oxidative and reductive TCA routes to conserve energy. *Acidilobales* another *Thermoprotei* order used S^0 and thiosulfate to metabolize protein and carbohydrate substrates [151]. The species used either EMP or the Entner-Doudoroff (E-D) [152] pathways to convert the substrates into acetate and $CO_{2(g)}$ [153]. The *Acidilobales* affiliates also were able to oxidize triacylglycerides and long-chain fatty acids [153] into acetate and $CO_{2(g)}$. Substrate-level and oxidative phosphorylation were the major routes of energy conservation in *Acidilobales*; and the two mechanisms occured when S^0 and acetyl-CoA concentrations were high. Due to S^0 respiration in *Acidilobales* species, we postulated that their member species closed anaerobic carbon cycle via complete mineralization of organic substrates and the variations of these substrates together with other determinsistics and non-determistics (Table 1 and 2) factors contributed to the observed abundant variation among the treatments and high β-diversity (SFig. 37b). We also hypothesized that soil type was another factor that exerted natural selection on the *Acidilobales* through altering the substrates' chemical composition.

The *Thaurmarchaeota* communities were classified into *Cenarchaeles* [154] and *Nitrosopumilales* [155] phylotypes (SFig. 38). The member species used modified 3-hydroxypropionate/4-hydroxybutyrate pathway to fix inorganic carbon to generate energy. The communities also were able to metabolize urea and its products into, $NH_{3(g)}$ [156, 157]. The *Nanoarchaeota* (SFig. 31a) species exploited *Crenarchaeota* mechanisms [158, 159] to metabolize amino acids and complex sugars [160, 161] while the *Korarchaeotes* were sole amino acid fermenters. The energy metabolism in the two phyla is still unclear and further biochemical studies are needed to unravel their full metabolic activities.

The fungi structure

Like *Bacterial* species, the identified fungi (SFig. 39a, 40a and 41a) converted macromolecular substrates into acetate, $CO_{2(g)}$ and $H_{2(g)}$ [162]. Among the fungal species, *Ascomycota* were the richest and most abundant guild, that conveyed <1% of the identified communites in the treatments. Among the *Ascomycota* species, majority were the affiliates of *Saccharomycetales* (*Saccharomycetes*, *Ascomycota*), that used EMP pathway to ferment glucose and other sugars into pyruvate. One of such identified example was *Saccharomyces cerevisiae* [163] that fermented carbohydrates into mixed fermentation products. Other *Saccharomycetes* species e.g *Hansenula polymorpha* yeast fermented monosaccharides, fatty acids and amino acids into acetate, ethanol, $CO_{2(g)}$ and $H_{2(g)}$ [164]. Furthermore, the species were also able to oxidize methanol into formaldehyde and $H_2O_{2(g)}$ [165]. The products acted as the sources for formate and $H_{2(g)}$. The observed abundant variations of the *Saccharomycetes* species were highly attributed to their metabolic plasticity, deterministics factors and other environmental conditions. Other *Ascomycota* affiliates belonging to

Pezazales (*Pezizomycetes*), *Pleosporales* (*Dothideomycetes*), and *Eurotilaes* (*Eurotiomycetes*) also were revealed previously to use mixed fermentation pathways to ferments carbohydrates [166, 167]. However, the metabolisms of these *Ascomycota* species were observed to vary within and among the phylotypes, despite the presence of a common enzyme for the pyruvate metabolism. One of such identified example was among the *Eurotiale*'s order; where the *Paecilomyces variotii* express low metabolic ability on plant-derived compounds compared to other *Eurotiale*'s members [168, 169]. All these explained genetic versatility among the *Eurotiomycetes*' species, which impacted on the abundance variations among the treatments. We also added that the communities of *Eurotiomycetes* were also influenced by the environmental temperature and soil type (Table 1 and 2 and SFig. 45a). Like *Fusobacteria* species and other amino acid metabolizers, *Dothideomycetes* were able to metabolize lipids and proteins into volatile organic acids [169] while the affiliates of *Pezizomycetes* used sulfate, as an electron sink, to cosume amino acids [170, 171]. Similarly, we attributed their community variation to the substrates composition and natural selection pressure differences among the treatments.

The *Sordariomycetes*' affiliates (SFig. 44a) produced enzymes that expressed assortment of degradative mechanisms and other fermetative roles. The *Hypocreales*' affiliates were such an example that produced powerful hydrolytic enzymes [172] that catabolized complex biomass in their vicinity. *Fusarium* sp. the affiliates of *Hypocreales*, metabolized $NH_{3(g)}$, lipids and proteins [173, 174] into the respective smaller molecules. The species were found to use reducing equivalent, the NAD(P)H to reduce $NO_3^-{}_{(aq)}$ into $NH_4^+{}_{(aq)}$. The production of $NH_4^+{}_{(aq)}$ by the *Hypocreales* were coupled to the substrate-level phosphorylation and ethanol oxidation; the processes that produced acetate [173, 174]. The two mechanisms ameliorated the $NH_4^+{}_{(aq)}$ toxicity to micro-organisms. Other *Hypocreales* species such as *Hirsutella thompsonii* conferred a wide metabolic activity that involved chitinases, lipases, proteases and carbohydrates degrading enzymes [175]. Monilia sp. [176] and *Neurospora crassa* [177] were among the identified *Sordariomycetes* and they converted cellulose into ethanol. *Magnaporthales*' such as *Pyricularia penniseti* were the converters of isoamylalcohol into isovaleraldehyde while other *Magnaporthales*' species converted fumaryl-acetoacetate into fumarate and acetoacetate [178, 179]. Other roles played by the *Sordariomycete*'s species were aromatic compounds metabolism, in which, phenolics and aryl alcohols/aldehydes were converted into acetate, formate and ethanol [180, 181]. Similarity the *Sordariomycete*'s community's abundances were also affected by the substrates and theenvironmental conditions (Table 1 and 2).

Contrasting the adaptability of the *Ascomycota* species, *Basidiomycota* affiliates estabilshed susceptible hyphal networks [182] that were constrained by the reactor pertubartions. The revealed trait explained why the *Ascomycota* species dominated over the *Basidiomycota*'s species. Nonetheless, like *Ascomycota* spcies, the affiliates of *Malasseziales*, *Ustilaginales*, *Agaricules* and *Tremellales* species, all belonging to *Basidiomycota*, were able to metabolize carbohydrates [183, 167], urea [183] and $NO_3^-{}_{(aq)}$ [183] in the systems. Other members like *Polyporales* metabolized woody substrates and their respective products [184]. All these indicated *Basidiomycota's* species played wide metabolic roles in the AD systems. Interestingly, we also identified *Chytridiomycota* affiliates in reactor 9 and 10 and the members were capable of catabolizing pollen, chitin, keratin, and cellulose [185, 186], which indicated the two treatments harbored the substrates. The detected *Spizellomycetale*s, *Cladochytriale*s and *Rhizophydiale*s species in the two treatments, further revealed the presence of high decayed plant tissue concentration with high cellulose content. The identified *Monoblepharidomycetes* members were close relatives of *Cladochytriales* and

Spizellomycetales [187] species and were detected only in reactor 9 [187] and were postulated to perform similar roles in the AD systems.

Generally, *Chytridiomycota* species were basically carbohydrate metabolizers as previously revealed by the presence of glucose transporters and invertase activity in their systems [185, 186]. The species used electron bifurcation mechanisms to conserve their energy and the mechanisms occured in the hydrogenosome [188]. The *Blastocladiales*, the close relatives of *Chytridiomycotes* and the affiliates of *Blastocladomycota*, class *Blastocladomycetes* were only detected in reactor 9. Their member species metabolized cellulose into acetate, $CO_{2(g)}$ and $H_{2(g)}$. Interestingly, *Blastocladiales* were previously reported to use Malate fermentation pathway to generate and conserve their energy [189]. *Mortierellales* formerly classified under *Mucoromycotina* [190] and latter assigned to *Mortierellomycotina* phylum [191] were the converters of organic compounds into poly-unsaturated fatty acids [192] while their relatives *Entomophthorales* and *Mucorales* metabolized decaying organic matter. The affiliates of the three orders were only detected in reactor 9, 10 and 11 respectively.

Like other organisms, fungi co-occurred with the species of *Bacteria* and methanogens [193, 194]. In these syntrophy, the fungal species catalyzed inter-species $H_{2(g)}$ transfer and re-generated oxidized reducing equivalents; NAD^+ and $NADP^+$ in the systems. All these reactions increased metabolism of the dry matter [195] in the treatments. The fungal syntrophy was more complex than simple cross-feeding mechanisms, because the species shifted more oxidized products into a more reduced form. For example the precursor, lactate and ethanol were converted into acetate and formate depending on the prevailing conditions. The involved interactions between fungi and other domain's species were very crucial to an extent of influencing isolation and culturing of some fungal species [196, 195].

Microbial abundances, β-diversity, and richness

To understand whether environmental conditions and perturbations influenced microbial populations, we collected the phenotypic traits of the treatment and compared it with the genotypic dataset. The phenotyphic datasets revealed an insight on the cause of the observed communities' β-diversities, (at $P \leq 0.05$) and biogas production variation. In this case we combined principles of four ecological mechanisms;-selection, drift, speciation, and dispersal to better understand the cause of varience among the treatments. While selection was deterministic and ecological drift was stochastic, dispersal and speciation contributed to the two ecological parameters. One of the challenges of examining communities' entity in any given niche is the difficulty of estimiting real dispersal among the sites. Nonetheless, our treatments were fluidic, naturally encouraging rapid microbial dispersion. Hence, species dispersal was not the major limiting factor that influenced our local communities. Similarly, we also excluded speciation based on our sampling process and experimental design. Therefore, selection (deterministics factors) and the ecological drift, (stochastic processes) were the two main factors that contributed to variation of community abundances among the treatments and the observed high β-diversities. The two factors largely emanated from the substrate inputs and disturbances from environmental conditions (Table 1 and 2). Temperature, agroecological zones and soil types were among the environmental factors that were observed to drive the specific communities' β-diversities and their relative abundances among the treatments. We postulated that environmental temperatures, combined with substrate composition and other deterministic factors such as the species metabolic traits, largely contributed to the lack of identification of the *Chytridiomycota*, *Blastocladomycota* and other rare fungal species in the majority of the treatments. Nonetheless, we never observed large reactor performances differences among the respective treatments, where these species were identified in low abundances, which indicated the importance of the

species abundance in the biogas production. Substrate composition was largely influenced by several environmental factors that included agroecological zones, soil type, rainfall, environmental temperatures to mention but a few, and its chemical compositions stimulated the growth of species, favouring colonization and speciation in the respective treatments. Similarly the same happened to other communities and the available resources reduced competition among the species, which weakened the niche selection abilities. These ecological factors strengthened ecological drift and prority effects [198]. The result was unpredictable site to site variation of microbial abundances and biogas systems' production abilities as well as high β-diversities, even under similar environmental conditions (Table 1 and 2).

Contrasting the substrate effects, the intensity and duration of the disturbances [199], led to high rate of species mortality, low growth rates and species extinction. All these caused decrease of stochasticity and in this case, we adopted the concepts of neutral theory [200] to explain our findings. The caveats reduced ecological drifts that resulted to low competition ability among the observed low abundances species. Persistence and intensities of these disturbances in the treatments eliminated species from the treatment. One of such identified case was revealed in reactor 5 in which the *Tenericutes* affiliates were absolutely extinct. Similarly the absence of *Mortierellales*, *Entomophthorales*, *Mucorales* to mention but a few, in other treatments was an indication of unfavorable disturbances that were exerted by either the substrates or other reactor perturbations. All these led to more predictable site-to-site variations and we used these concepts to explain the observed low β-diversity in the few reactors (SFig. 51). However, to understand better the links between the environmental conditions and the microbial communites in these AD systems, we suggested the utilization of metacommunity theory that considers the world as a collection of patches that are connected to form a metacommunity through organisms' dispersal. The model depends on the specific species traits, the varience of the patches in the environmental conditions as well as frequency and extent of species dispersal in the systems. Other ecological models lack these features. Therefore, the model can used to ask and test whether dispersal, diversification, environmental selection and ecological drift were the vital eements in the biogas production process. In this case, a detailed understanding of the contribution of these factors on specific species abundances and community assemblage can be used to tailor the reactor management from an ecological point of view.

Conclusion

The results provided an insight into the reactors' microbiome composition, richness, abundances and variations within and among the taxa and the treatments. Interestingly, microbiome composition and richness were similar among the twelve treatments, except the affiliates of *Rhizophydiales*, *Cladochytriales*, *Monoblepharidales*, *Blastocladiales* and *Entomophthorales*, which were detected only in reactor 9. Contrasting the findings were the affiliates of *Mortierellales* and *Mucorales* that were identified in reactor 10 and reactor 11 respectively while the *Spizellomycetales* species were detected in both reactor 9 and 10. We also revealed that the affiliates of *Tenericutes*, were the only bacterial guild that was absent in reactor 5. Substrate and metabolic capabilities of these species were the main elements that were considered to limit their dispersal to other treatments. Nonetheless, the identified phylotypes richness within and among the phyla were found to vary, except for the phyla, that comprised a single phylotypes. Similarly, all the identified species abundances among the treatments and the community's β-diversities in majority of the treatments significantly varied. All these contributed to our conclusion that environmental and spatial conditions largely influenced biogas microbiome populations and thus the amount of biogas production. Therefore, we further suggested studies on species responses on the natural conditions for

optimal biogas production and microbial community managements. The discoveries highlight the importance of identifying the involved natural processes that give rise to such identified variations. We also suggested the application of metacommunity theory among other ecological theories in developing management strategies for the biochemical issues in the AD systems.

Acknowledgement
This work was supported by the BecA-ILRI-hub through ABCF Program and University of Eldoret through Annual research grants. We also appreciate Vincent Njunge, Collin mutai, Tina Kyalo, Joyce Njuguna, Eunice machuka, Muga Fredrick Nganga and Emilda of BecA-ILRI hub and Donald Otieno and Linet Gohole of University of Eldoret for their technical surport.

Authors Contribution
Conceived idea: MSM
Designed experiments: MSM, SOO
Performed experiments: MSM
Performed Sequencing: MSM, ML
Analyzed the data: MSM
Drafted manuscrispt: MSM
Reviewed the manuscript: SMM, SOO, RP, FNW, JMO, PMN

Supplementary Material
STable 1: The quality control statistics of the obtained scaffolds for the twelve treatments
SFig. 1: Bar charts showing a). Unfiltered and filtered sequencing reads and b). The known and unknown protein genes in our samples.
SFig. 2: Stacked barchat (a) showing some (21 phyla) Bacteria domain phyla, relative abundances and the PCoA plot (phylum level) (b), based on the Euclidean model.
SFig. 3: The PCoA plot showing (a) the community composition variation at the class and (b) order level.
SFig. 4: The stacked barchart revealing six classes of *Proteobacteria* communities, the proportions of their relative abundances (a) and their PCoA plots (b) based on the Euclidean model.
SFig. 5: The PCoA plot based on the Euclidean model for all the identified *Proteobacteria* orders.
SFig. 6: Stacked barchat (a) showing seven *δ-Proteobacteria* orders, relative abundances and their PCoA plot (b) revealing nucleotide composition variations, based on the Euclidean model.
SFig.7: Stacked barchart showing fourteen *γ-Proteobacteria* orders (a) and their PCoA plot based on the Eucleaden model.
SFig. 8: Stacked barchat showing seven *β-proteobacteria* orders, relative abundances (a) and their PCoA plot (b) revealing nucleotide composition (dis)similarities among the reactors.
SFig. 9: Stacked barchat showing the (a) seven *α-Proteobacteria* orders, relative abundances and their PCoA plots (b), revealing the nucleotide composition variation among the reactors based on the Euclidean model.
SFig. 10: Stacked barchat (a) showing two *ε-Proteobacteria* orders, relative abundances and their PCoA plot (b) based on Euclidean model.
SFig. 11: The stacked barchat (a) revealing four *Firmicutes* classes, the relative abundances and their PCoA plots (b), revealing variance among the twelve reactors based on the Euclidean model.
SFig. 12: Stacked barchat (a) showing eight *Firmicute'*s orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 13: Stacked barchat (a) showing the four *Clostridia* orders and their PCoA plots (b), revealing nucleotide composition variations among the twelve reactors based on the Euclidean model.
SFig. 14: Stacked barchat (a) showing two *Bacilli* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 15: The stacked barchat (a) revealing four *Bacteroidete*s classes, relative abundances and their

PCoA plots (b) revealing nucleotide composition variation, based on the Euclidean model.
SFig. 16: Stacked barchat (a) showing the four *Bacteroidete*'s orders, the relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 17: Stacked barchat (a) showing six *Actinobacteria* orders, the relative abundance and their PCoA plot (b) based on the Euclidean model.
SFig. 18: The stacked barchat (a) showing four *Chloroflexi* classes, relative abundances and PCoA plot (b) revealing variance among the twelve reactors, based on the Euclidean model.
SFig. 19: Stacked barchat (a) showing five *Chloroflexi* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 20: Stacked barchat (a) showing two *Chloroflexi* class orders, relative abundances and their PCoA plot (b) based on their Euclidean model.
SFig. 21: Stacked barchat (a) showing the two *Thermomicrobia* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 22: Stacked barchat (a) showing *Cyanobacteria* class, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 23: Stacked barchat (a) showing the five *Cyanobacteria* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 24: Stacked barchat (a) showing the four affiliates of the Unclassified *Cyanobacteria* nucleotide reads, relative abundances and their PCoA plots (b), based on the Euclidean model.
SFig. 25: The stacked barchat (a) showing the two *Acidobacteria* classes, relative abundances and their PCoA plots (b) for their nucleotide composition.
SFig. 26: The stacked barchat (a) showing the two *Acidobacteria* orders, relative abundances and their PCoA plots (b) for their nucleotide composition.
SFig. 27: Stacked barchat (a) showing two *Deinococcus-Thermus* orders, the relative abundances and the PCoA plots (b) based on the Euclidean model.
SFig. 28: Stacked barchat (a) showing *Verrucomicrobia* classes, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 29: Stacked barchat (a) showing three *Verrucomicrobia* orders, relative abundances and their PCoA plot (b) based Euclidean model.
SFig. 30: Stacked barchat (a) showing three *Terneicutes* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 31: The stacked barchart (a) showing the four *Archaea* phyla, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 32: Stacked barchat (a) showing nine *Archaea* classes, and the affiliates of the unclassified reads, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 33: Stacked barchat showing sixteen *Archaea* orders, relative abundances and their PCoA plot based on the Euclidean model.
SFig. 34: The stacked barchat (a) showing the eight *Eurychaeota* classes, the proportion of relative abundances and their PCoA plots (b) revealing dissimilarities of the nucleotide composition among the twelve studied reactors
SFig. 35: Stacked barchat (a) showing ten *Euryarchaeota* orders, relative abundances and their PCoA plot (b) based Euclidean model.
SFig. 36: The stacked barchat (a) showing three *Methanomicrobia* orders, the proportion of relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 37: Stacked barchat (a) showing four *Thermoprotei* orders and their PCoA plot (b) based on the Euclidean model.
SFig. 38: Stacked barchat (a) showing two *Thaurmarchaeota* order, the relative abundances, and their PCoA plot (b) based on the Euclidean model.
SFig. 39: Stacked barchat (a) showing five fungal phyla, proportion of the relative abundances and their PCoA plots (b) based on Euclidean model.
SFig. 40: Stacked barchat (a) showing thirteen fungal classes, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 41: Stacked barchat (a) showing 23 fungal orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 42: The stacked barchat (a) showing the five *Ascomycota* classes, relative abaundances and

their PCoA plots (b), based on the Euclidean model at the class level.
SFig. 43: The stacked barchat (a) revealing eleven *Ascomycota* orders, proportions of relative abundances and their PCoA plot (b); based on the Euclidean model revealed dissimilarities among the 12 reactors
SFig. 44: Stacked barchat (a) showing three *Sordariomycetes* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 45: Stacked barchat (a) showing the two *Eurotiomycetes* orders, the proportion of relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 46: Stacked barchat (a) showing four *Basidiomycota* classes, the proportion of the relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 47: Stacked barchat (a) showing four *Basidiomycota* orders, relative abundances and their PCoA plot (b) based on the Euclidean model.
SFig. 48: Stacked barchart (a) showing the proportion of the *Agaricomycetes* order relative abundances and their PCoA plot (b) based on the Euclidean model at the order level.
SFig. 49 : Stacked barchat showing the two *Chytridiomycota* classes, and their relative abundances in the two eactors.
SFig. 50: Stacked barchat showing the four *Chytridiomycota* orders, and their relative abundances in the two reactors.
SFig. 51: Stacked barchat showing the three orders affiliated to unclassified fungal nucleotide reads.
SFig. 52: The PCoA analysis revealing β-diversity of the twelve reactors at the three taxa level: a) Domain, b) Phylum, and c) Class level.

References

1. Calusinska M., Goux X., Fossépré M., Muller E., Wilmes P. and Delfosse P. (2018). A year of monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but different core microbiomes in bio-waste and wastewater anaerobic digestion systems. *BMC Biotechnol. Biofuel*, **11:**196.
2. Weiland P. (2010). Biogas production: current state and perspectives. *Appl. Microbiol. Biotechnol.*, **85:** 849-860.
3. Verstraete W. (2015). The manufacturing microbe. *Microb. Biotechnol*. **8:** 36-7.
4. Austin G. and Morris G. (2012). Biogas production in Africa, in: R. Janssen, D. Rutz (Eds.), Bioenergy Sustain. Dev. Africa, Springer, Dordrecht, 2012.
5. Scarlat N., Dallemand J-F. and Fahl F. (2018). Biogas: Developments and perspectives in Europe. *Renew. Energ*., **129:** 457-472.
6. Tiwary A., Williams I., Pant D. and Kishore W. (2015). Emerging perspectives on environmental burden minimiasation initiatives from anaerobic digestion technologies for community scale biomass valorisation. *Renew. Sustain. Energ. Rev*., **42:** 883-901.
7. Africa Biogas Partnership Programme, (2017). http://africabiogas.org/.
8. Mizrahi-Man O., Davenport E. and Gilad Y. (2013). Taxonomic classification of *Bacterial* 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS one*, **8:** e53608.
9. Campanaro S., Treu L., Kougias G., Zhu X. and Angelidaki I. (2018). Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Scient. Rep*., **8:** 1926.
10. Fox G., Wisotzkey J. and Jurtshuk P. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol*., **42:** 166-170.
11. Locey K. and Lennon J. (2016). Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. USA*. **113:** 5970-5975.
12. Locey K. and Lennon J. (2016). Reply to Willis: Powerful predictions of biodiversity from ecological models and scaling laws. *Proc. Natl. Acad. Sci. USA*., **113:** E5097-E5097.
13. Shoemaker W., Locey K. and Lennon J. (2017). A macroecological theory of microbial biodiversity. *Nat. Ecol. Evol*., **1:** 0107.
14. Dykhuizen D. (1998). Santa Rosalia revisited: why are there so many species of Bacteria? *Antonie Van Leeuwenhoek*, **73:** 25-33.
15. Schloss P., Girard R., Martin T., Edwards J. and Thrash J. (2016). Status of the Archaeal and Bacterial census: an update. *MBio*., **7:** e00201-16.
16. Amann R. and Rosselló-Móra R. (2016). After all, only millions? *MBio*., **7:** e000999-16.
17. Mora C., Tittensor D., Adl S., Simpson A. and Worm B. (2011). How Many Species Are There on Earth and in the Ocean? *PLOS Biol*., **9:** e1001127.
18. Wittebolle L., Marzorati M., Clement L., Balloi A., Daffonchio D., Heylen K., …, Boon N. (2009). Initial community evenness favours functionality under selective stress. *Nat*. **458:** 623-6.
19. Carballa M., Regueiro L. and Lema J. (2015). Microbial management of anaerobic digestion: exploiting the microbiome-functionality nexus. *Curr. Opin. Biotechnol*., **33:** 103-11.
20. Huse S., Huber J., Morrison H., Sogin M. and Welch D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8:** R143.
21. Grant T. (2010). *The Randomized Complete Block Design (RCBD).* Department of Horticulture and Crop Science, OARDC, The Ohio State University, USA.
22. Boaro A., Kim Y., Konopka A., Callister S. and Ahring B. (2014). Integrated 'Omics' analysis for studying the microbial community response to a pH perturbation of a cellulose-degrading bioreactor culture. *FEMS Microbiol. Ecol*., **90:** 802-815.
23. Menon, S. (2014). ArcGIS 10.3: The Next Generation of GIS Is Here. Accessed June, 2019.
24. Oksenberg L., Cannell, C. and Kalton G. (1991). New strategies for pretesting survey questions". *J. Off. Statist*., **7:** 349-365.
25. Campanaro, S., Treu L., Kougias P., 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. *BMC Biotechnol. Biofuel*, **9:** 26.
26. Caporaso J., Lauber C., Walters W., Berg-Lyons D., Huntley J., Fierer N., …, Knight R. (2012).

Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6:** 1621-4.
27. Zhou Q., Su X., and Ning K. (2014). Assessment of quality control approaches for metagenomic data analysis. *Scient. Rep.* **4:** 6957.
28. Bolger A., Lohse M. and Usadel B. (2014). Trimmomatic: A flexible trimmer for 588 Illumina sequence data. *BMC Bioinformatic*, **30:** 2114-2120.
29. Nurk S., Meleshko D., Korobeynikov A., and Pevzner P. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res*., **27:** 824-834.
30. Cox M., Peterson D. and Biggs P. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatic*, **11:** 485.
31. Keegan K., Trimble W., Wilkening J., Wilke A., Harrison T., D'Souza M. and Meyer F. (2012). A platform-independent method for detecting errors in metagenomic sequencing data: DRISEE. *PLOS Comput. Biol*., 8(6):e1002541.
32. Langmead B., Trapnell C., Pop M. and Salzberg S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10:** R25.
33. Wilke A., Harrison T., Wilkening J., Field D., Glass E., Kyrpides N., …, Meyer F. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatic*, **13:** 141.
34. Rho M., Tang H. and Ye Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucl. Acids Res*., **38:** e191
35. Kent W. (2002). BLAT-the BLAST-like alignment tool. *Genome Res*., **12:** 656-664.
36. Field D., Sterk P., Kottmann R., Wim De Smet J., Amaral-Zettler L., Cochrane G., …, *et al*. (2014). Genomic standards consortium projects. *Stand. Genome Sci*., 9: 599-601.
37. Glass E., Dribinsky Y., Yilmaz P., Levin H., Van Pelt R., Wendel D., …, Schriml L. (2014). MIxS-BE: a MIxS extension defining a minimum information standard for sequence data from the built environment. *ISME J*., **8:** 1-3.
38. Angiuoli S., Matalka M., Gussman A., Galens K., Vangala M., Riley D., …, Fricke W. (2011). CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatic*, **12:** 356.
39. Terry Speed (2003). *Statistical Analysis of Gene Expression Microarray Data*. Chapman and Hall/CRC, 2003.
40. Meyer F., Paarmann D., D'Souza M., Olson1 R., Glass E., Kubal M., …, Edwards R. (2008).The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatic*, **9:** 386.
41. Ondov B., Bergman N. and Phillippy A. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatic*, **12:** 385.
42. Gomez-Alvarez, V., Teal, T. and Schmidt T. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J*., **3:** 1314-7.
43. Shankar V. (2016). Extension of multivariate analyses to the field of microbial ecology. Thesis, Wright State University.
44. Overbeek R., Begley T., Butler R., Choudhuri J., Diaz N., Chuang H-Y., …, Vonstein V. (2005). The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. *Nucl. Acids Res*., **33:** 17.
45. Luo G., Fotidis I. and Angelidaki I. (2016). Comparative analysis of taxonomic, functional, and metabolic patterns of microbiomes from 14 full-scale biogas reactors by metagenomic sequencing and radioisotopic analysis. *BMC Biotechnol. Biofuel*., **9:** 1-12.
46. Knight, R. et al. (2012). Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol*. **30:** 513-520.
47. Stolze Y., Zakrzewski M., Maus I., Eikmeyer F., Jaenicke S., Rottmann N., …, Schlüter A. (2015). Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. *BMC Biotechnol. Biofuel*, **8:** 14.
48. Lynd L., Weimer P., van Zyl W. and Pretorius I. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev*., **66:** 506-577.

49. Gold D. and Martin J. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol*., **189:** 6787-6795.
50. Seedorf H., Fricke W., Veith B., Brüggemann H., Liesegang H., Strittmatter A., …, Gottschalk G. (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Proc. Natl. Acad. Sci. USA*, **105:** 2128-2133.
51. Kaji M., Taniguchi Y., Matsushita O., Katayama S., Miyata S., Morita S. and Okabe A. (1999). The hydA gene encoding the H_2 evolving hydrogenase of *Clostridium perfingens:* molecular characterization and expression of the gene. *FEMS Microbiol. Lett*., **181:** 329-336.
52. Fonknechten N., Chaussonnerie S., Tricot S., Lajus A., Andreesen J., Perchat N., …, Kreimeyer A. (2010). *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomic*, **11:** 555.
53. Lawson P., Song Y., Chengxu L., Denise R., Vaisanen M., Collins M. and Finegold S. (2004). *Anaerotruncus colihominis* gen. nov. sp. nov. from human feces. *IJSEM*., **54:** 413-417.
54. Duncan S., Hold G., Harmsen H., Stewart C. and Flimt H. (2002). Growth requirements and fermentation products of *Fusobacterium prausnitzii* and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov. comb. nov. *IJSEM*, **52:** 2141-2146.
55. Nokaga H., Keresztes G., Shinoda Y., Ikenaga Y., Abe M., Naito K., …, Yukawa H. (2006). Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense Y51* and comparison with *Dehalococcoides ethenogenes*. *J. Bacteriol*., **188:** 2262-2274.
56. Debarati P., Frank W., Arick T., Bridges S., Burgess S., Yoginder S. and Lawrence M. (2010). Genome sequence of the solvent-producing bacterium *Clostridium carboxidivorans* strain P7. *J. Bacteriol.*, **192:** 5554-5.
57. Rother M. and Oelgeschläger E. (2008). Carbon monoxide-dependent energy metabolism in anaerobic Bacteria. *Arch. Microbiol*., **190:** 257-269.
58. Simankova M., Chernych N., Osipov G. and Zavarzin G. (1993). *Halocella cellulolytica* gen. nov., sp. nov., a new obligately anaerobic, halophilic, cellulolytic bacterium. *Syst. Appl. Microbiol.* 16:385-389.
59. Wei C., Kunio O. and Shoichi S. (1987). Intergeneric protoplast fusion between *Fusobacterium varium* and *Enterococcus faecium* for enhancing dehydrodivanillin degradation. *Appl. Environ. Microbiol*., **53:** 542-548.
60. Guerra N., Fajardo P., Fuciños C., Amado R., Alonso E., Torrado A. and Pastrana L. (2010). Modelling the biphasic growth and product formation by *Enterococcus faecium* CECT 410 in realkalized fed-batch fermentations in whey. *J. Biomed. Biotechnol*., 290286.
61. *Enterococcus faecalis* hydrogenase 3. http://biocyc.org/EFAE749511-HMP/NEWIMAGE?object = Hydrogenase-3.
62. de Bok F., Stams A., Dijkema C. and Boone D. (2001). Pathway of Propionate Oxidation by a Syntrophic Culture of *Smithella propionica* and *Methanospirillum hungatei*. *Appl. Environ. Microbiol*., **67:** 1800-1804.
63. Schmidt A., Muller N., Schink B. and Schleheck D. (2013). A Proteomic View at the Biochemistry of Syntrophic Butyrate Oxidation in Syntrophomonas wolfei. PLoS ONE **8:** e56905.
64. Dubbels B., Sayavedra-Soto L., Bottomley P. and Arp D. (2009). *Thauera butanivorans* sp. nov., a C2- C9 alkane-oxidizing bacterium previously referred to as '*Pseudomonas butanovora*'. *Int. J. Syst. Evol. Microbiol*., **59:** 1577-1578.
65. Baena S., Fardeau M., Labat M., Ollivier B., Thomas P., Garcia J. and Patel B. (1998). *Aminobacterium colombiense* gen. nov. sp. nov., an amino acid-degrading anaerobe isolated from anaerobic sludge. *Anaerobe*, **4:** 241-50.
66. Chertkov O., Sikorski J., Brambilla E., Lapidus A., Copeland A., Glavina Del Rio T., …, Klenk H-P. (2010). Complete genome sequence of *Aminobacterium colombiense* type strain (ALA-1). *Stand Genomic Sci*., **2:** 280-9.
67. Wei C., Kunio O. and Shoichi S. (1987). Intergeneric protoplast fusion between *Fusobacterium varium* and *Enterococcus faecium* for enhancing dehydrodivanillin degradation. *Appl. Environ. Microbiol*., **53:** 542-548.
68. Kindaichi T., Ito T. and Okabe S. (2004). Ecophysiological interaction between nitrifying Bacteria and heterotrophic Bacteria in autotrophic nitrifying biofilms as determined by

microautoradiography-fluorescence in situ hybridization. *Appl. Environ. Microbiol.*, **70:** 1641-1650.
69. Fonknechten N., Chaussonnerie S., Tricot S., Lajus A., Andreesen J., Perchat N., …, Kreimeyer A. (2010). *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomic*, **11:** 555.
70. Bjursel M., Martens E. and Gordon J. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J. Biol. Chem*., **281**: 36269-36279.
71. Li H., Shen T., Wang X., Lin K., Liu Y., Lu S., …, Du X. (2013). Biodegradation of perchloroethylene and chlorophenol co-contamination and toxic effect on activated sludge performance. *Bioresour. Technol*., **137:** 286-93.
72. Xie G, et al. (2007). Genome sequence of the cellulolytic gliding bacterium, *Cytophaga hutchinsonii*. *Appl. Environ. Microbiol*. **73:** 3536-3546.
73. Hiras J., Wu Y-W., Eichorst S., Simmons B. and Singer S. (2016). Refining the phylum Chlorobi by resolving the phylogeny and metabolic potential of the representative of a deeply branching, uncultivated lineage. *The ISME J*., **10**: 833-845.
74. Wahlund TM, Madigan MT. (1993). Nitrogen-fixation by the thermophilic green sulfur bacterium *Chlorobium tepidum*. *J. Bacteriol.*, **175:** 474-478.
75. Wahlund T. and Tabita F. (1997). The reductive tricarboxylic acid cycle of carbon dioxide assimilation: Initial studies and purification of ATP-citrate lyase from the green sulfur bacterium *Chlorobium tepidum*. *J. Bacteriol*. **179:** 4859-4867.
76. Ghilarov D., Serebryakova M., Stevenson C., Hearnshaw S., Volkov D., Maxwell, A., …, Severinov, K. (2017). The origins of specificity in the microcin-processing protease TldD/E. *Structure*, **25:** 1549-1561.
77. Moon C., Young W., Maclean P., Cookson A. and Bermingham E. (2018). Metagenomic insights into the roles of proteoBacteria in the gastrointestinal microbiomes of healthy dogs and cats. *MicrobiologyOpen*, **7:** e677.
78. Han K., Lim H. and Hong (1992). Acetic acid formation in *Escherichia coli* fermentation. *Biotechnol. Bioeng*., **15:**663-671.
79. Textor S., Wendisch V., de Graaf A., Mu¨ller U., Linder M., Linder D. and Buckel W. (1997). Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in Bacteria. *Arch. Microbiol*. **168:** 428-436.
80. Hippe H., Hagelstein A., Kramer I., Swiderski J and Stackebrandt E. (1999). Phylogenetic analysis of *Formivibrio citricus*, *Propionivibrio dicarboxylicus*, *Anaerobiospirillum thomasii*, *Succinimonas amylolytica* and *Succinivibrio dextrinosolvens* and proposal of Succinivibrionaceae fam. nov. *Int. J. Syst. Bacteriol*., **49:** 779-82.
81. Greetham H., Collins M., Gibson G. Giffard C., Falsen E. and Lawson P. (2004). *Sutterella stercoricanis* sp. nov., isolated from canine faeces. *Int. J.Syst. Evol. Microbiol*., **54**: 1581-1584.
82. Hottes AK, et al. (2004). Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. *J. Bacteriol*., **186:** 1448-1461.
83. Müller V. and Grüber G. (2003). ATP synthases: structure, function and evolution of unique energy converters. *Cell. Mol. Life Sci*., **60:** 474-94.
84. Manconi R. and Pronzato R. (2008). Global diversity of sponges (Porifera: Spongillina) in freshwater. *Hydrobiol.*, **595:** 27-33.
85. Mclnernery M., Mackie I. and Bryant M. (1981). Syntrophic association of butyrate degrading bacterium and methanosarcina enriched from bovine rumen fluid: *Appl. Environ. Microbiol*., **42:** 826-828.
86. Meher K. and Ranade D. (1993). Isolation of propionate degrading bacterium in co-culture with a methanogen from a cattle dung biogas plant. *J. Biosci.*, **18:** 271-277.
87. Boone D. and Bryant M. (1980). Propionate-degrading bacterium, *Syntrophobacter wolinii* sp. nov. gen. nov., from methanogenic ecosystems. *Appl. Environ. Microbiol*., **40:** 626-632.
88. Mountfort D. and Bryant M. (1982). Isolation and characterization of an anaerobic syntrophic benzoate-degrading bacterium from sewage sludge. *Arch. Microbiol*., **133:** 249-256.

89. Spring S., Sorokin D., Verbarg S., Rohde M., Woyke T. and Kyrpides N. (2019). Sulfate-reducing Bacteria that produce exopolymers thrive in the calcifying zone of a Hypersaline CyanoBacterial Mat. 1
90. Reichenbach H. and Dworkin M. (1991). *The myxoBacteria*. In: The Prokaryotes (Starr, K.P., Stolp, H., Truper, H.G., Belows, A. and Schlegel, H.G., Eds.), Vol. 4, pp. 3415-3487. Springer, New York.
91. Dawid W. (2000). Biology and global distribution of myxoBacteria in soils. *FEMS Microbiol. Rev*., **24:** 403-427.
92. Johannessen K., Vander-Roost J., Dahle H. and Thorseth I. (2017). Biomineralization and Fe-oxyhydroxide mound formation in a diffuse hydrothermal system at the Jan Mayen vent fields. *Geochim. Cosmoch. Acta*. **202:** 101-123.
93. Vander Roost J., Thorseth I. and Dahle H. (2017). Microbial analysis of Z-proteoBacteria and co-colonizers of iron mats in the Troll Wall Vent Field, Arctic Mid-Ocean Ridge. *PLOS ONE*, **12:** e0185008.
94. Hoshino T., Kuratomi T., Morono Y., Hori T., Oiwane H., Kiyokawa S. and Inagaki F. (2016). Ecophysiology of *Z-Proteobacteria* associated with shallow hydrothermal iron-oxyhydroxide deposits in Nagahama Bay of Satsuma Iwo-Jima, Japan. *Front. Microbiol*., **6:** 1554.
95. Pankratov T. and Dedysh S. (2010). *Granulicella paludicola* gen. nov., sp. nov., *Granulicella pectinivorans* sp. nov., *Granulicella aggregans* sp. nov. and *Granulicella rosea* nov., acidophilic, polymer-degrading acidoBacteria from Sphagnum peat bogs. *Int. J. Syst. Evol. Microbiol*., **60:** 2951-2959.
96. Eichorst S., Trojan D., Roux S., Herbold C., Rattei T. and Woebken D. (2018). Genomic insights into the *Acidobacteria* reveal strategies for their success in terrestrial environments. *Environ. Microbiol*., **20:** 1041-1063.
97. Greening C., Carere C., Rushton-Green R., Harold L., Hards K., Taylor M., …, Cook G. (2015). Persistence of the dominant soil phylum *Acidobacteria* by trace gas scavenging. *Proc. Natl. Acad. Sci. USA*., **112:** 10497-10502.
98. Lazarev V., Levitskii S., Basovskii Y., Chukin M., Akopian T., Vereshchagin V., …, Govorun V. (2011). Complete genome and proteome of *Acoleplasma laidlawii*. *J. Bacteriol*., **193:** 4943-4953.
99. Freundt E., Whitcomb R., Barile M., Razin S. and Tully J. (1984). Proposal for elevation of the family *Acholeplasmataceae* to ordinal rank: *Acholeplasmatales*. *Int. J. Syst. Bacteriol*., **34:** 346-349.
100. Ventura M., Canchaya C., Tauch A., Chandra G., Fitzgerald F., Chater F. and van Sinderen D. (2007). Genomics of *Actinobacteria*: Tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev*., **71:** 495-548.
101. Schell M., Karmrabűntzon M., Snel B., Vilanova D., Berger B., Pessi G., …, Arigoni F. (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. USA*., **2002:** 99.
102. Zakrzewski M., Goesmann A., Jaenicke S., Jünemann S., Eikmeyer F., Szczepanowski R., Al-Soud W., Sørensen S., Pühler A. and Schlüter A. (2012). Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. *J. Biotechnol*., **158:** 248-58.
103. Fuerst J. and Sagulenko E. (2011). Beyond the bacterium: *Planctomycetes* challenge our concepts of microbial structure and function. *Nat. Rev. Microbiol*., **9:** 403-413.
104. Kampmann K., Ratering S., Kramer I., Schmidt M., Zerr W. and Schnell S. (2012). Unexpected stability of *Bacteroidetes* and *Firmicutes* communities in laboratory biogas reactors fed with different defined substrates. *Appl. Environ. Microbiol*., **78:** 2106-2119.
105. Krakat N., Westphal A., Schmidt S. and Scherer P. (2010). Anaerobic digestion of renewable biomass: thermophilic temperature governs population dynamics of methanogens. *Appl. Environ. Microbiol*., **77:** 1842-1850.
106. Weiss A., Jérôme V., Freitag R. and Mayer H. (2008). Diversity of the resident microbiota in a thermophilic municipal biogas plant. *Appl. Microbiol. Biotechnol*., **81:** 163-73.

107. Maus I., Cibis K., Wibberg D., Winkler A., Stolze Y., König H., Pühler A. and Schlüter A. (2015). Complete genome sequence of the strain *Defluviitoga tunisiensis* L3, isolated from a thermophilic, production-scale biogas plant. *J. Biotechnol*., **203:** 17-8.
108. Xia Y., Wang Y., Fang H., Jin T., Zhong H. and Zhang T. (2014). Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Scient. Rep*., **4:** 6708.
109. Wang W., Xie L., Luo G., Zhou Q. and Angelidaki I. (2013). Performance and microbial community analysis of the anaerobic reactor with coke oven gas biomethanation and in situ biogas upgrading. *Bioresour. Technol*., **146:** 23-9.
110. Siriwongrungson V., Zeng R. and Angelidaki I. (2007). Homoacetogenesis as the alternative pathway for H_2 sink during thermophilic anaerobic degradation of butyrate under suppressed methanogenesis. Water Res. **41:** 4204-10.
111. Eichorst S. and Kuske C. (2012). Identification of cellulose-responsive Bacterial and fungal communities in geographically and edaphically different soils by using stable isotope probing. *Appl. Environ. Microbiol*., **78:** 2316-2327.
112. Appel J. and Schulz R. (1998). Hydrogen metabolism in organisms with oxygenic photosynthesis: hydrogenases as important regulatory devices for a proper redox poising? *J. Photochem. Photobiol. Ser*. **B47:** 1-11.
113. Bergman B., Gallon J., Rai A. and Stal L. (1997). N_2 fixation by non-heterocystous cyanoBacteria. *FEMS Microbiol. Rev*., **19:** 139-185.
114. Lindblad P. (1999). CyanoBacterial H_2-metabolism: knowledge and potential/strategies for a photo-biotechnological production of H_2. *Biotecnol. Apl*. **16:** 141-144.
115. Howarth D. and Codd A. (1985). The uptake and production of molecular hydrogen by unicellular CyanoBacteria. *J. Gen. Microbiol*., **131:** 1561-1569.
116. Weisshaar H. and Böger P. (1985). Pathways of hydrogen uptake in the cyanobacterium *Nostoc muscorum*. *Arch. Microbiol*., **142:** 349-353.
117. Siegl A., Kamke J., Hochmuth T., Piel J., Richter M., Liang C., Dandekar T. and Hentschel U. (2011). Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. *The ISME J.* **5:** 61-70.
118. Pelletier E., Kreimeyer A., Bocs S., Rouy Z., Gyapay G., Chouari R., …, Denis Le Paslier (2008). "*Candidatus Cloacamonas acidaminovorans*": genome sequence reconstruction provides a first glimpse of a new Bacterial division. *J. Bacteriol*., **190:** 2572-2579.
119. Guiavarch E., Pons A., Christophe G., Creuly C. and Dussap C-G. (2010). Analysis of a continuous culture of *Fibrobacter succinogenes* S85 on a standardized glucose medium. *Bioprocess Biosyst. Eng*. **33:** 417-25
120. Ye Q., Roh Y., Carroll S., Blair B., Zhou J., Zhang C. and Fields M. (2004). Alkaline anaerobic respiration: isolation and characterization of a novel alkaliphilic and metal-reducing bacterium. *Appl. Environ. Microbiol.*, **70:** 5595-5602.
121. Fisher E., Dawson A., Polshyna G., Lisak J., Crable B., Detter J., …, Fields M. (2008). Transformation of inorganic and organic arsenic by *Alkaliphilus oremlandii* sp. nov. strain OhILAs. *Ann. NY. Acad. Sci.* **1125:** 230-241.
122. Daims H., Lebedeva E., Pjevac P., Han P., Herbold C., Albertsen M., …, Wagner M. (2015). Complete nitrification by *Nitrospira* bacteria. *Nat*., **528:** 504-509.
123. Koch H., van Kessel M. and Lücke S. (2019). Complete nitrification: insights into the ecophysiology of comammox *Nitrospira*. *Appl. Microbiol. Biotechnol*., **103:**177-189
124. Kotsyurbenko O., Glagolev M., Nozhevnikova A. and Conrad R. (2001). Competition between homoacetogenic *Bacteria* and methanogenic *Archaea* for hydrogen at low temperature. *FEMS Microbiol. Ecol*., **38:** 153-159.
125. Thauer R., Kaster A., Seedorf H., Buckel W. and Hedderich R. (2008). Methanogenic *Archaea:* ecological relevant differences in energy conservation. *Nat. Rev. Microbiol*., 6:579-589.
126. Deppenmeier U. and Müller V. (2008). Life close to the thermodynamic limit: how methanogenic Archaea conserve energy. *Result and problem cell different*., **45:**123-152.

127. Rother M. and Melchaf W. (2004). Anaerobic growth of *Methanosarcina acetivorans C2A* on carbon monoxide: An unusual way of life for a methanogenic archaeon. *PNAS*, **101:** 16929-16934.
128. Barry K, Pitluck S, et al. (2009). Complete genome sequence of *Methanocorpusculum labreanum* type strain Z. *Stand Genomic Sci*., **1:**197-203.
129. Gilmore S., Henske J., Sexton J., Solomon K., Seppälä S., Yoo J., …, O'Malley M. (2017). Genomic analysis of methanogenic *Archaea* reveals a shift towards energy conservation. *BMC Genomic*, **18:** 639.
130. Zellner G., Stackebrandt E., Messner P., Tindall B., Demacario E., Kneifel H., Sleytr U. and Winter J. (1989). *Methanocorpusculaceae* fam-nov, represented by *Methanocorpusculum parvum*, *Methanocorpusculum sinense* spec-nov and *Methanocorpusculum bavaricum* spec-nov. *Arch. Microbiol*., **151:** 381-90.
131. Anderson I., Sieprawska-Lupa M., Lapidus A., Nolan M., Copeland A., Glavina Del Rio T., …, Kyrpides N. (2009).Complete genome sequence of Methanoculleus marisnigri Romesser et al. 1981 type strain JR1. *Stand.Genomic Sci.*, **1:** 189-196.
132. Rivera M. and Lake J. (1996). The phylogeny of *Methanopyrus kandleri*. *Int. J.Syst. Bacteriol.*, **46:** 348-351.
133. Burggraf S., Fricke H., Neuner A., Kristjansson J., Rouvier P., Mandelco L., …, Stetter K. (1990). *Methanococcus igneus* sp. nov., a Novel Hyperthermophilic Methanogen from a Shallow Submarine Hydrothermal System. *Syst. Appl. Microbiol.*, **13:** 263-269.
134. Ding J., Chen S., Lai M. and Liao T. (2017). *Haloterrigena mahii* sp. nov., an extremely halophilic archaeon from a solar saltern. *Int. J. Syst. and Evolut. Microbiol*., **67:** 1333-1338.
135. Oren A. (2006). *The order HaloBacteriales*. In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. A Handbook on the Biology of Bacteria. Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes. Springer; pp. 113-164.
136. Huber H. and Stetter K. (2006). *Thermoplasmatales*. In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. A Handbook on the Biology of Bacteria. Volume 3: *Archaea*. *Bacteria*: *Firmicutes*, *Actinomycetes*. Springer; pp. 101-112. 57
137. Lang, K., Schuldes, J., Kling A., Poehlein, A., Daniel R., and Brunea A. (2015). New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of "*Candidatus Methanoplasma termitum*". *Appl. Environ. Microbiol*., **81:** 1338-52.
138. Gorlas A., Alain K., Bienvenu N. and Geslin C. (2013). *Thermococcus prieurii* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol*., **63:** 2920e2926.
139. Cossu M., Da-Cunha V., Toffano-Nioche C., Forterre P. and Oberto J. (2015). Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged *Thermococcales* chromosomes. *Biochimie*, **118:** 313e321.
140. Brasen C., Esser D., Rauch B. and Siebers B. (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev*., **78:** 89e175.
141. Wang X., Gao Z., Xu X. and Ruan L. (2011). Complete genome sequence of *Thermococcus sp*. Strain 4557, a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent area. *J. Bacteriol*., **193:** 5544-5545
142. Lee H., Kang S., Bae S., Lim J.,Cho Y., Kim Y., …, Lee J-H. (2008).The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydotrophic metabolism. *J. Bacteriol*., **190:** 7491-7499
143. Hartzell P. and Reed D. (2006). *The genus Archaeoglobus*. In Dworkin M, Falkow S, Rosenberg E., Schleifer K., Stackebrandt E., editors. The Prokaryotes. A Handbook on the Biology of Bacteria. Volume 3: *Archaea*. *Bacteria*: *Firmicutes*, *Actinomycetes*. Springer, pp. 82-100.
144. Kashefi K., Tor J., Holmes D., Gaw Van Praagh C., Reysenbach A-L and Lovley D. (2002). *Geoglobus ahangari* gen. nov., sp. nov., a novel hyperthermophilic archaeon capable of oxidizing organic acids and growing autotrophically on hydrogen with Fe(III) serving as the sole electron acceptor. *Int. J. Syst. Evol. Microbiol*., **52:** 719-28.

145. Dai X., Wang H., Zhang Z., Li K., Zhang X., Mora-López M., …, Huang L. (2016). Genome sequencing of *Sulfolobus sp. A20* from costa rica and comparative analyses of the putative pathways of carbon, nitrogen, and sulfur metabolism in various sulfolobus strains. *Front. Microbiol*., **7:** 1902.
146. Sakai H. and Kurosawa N. (2019). Complete genome sequence of the *Sulfodiicoccus acidiphilus* strain HS-1T, the frst crenarchaeon that lacks polB3, isolated from an acidic hot spring in Ohwaku-dani, Hakone, Japan. *BMC Res., Notes*, **12:** 444.
147. Berg I., Kockelkorn D., Buckel W and Fuchs G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea*. *Sci*. **318:** 1782-1786.
148. Siebers B., Tjaden B., Michalke K., Do¨rr C., Ahmed H., Zaparty M., …, Hensel R. (2004). Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. *J. Bacteriol*., **186:** 2179-2194.
149. Zaparty M., Tjaden B., Hensel R. and Siebers B. (2008). The central carbohydrate metabolism of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: pathways and insights into their regulation. *Arch. Microbiol.*, **190:** 231-245.
150. Cossu M., Cunha V., Toffano-Nioche C., Forterre P. and Oberto J. (2015). Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged Thermococcales chromosomes. Biochimie, **118:** 313e321.
151. Prokofeva M., Kostrikina N., Kolganova T., Tourova T., Lysenko A., Lebedinsky A., and Bonch-Osmolovskaya E. (2009). Isolation of the anaerobic thermoacidophilic crenarchaeote *Acidilobus saccharovorans* sp. nov. and proposal of *Acidilobales* ord. nov., including *Acidilobaceae* fam. nov. and *Caldisphaeraceae* fam. nov. *Int. J. Syst. Evol. Microbiol*., **59:** 3116-3122.
152. Mardanov A., Svetlitchnyi V., Beletsky A., Prokofeva M., Bonch-Osmolovskaya, E., Ravin N. and Skryabin K. (2009). The genome sequence of the Crenarchaeon *Acidilobus saccharovorans* Supports a New Order, *Acidilobales*, and suggests an important ecological role in terrestrial acidic hot springs. *Appl. Environ. Microbiol*., **77:** 5652-5657.
153. Hallam S., Mincer T., Schleper C., Preston C., Roberts K., Richardson P. and DeLong E. (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLOS Biol.* **4:** e95.
154. Hallam S., Konstantinidis K., Putnam N., Schleper C., Watanabe Y., Sugahara J., …, DeLong E. (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *PNAS*, **103:** 18296-18301.
155. Walker C., de la Torre J., Klotz M., Urakawa H., Pinel N., Arp D., …, Stahl D. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenArchaea. *Proc. Natl Acad. Sci*., **107:** 8818-8823.
156. Hu¨gler M., Huber H., Stetter K. and Fuchs G. (2003). Autotrophic CO_2 fixation pathways in *Archaea* (*Crenarchaeota*). *Arch. Microbiol*., **179:** 160-173.
157. Alonso-Sa´ez L., Waller A., Mende D., Bakker K., Farnelid H., Yager PL., …, Bertilsson S. (2012). Role for urea in nitrification by polar marine Archaea. *Proc. Natl. Acad. Sci. USA*, **109:** 17989-17994.
158. Podar M., Podar M., Makarova K., Graham D., Wolf Y., Koonin E. and Reysenbach A-L. (2013). Insights into a*rchaeal* evolution and symbiosis from the genomes of a Nanoarchaeon and its inferred Crenarchaeal host from Obsidian Pool, Yellowstone national park. *Biol. Direct.*, **8:** 9.
159. Munson-McGee J., Field E., Bateson M., Rooney C., Stepanauskas R. and Young M. (2015). *Nanoarchaeota*, their *sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone national park hot springs. *Appl. Environ. Microbiol*., **81:** 7860-7868.
160. Wurch L., Giannone R., Belisle B., Swift C., Utturkar S., Hettich R., Reysenbach A-L. and Podar M. (2016). Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. *Nat. Commun.* **7:** 1-10.
161. Pinto O., Silva T., Vizzotto C., Santana R., Lopes F., Silva B., Thompson F. and Kruger R. (2020). Genome-resolved metagenomics analysis provides insights into the ecological role of *Thaumarchaeota* in the Amazon river and its plume. *BMC Microbiol*. **20:**13

162. Bernalier A., Fonty G., Bonnemoy F. and Gouet P. (1992). Degradation and fermentation of cellulose by the rumen anaerobic fungi in axenic cultures or in association with cellulolytic Bacteria. *Curr. Microbiol*., **25:** 143-148.
163. Costa V., Amorim M., Reis E., Quintanilha A. and Moradas-Ferreira P. (1997). Mitochondrial superoxide dismutase is essential for ethanol tolerance of *Saccharomyces cerevisiae* in the post-diauxic phase. *Microbiol*., **143:** 1649-1656.
164. Viigand K., Tammus K. and Alamäe T. (2005). Clustering of MAL genes in *Hansenula polymorpha*: cloning of the maltose permease gene and expression from the divergent intergenic region between the maltose permease and maltase genes. *FEMS Yeast Res*., **5:** 1019-1028.
165. Yurimoto H, Oku M. and Sakai Y. (2011). Yeast methylotrophy: metabolism. *Gene Regul. and Peroxisome Homeost. Int. J. Microbiol*., **2011:** 101298.
166. McHardy AC., Fan L., and Yu X. (2017). Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. *DNA Res*., **24:** 635-648.
167. Ren X., Liu Y., Tan Y., Huang Y., Liu Z. and Jiang X. (2019). Sequencing and functional annotation of the whole genome of *Shiraia bambusicola*. *Genes, Genome, Genetics*.
168. Urquhart A., Mondo S., Mäkelä M., Hane J., Wiebenga A., He G., …, Idnurm A. (2018). Genomic and genetic insights into a cosmopolitan fungus, *Paecilomyces variotii* (*Eurotiales*). *Front. Microbiol*., **9:** 3058.
169. Ohm R., Feau N., Henrissat B., Schoch C., Horwitz B., Barry K., …, Grigoriev I. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLOS Pathog*. **8:** e1003037.
170. Li J., Yu L., Yang J., Dong L., Tian B., Yu Z., …, Zhang K. (2010). New insights into the evolution of subtilisin-like serine protease genes in *Pezizomycotina*. *BMC Evol. Biol*., **10:** 68.
171. Murat C., Payen T., Noel B., Kuo A., Morin E., Chen J., Kohler A. Krizsán K., …, Martin F. (2018). *Pezizomycetes* genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nat. Ecol. Evol.* **2:** 1956-1965.
172. Larriba E., Martín-Nieto J. and Lopez-Llorca L. (2012). Gene cloning, molecular modeling, and phylogenetics of serine protease P32 and serine carboxypeptidase SCP1 from nematophagous fungi *Pochonia rubescens* and *Pochonia chlamydosporia*. *Can. J. Microbiol*., **58:** 815-27.
173. Voigt C., Scha¨fer W. and Salomon S. (2005). A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. *The Plant J.*, **42:** 364-375.
174. Olivieri F., Eugenia Zanetti M., Oliva C., Covarrubias A. and Casalongue C. (2002). Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. *Eur. J. Plant. Pathol*., **108:** 63-72.
175. Agrawa Y., Khatri I., Subramanian S. and Shenoy B. (2015). Genome sequence, comparative analysis and evolutionary insights into chitinases of entomopathogenic fungus, *Hirsutella thompsonii*. *Genome Biol. Evol*., 916-930.
176. Yang Y., Liu B., Du X., Li P., Liang B., Cheng X., …, Wang S. (2014). Complete genome sequence and transcriptomics analyses reveal pigment biosynthesis and regulatory mechanisms in an industrial strain, *Monascus purpureus* YY-1. *Scient. Rep*., **5:** 8331.
177. Galagan J., Calvo S., Borkovich K., Selker E., Read N., Jaffe D., …, Birren B. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nat*., **422:** 859-68.
178. Zheng H., Zhong Z., Shi M., Zhang L., Lin L., Hong Y., …, Wang Z. (2018). Comparative genomic analysis revealed rapid differentiation in the pathogenicity related gene repertoires between *Pyricularia oryzae* and *Pyricularia penniseti* isolated from a *Pennisetum grass*. *BMC Genome*, 1**9:** 927.
179. Zhang N., Cai G., Price D., Crouch J., Gladieux P., Hillman B., …, Bhattacharya D. (2018). Genome wide analysis of the transition to pathogenic lifestyles in *Magnaporthales* fungi. *Scient. Rep*., **8:** 5862.
180. Bugos R., Sutherland J. and Adler J. (1988). Phenolic compound utilization by the soft rot fungus *Lecythophora hoffmannii*. *Appl. Environ. Microbiol*., **54:** 1882-1885.
181. Leonhardt S., Büttner E., Gebauer A., Hofrichter M. and Kellner H. (2018). Draft genome sequence of the *sordariomycete Lecythophora* (Coniochaeta) *hoffmannii* CBS 245.38. *Genome Announc*., **6:** e01510-17.

182. Lindahl B., de Boer W. and Finlay R. (2010). Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. ISME J., **4:** 872-881.
183. Wang Q-M., Theelen B., Groenewald M., Bai F.-Y. and Boekhout T. (2014). *Moniliellomycetes* and *Malasseziomycetes*, two new classes in *Ustilaginomycotina*. *Persoonia*. **33:** 41-47.
184. Sánchez C. (2009) Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnol. Advances,* **27:** 185-194.
185. James T., Letcher P., Longcore J., Mozley-Standridge S. Porter D., Powell M., Griffith G. and Vilgalys R. (2006). A molecular phylogeny of the flagellated fungi (*Chytridiomycota*) and description of a new phylum (*Blastocladiomycota*). *Mycologia,* **98:** 860-871.
186. Bart T. van de Vossenberg L., Warris S., Nguyen H., van Gent-Pelzer M., Joly D., …, van der Lee T. (2019). Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. *Scient. Rep.*, 9:8672.
187. Karov S., Mamanazarova K., Popova O., Aleoshin V., James T., Mamkaeva M., …, Vishnyakov A. (2017). *Monoblepharidomycetes* diversity includes new parasitic and saprotrophic species with highly intronized rDNA. *Fungal Biol.*, **121:** 729-741.
188. Van der Giezen M. (2009). Hydrogenosomes and mitosomes: conservation and evolution of functions. *J. Euk. Microbiol.,* **56:** 221-231.
189. Embley T., van der Giezen M., Horner D., Dyal P. and Foster P. (2002). Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Phil. Trans. R. Soc. Lond. B.* **358**: 191-203.
190. Hibbett D., Binder M., Bischoff J., Blackwell M., Cannon P., Eriksson O., …, Zhang N. (2007). A higher-level phylogenetic classification of the Fungi. *Mycolog. Res*., **111:** 509-547.
191. Hoffmann K., Voigt K. and Kirk P. (2011). *Mortierellomycotina* subphyl. nov., based on multi-gene genealogies. *Mycotaxon,* **115:** 353-363.
192. Wagner L., Stielow B., Hoffmann K., Petkovits T., Papp T., Vágvölgyi C., …, Voigt K. (2013). A comprehensive molecular phylogeny of the *Mortierellales* (*Mortierellomycotina*) based on nuclear ribosomal DNA. *Persoonia,* **30**: 77-93.
193. Cheng Y., Edwards J., Allison G., Zhu W. and Theodorou M. (2009). Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. *Bioresour. Technol.,* **100:** 4821-4828.
194. Nakashimada Y., Srinivasan K., Murakami M. and Nishio N. (2000). Direct conversion of cellulose to methane by anaerobic fungus *Neocallimastix frontalis* and defined methanogens. *Biotechnol. Lett.,* **22:** 223-227.
195. Lee SS, Ha JK, Cheng K-J (2000). Relative contributions of *Bacteria*, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions. *Appl. Environ. Microbiol*., **66:** 3807-3813.
196. Leis S., Dresch P., Peintner U., Fliegerovà K., Sandbichler A., Insam H. and Podmirseg S. (2014). Finding a robust strain for biomethanation: anaerobic fungi (*Neocallimastigomycota*) from the Alpine ibex (Capra ibex) and their associated methanogens. *Anaerobe,* **29:** 34-43.
197. Zhou J., Liu W., Deng Y., Jiang Y-H., Xue K., He Z., …, Wang A. (2013). Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. *MBio.,* **4:** pii: e00584-12.
198. Zhou J., Deng Ye., Zhang P., Xue K., Lianga, Y., Van Nostrand J. …, Arkin A. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *PNAS*, E836-E845.
199. Ofiţeru I., Lunnc M., Curtisa T., Wellsd G., Criddled C., Francise C., and Sloan W. (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *PNAS,* **107:** 15345-15350.