PLOS ONE

Metagenomics survey unravels diversity of biogas microbiomes with potential to enhance productivity in Kenya

Samuel Mwangangi Muturi [1,2]*, Lucy Wangui Muthui [3], Paul Mwangi Njogu[4], Justus Mong'are Onguso[2], Francis Nyamu Wachira[5], Stephen Obol Opiyo[6,7], Roger Pelle [3]

1 Department of Biological Sciences, University of Eldoret, Eldoret, Kenya, 2 Institute for Bioteschnology Research, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya, 3 Biosciences Eastern and Central Africa—International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya, 4 Institute for Energy and Environmental Technology, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya, 5 Department of Life Sciences, South Eastern Kenya University, Kitui, Kenya, 6 OARDC, Molecular and Cellular Imaging Center-Columbus, Ohio State University, Columbus, Ohio, United States of America, 7 The University of Sacread Heart, Gulu, Uganda

* samuelmuturi37@gmail.com

Abstract

The obstacle to optimal utilization of biogas technology is poor understanding of biogas microbiomes 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-diversity studies. We identified 37 phyla, 65 classes and 132 orders. *Bacteria* dominated and comprised 28 phyla, 42 classes and 92 orders, conveying substrate's versatility in the treatments. Though, *Fungi* and *Archaea* comprised 5 phyla, the *Fungi* were richer; suggesting the 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, metabolize organic macromolecules. The identified *Cytophagales*, *Alteromonadales*, *Flavobacteriales*, *Fusobacteriales*, *Deferribacterales*, *Elusimicrobiales*, *Chlamydiales*, *Synergistales* to mention but few, also catabolize macromolecules into smaller substrates to conserve energy. Furthermore, *δ-Proteobacteria*, *Gloeobacteria* and *Clostridia* affiliates syntrophically regulate P_{H2} and reduce 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, comprising 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 variables and its' productivity provided insights into better management strategies that ameliorate biochemical limitations to effective biogas production.

Introduction

The optimal utilization of micro-organisms is a critical strategy in the development of new products and improvement of the existing bioprocesses [1]. One of the widely applied technologies that utilises micro-organisms is anaerobic digestion (AD) of waste 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 power, producing up to 10,000 MW [3]. Apart from power production, stable AD processes provide one of the most efficient environmental models for organic waste valorization that enables nutrient recovery from the digestate [4], and hence providing additional economic benefits [5], when compared to other methods of waste management [6]. Many African countries, including Kenya produce huge amounts of waste [5] with high biogas generating potential. In spite of the efforts made to exploit this resource through the Biogas Partnership Programme [7], the biogas systems in African countries remain fairly rudimentary. The Programmes have installed approximately 60,000 biogas plants in Africa which include 16,419 in Kenya, 13,584 in Ethiopia, and 13,037; 7518, and 6504 in Tanzania, Burkina Faso and Uganda, repectively [5, 7]. Though Kenya leads in biogas exploitation in Africa, adoption of the technology remains poor, and is not commensurate with the country's energy needs. However, because of the relative popularity of the technology in the country, compared to its' counterparts, Kenya was justified as the appropriate pilot site for biogas studies among other African countries. Similar guidelines were also followed to select the Kenyan regions in this study.

Overall, it is thought that the poor adoption of the biogas technology in Africa is partly due to poor knowledge on the microbiomes that mediate the AD process. These processes comprise thermodynamic unfavorable reactions that contribute to stability and performance of the treatments. The microbial populations and the reactions involved are highy sensitive to environmental and reactor operation conditions; traits that prompt the need for development of sustainable management strategies for effective biogas production. Without such strategies, a complete breakdown of the whole system occurs. In the event of complete failure, farmers are forced to abandon the systems or invest in costly restoration processes, all of which end up reducing the economic and environmental benefits of the AD systems; thereby constraining efforts towards greening the economy.

Although, previous reports from other continents have indicated dominance of *Bacterial* and *Archaeal* species in tolerating 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 16S rRNA techniques used thus far provide low resolution [10–12], and are not robust enough to unravel the high level of microbial diversity in complex biogas ecosystems. All these limit the information available on microbial composition and diversity which would be critical for improvement and better management of biogas systems. The 16S rRNA gene does not reveal the full extent of microbial diversity in any ecosystem [13–17] and most current diversity estimates are extrapolations of empirical scaling laws [13], and other theoretical models of biodiversity [14]. The above limitation constrains the capacity to scrutinize and apply the biodiversity scaling laws,

macroecological theories and other ecological theories [12–14] in the context of biogas production. This ends up directly limiting the available microbial diversity knowledge which would be useful to develop viable mitigition strategies to address biochemical reaction failures. Despite covering a small fraction of microbial diversity in a dataset, marker gene based studies, annotate the detected species against biological databases that comprise sequences of interest. This approach has the potential to lead to conflicts on the links between microbial populations and reactor performance. Such conflicts were reported by 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 contrasting views prompted us to ask whether microbial diversity in biogas reactors matters or it is the existence of the core microbiomes that determine the AD stability and maximum biogas production. In an effort to seek answers to these questions, we were provoked to further characterize the biogas system's micro-organisms using next generation sequencing platforms also termed "massively parallel sequencing" over a wide geographic area. Specifically, the technology employed sequencing-by-synthesis chemistry to generate the dataset. Previously, this technology has enabled massive sequencing of environmental genomes with high quality and accurancy [20]. However, the application of the technology in the biogas systems is still at its infancy. Furthermore, eukaryotes, mainly *Fungi* have normally been neglected by the majority of the studies, yet these species, are also crucial in the metabolism of large macromolecules in the AD systems. Our study was therefore designed to address the limitations enumerated above. We expect our findings to contribute to the formation of testable hypotheses for future application of metacommunity theory on the links between microbal diversity and environmental conditions. Further the study will contribute to a greater understanding of biogas microbiomes for improvement of biogas production.

Results

Annotations against the 16 biological databases

We used Illumina based shotgun metagenomics to identify and compare microbial communities within and among the biogas reactor treatments. In total 12 sludge samples assigned to twelve different treatments operating at a steady-state were assayed. The procedure yielded a total of 44 million raw reads and out of these, 2 million reads were of poor quality. We revealed 24,189,921 reads of 35-151bps from the remaining 42 million reads, which were filtered based on the % G:C content bias and duplicates according to Gomez-Alvarez *et al.* [21]. Approximately 23,373,513 reads conveying 1,261833 to 3,986675 reads per the treatment were realized (S1A Fig). Thereafter, we trimmed and assembled the contigs using *de novo* approach and the obtained high quality scaffolds (3,819,927 scaffolds, 1,119,465,586 bps, with mean average length of 291bps, and mean average standard deviation of 410 bps, S1 Table) were annotated against sixteen databases using 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 were the least (<8000 scaffolds). Among the databases, IMG, Genbank and TrEMBL conveyed equal scaffold annotations (144,000 scaffolds). Other databases annotated the reads as indicated in Fig 1. Between the two ontologies, the large lake biome revealed the highest annotation compared to the terrestrial biome with a precision of $P{\leq}0.05$ across the treatments. Furthermore, 39.75% to 46.70% and 53.07% to 60.11% genes were conveyed to known and unknown proteins.
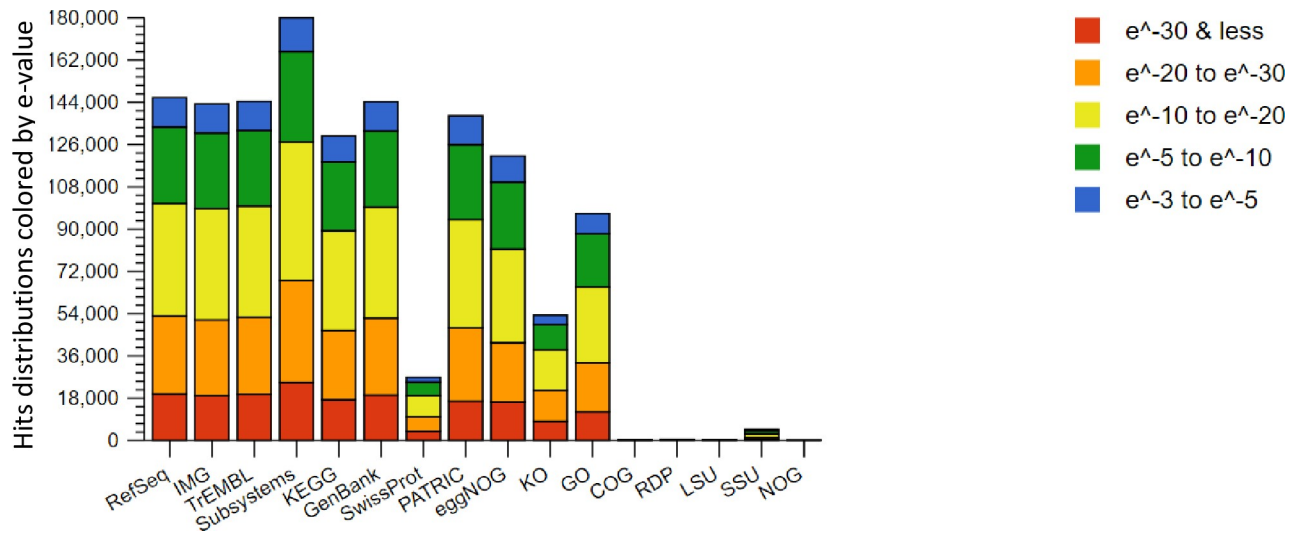
Fig 1. Stacked barchart showing the sources of hits distributions from the sixteen databases. These include nucleotide and protein databases, protein databases with functional hierarchy information, and ribosomal RNA databases. The bars representing annotated reads are colored by e-value range. Different databases have different numbers of hits but can also have different types of annotation data.

https://doi.org/10.1371/journal.pone.0244755.g001

Annotations against the archived taxonomic profiles

Taxonomic assignments revealed four broad but relatively simple distributions of microorganisms, with the species of three domains being responsible for methanogenesis. In summary, 2, 617,791 reads hits were generated, with 213681 hits being assigned to *Archaea*, 2379293 hits to *Bacteria* and 21886 hits to *Eukaryotes* (S1A Fig). Thus, the revealed abundance hits ratio of *Bacteria* to *Archaea* to *Eukaryotes* in the reactor treatments was 109:9:1. The higher abundance of *Bacterial* species was linked to their metabolic capabilities. Unlike other cellular organisms, *Eukaryotes* mainly *Fungi* had the least abundances ($< 1\%$ of the total hits) among the treatments. Despite these variations of species abundances, we observed consistent microbial richness among the studied treatments. In total 37 phyla, 73 classes and 132 orders were indentified. A majority of the reactor treatments exhibited high community diversity, 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 identified *Bacterial* biomes

We evaluated the variation in abundance and composition of 92 *Bacterial* domain orders that were assigned to 41 classes and 21 phyla (S2A Fig, Table 1). 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 (S2B Fig). In addition, we further observed low diversity variabilities between communities of reactor 3 and 6 and those of reactor 8 and 11 when compared at lower (class and order) taxonomic ranks (S3A and S3B Fig). Hence, only four out of twelve treatments had high β-diversities, which could be indicating genetic plasticity due to species evolution [22, 23].

Specifically, we revealed *δ-Proteobacteria* species as the most abundant guild, constituting 9–14% of the total microbial communities. Globally the species were the second most
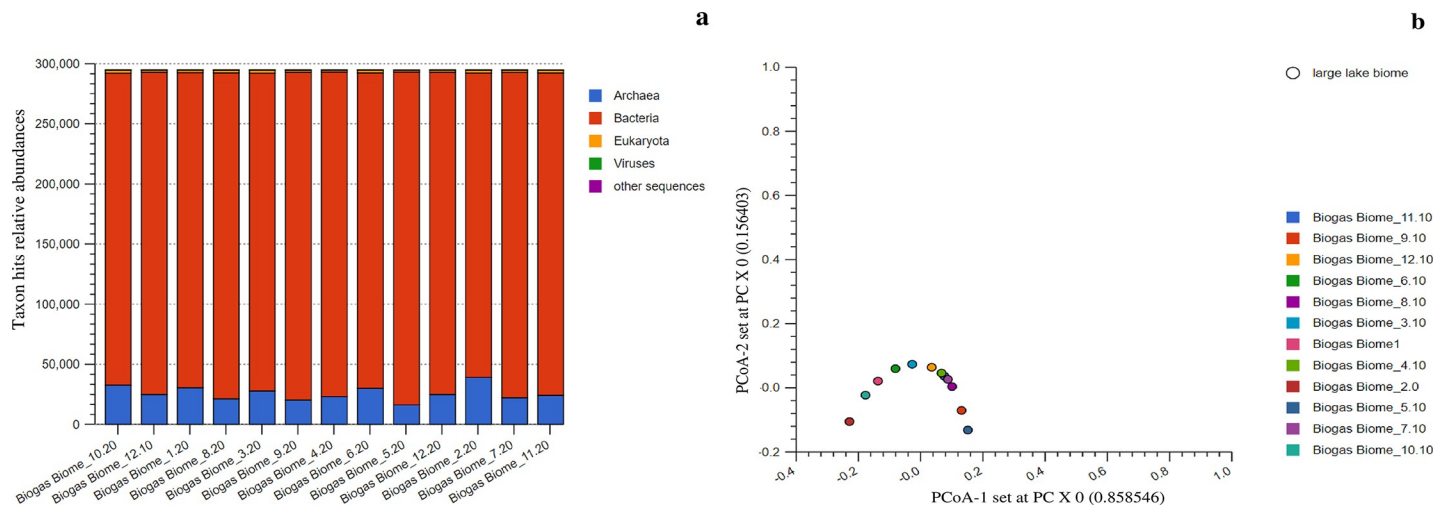
Fig 2. The stacked barchart (a) showing the three main domains, proportion of relative abundances and their PCoA plot (b) based on the Euclidean model. The PCoA plot revealed dissimilarities among the studied reactors. However, only three reactors partially clustered in the upper right quadrant of the plot. Though the Viruses nucleotide contributed to the curve they were not considered in the downstream analysis.

https://doi.org/10.1371/journal.pone.0244755.g002

abundant after the *Clostridial* species that composed 11–26% of the total communities. The δ-*Proteobacteria* communities were classified into 7 orders (S6A Fig). Among them were *Syntrophobacteriales* that comprised 2–4% of the total identified communities, followed by *Desulfuromonadales* and *Myxococcales* species that comprised 2–4% and 0.9–4% of the total microbial compositions respectively. *Desulfobacterales* and *Desulfovibronales* communities were found to comprise 1–2% of the total species abundances. Among the least abundant δ-*Proteobacterial* species were the affiliates of *Desulforculales* and *Bdellovibrionales* that comprised only < 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 (S6B Fig).

Among the identified species, the affiliates of *Firmicutes* were the most abundant in seven out of twelve reactor treatments and comprised 15–34% of total hits (Table 1) while the *Proteobacteria* species (Fig 3) dominated in the other five treatments, comprising of 19–32% of total hits (Table 1). *Proteobacteria* richness was significantly higher compared to that of *Firmicutes* species. The *Proteobacteria* communities were classified into 38 orders, which were affiliated to 7 classes (S4A Fig). We linked their versatility with their ability to circumvent the oxidative stress in the systems. Significantly high levels of similarity of microbial composition were revealed among the treatments 1, 3 and 6, and treatments 4, 7 and 12, all located on the upper and lower right quadrant of the plot, (S4B and S5 Figs; at the class and order level). However, low β-diversities were observed for the identified *Proteobacteria* communities of reactor 4, 7, 8 and 9, (S5 Fig) at the order level. Other treatments were found to have high β-diversities of *Proteobacteria* species.

The communities' affilliated to the class γ-*Proteobacteria* comprised 4–11% of the total identified reads. The affiliates were the fifth most abundant group in the entire microbiome and were classified into 14 orders (S7A Fig). Among the identified orders was *Pseudomonadales*, that dominated the class, comprising of <1 to 6% of the total identified communities. This phylotype was followed by *Alteromonadales* and *Enterobacterales* species in abundance. The members of *Chromatiales*, *Vibrionales*, *Methylococcales*, *Oceanspirillales* and *Pasteurelales* respectively were also among the detected species. Other identified orders were

Table 1. The Krona software generated *Bacterial* phyla biomes of the twelve treatments 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%

S_1: Biogas Biome1; **S_2.0:** Biogas Biome 2.0; **S_3.10:** Biogas Biome 3.10; **S_4.10:** Biogas Biome 4.10; **S_5.10:** Biogas Biome 5.10; **S_6.10:** Biogas Biome 6.10; **S_7.10:** Biogas Biome 7.10; **S_8.10:** Biogas Biome 8.10; **S_9.10:** Biogas Biome 9.10; **S_10.10:** Biogas Biome 10.10; **S_11.10:** Biogas Biome 11.10; and **S_12.10:** Biogas Biome 12.10.

https://doi.org/10.1371/journal.pone.0244755.t001
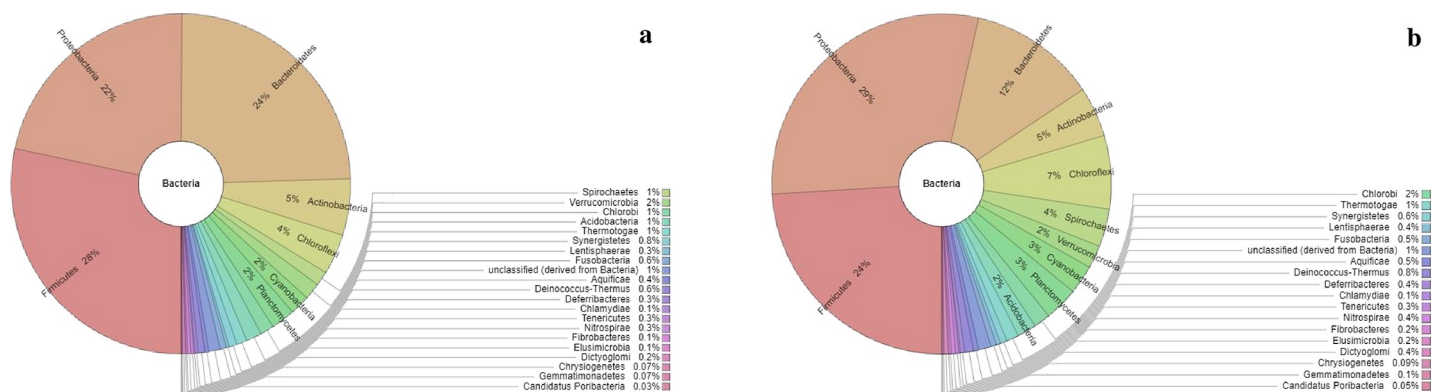
Fig 3. The Krona radial space filling display showing the source evidences of the stated relative abundance of the identified communities in the text (e.g Table 3 and within the text). The display reveals the abundances of the Bacterial reads for the communities identified in reactor 12 and 3: a) Reveals community abundances of reactor 12 while b) reveals community abundances of reactor 3, where *Firmicutes* and *Proteobacteria* dominated.

https://doi.org/10.1371/journal.pone.0244755.g003

Table 2. The geographic positioning systems' co-ordinate for the sampled reactor treatments, assigned to the four blocks of 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

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.

Xanthomonadales, Thiotrichales, Legionellales, Aeromonadales, Acidithiobacillales and *Cardiobacterales*, all comprising <1% of the total identified communities. Unclassified *γ-Proteobacteria* reads were also identified in this study. Our results also revealed significantly high β-diversity ($P \leq 0.05$) of the microbial communities in the following six reactor treatments (reactor 2, 5, 6, 10, 11, 12). In contrast, the communities identified in reactors 1, 3, 4, 7, 8 and 9 were found to cluster on the upper right quadrant of the PCoA plot (S7B Fig), indicating homogeneity and genetic composition similarity within the microbial populations.

Communities of *β-Proteobacteria* were detected in low abundances (2–4% of the identified microbial communities). Affiliates of *Burkholderiales* comprised of 1–2% of the microbial populations while the affiliates of *Rhodocyclales, Neisseriales,* and *Nitrosomonadales* species were found to comprise of <1% of the total identified populations across the treatments. Other identified *β-Proteobacteria* species were assigned to the order *Methylophilales, Gallionellales* and *Hydrogenophilales* which comprised of <1% of the total identified communities in the twelve treatments. High sample to sample variation was also observed for the *β-Proteobacteria* species (S8B Fig).

The affiliates of *α-Proteobacteria* were revealed to comprise of 3–8% of the total microbial communities, and were classified into 7 orders (S9A Fig). The *Rhizobiales* dominated this group, comprising of 1–2% of the total microbial population in the treatments. The relative abundances of *Rhodobacterales, Rhodospirillales,* and *Sphingomonadales* followed (representing <1% of the total population). Other *α-Proteobacteria* orders identified included *Caulobacterales, Rickettsiales* and *Parvularculales*. We also found significantly high dissimilarities among the *α-Proteobacteria* within the studied treatments. Only the *α-Proteobacteria* communities of reactor 3 and 11 were revealed to express low β-diversity (S9B Fig), a trait that we postulate was attributed by similar reactor operating conditions rather than the observed environmental conditions (Tables 2 and 3).

The ε and Z-*Proteobacteria* were among the least abundant *Proteobacteria* phylotypes in our study and were found to comprise of ≤1% of the total identified microbial populations in the treatments. The *ε-Proteobacteria* communities were classified into the orders *Campylobacteriales* and *Nautiliales* (S10A Fig), both comprising of <1% of the total identified populations. We found similarities between the *ε-Proteobacteria* communities of reactor 4 and 8 (clustered)

Table 3. The ArcGIS software and the questionnaire generated environmental variables and phenotypic traits for the sampled treatments within the identified blocks of Kenya.

Sam_id	Alt	S-T	AEZ	AAR	AMT ℃	Substrate source			Family size			Reactor Performance				
						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	16 M^3	14 M^3	2
S_2	1674 M	Ferralsol	UM 2	1100–1400	19.5–18.4	3	-	-	2	1	3	2.4	2.0	8 M^3	7 M^3	1
S_3	1443 M	Nitosol	UM 3	1400–2200	20.6–19.2	11	-	-	3	4	7	2.4	2.0	8 M^3	5 M^3	3
S_4	1364 M	Andosol	UM 2	1220–1500	20.1–19.0	6	-	-	4	2	6	1.8	1.5	6 M^3	4.5 M^3	1.5
S_5	1958 M	Phaeozem	L H 1	1700–2600	17.4–14.9	3	-	-	3	1	4	4.8	4.0	16 M^3	12 M^3	4
S_6	1650 M	Nitosol	UM1	1650–2400	19.2–17.6	2	30	-	2	2	4	3.0	2.5	10 M^3	6 M^3	4
S_7	1987 M	Phaeozem	LH 1	1700–2600	17.4–14.9	5	-	-	5	2	7	4.8	4.0	16 M^3	10 M^3	6
S_8	1306 M	Nitosol	UM 3	1000–1250	20.7–19.6	5	-	6	5	4	9	4.8	4.0	16 M^3	11 M^3	5
S_9	1417 M	Andosol	UM 2	1200–1500	20.1–18.9	2	-	-	3	2	5	3.0	2.5	10 M^3	8 M^3	2
S_10	1589 M	Acrisols	UM 2	1100–1400	19.5–18.4	3	-	-	2	2	4	4.8	4.0	16 M^3	13 M^3	3
S_11	1577 M	Cambisol	UM 3	800–1200	19.9–19.5	14	-	-	5	4	9	2.4	2.0	8 M^3	6 M^3	2
S_12	1539 M	Ferralsol	UM1	1400–1700	19.3–17.5	3	-	-	4	3	7	3.0	2.5	10 M^3	8 M^3	2

Sam_id: Sample id; **Alt:** Altitude (M a.m.s.l); **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; **AAR**: Annual Average rainfall (mm); **AMT**: Annual mean Temperature (^{0}C); **CO**: Cow; **PO**: Poultry; **PI**: Pigs; **M**: Male; **F**: Female; **T**: Total; **IG**: Ideal gas (M^3); **ACG**: Actual calculated gas (M^3); **RV**: Reactor Volume (M^3); **FACoG**: Farmers Actual Collected gas (M^3); **DGV:** Difference between the RV and FACoG.

and those of reactor 5 and 9 (clustered partially) revealing low β-diversities (S10B Fig). All the identified *Z-Proteobacteria* were from the order *Mariprofundales*, and accounted for <1% of the total microbial populations in the studied treatments.

The *Firmicutes* communities were the second most diverse community identified in this study. They comprised of 15–34% of the total microbial communities identified. The community was classified into 8 orders and 4 classes (S11A and S12A Figs). The *Firmicutes* in the three treatments of reactor 2, 10 and 12 revealed significant genetic similarity and were revealed to cluster in the lower left quadrant of the PCoA plot (S11B Fig). The *Firmicutes* in reactor 1 and 7 were found to cluster partially on the upper right quadrant of the PCoA plot (significant threshold; *P≤0.05*; S11B Fig), 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 also exhibited low β-diversities (S12B Fig). Other local communities were revealed to express high β-diversities at class and order levels.

The *Clostridia* was revealed to be one of the most abundant *Firmicutes* phylotype, comprising 8–22% of the identified total microbial communities. It was classified into 4 orders (S13A Fig). The order *Clostridiales* dominanted the phylum and constituted 8–22% of the total communities. *Thermoanaerobiales* species constituted 2–3% of the total identified communities. Other identified *Clostridia* species were the affiliates of *Halanaerobiales* and *Natranaerobiales*, all representing <1% of the total microbial populations, except for the *Halanaerobiales* communities in reactor 11 that comprised of 2% of the total identified microbial populations. Among the twelve biogas reactor treatments, three treatments (i.e reactor 7, 11 and 12) clustered on the lower left quadrant of the PCoA plot (S13B Fig), indicating high community similarities. Communities of reactor 5 and 9 were found to cluster partially on the upper left quandrant of the PCoA plot, revealing significantly low β-diversity. The other seven treatments had high β-diversities, implying metabolic and genetic plasticity [22, 23] among the identified *Clostridia* species. The class *Bacilli* comprised of 4–9% of the total identified communities and was the fourth most abundant *Clostridia*. The *Bacilli* were classified into *Bacillales* and

Lactobacillales (S14A Fig), which comprised of 3–4% and 1–4% of the total identified microbiome communities, respectively. Among the *Bacilli* communities, those of reactor 5 and 8 were found to exhibit significant dissimilarity, while the other treatments exhibited low β-diversities (S14B Fig). Other detected low abundance *Firmicutes* were *Erysipelotrichales* (*Erysipelotrichi*) and *Selenomonadales* (*Negativicutes*) which comprised of <1% of the total identified communities. The *Erysipelotrichales* species in reactor 5 and 9 were however found to comprise of 1% of the total identified microbial populations.

*Bacteroidete*s were the third most abundant phylotype comprising of 10–22% of the total identified microbial populations. These were classified into 4 classes and orders (S15A and S16A Figs). Among them, were the *Bacteroidia* populations that comprised of 5–12% of the identified microbial communities. All the identified *Bacteroidia* species belonged to the order *Bacteroidales*. Other identified members of *Bacteroidete*s were *Flavobacteria* that belonged to the order *Flavobacterales*. The communities of *Cytophagia* were all assigned to *Cytophagales* species and they comprised of 1–2% of the total identified microbial communities. The other *Bacteroidete*s were *Sphingobacteria*, mainly of the order *Sphingobacteriales*, which comprised of 2–3% of the total identified microbial communities. We observed partial clustering of the *Bacteroidetes*' communities particularly in reactor 2 and 10 and in reactor 8 and 12, which indicated low β-diversity between the respective communities (S15B and S16B Figs). The *Bacteroidete*s communities of reactor 4 and 9 also clustered on the lower left quandrant of the PCoA plot (S15B and S16B Figs), implying significant similarities of the *Bacteroidete*s in the two treatments.

We identified 6 orders of *Actinobacteria* species (S17A Fig) which comprised of 3–7% of the identified microbial reactor communities. The *Actinomycetales*' were among the identified affiliates, comprising of 2–6% of the total identified communities. The species of order *Coriobacteriales* and *Bifidobacteriales* accounted for <1% of the total identified microbial composition. Other less abundant *Actinobacteria* species were assigned to the order *Rubrobacteriales*, *Solirubrobacteriales* and *Acidimicrobiales*. Out of the twelve biogas reactor treatments, the microbial communities of the seven treatments (reactor 1, 3 and 12; reactor 5 and 11 and reactor 7 and 8; S17B Fig) were found to exhibit significantly ($P{\leq}0.05$) low β-diversities. Only *Actinobacteria species* of the remaining three out of the twelve treatments were found to exhibit high β-diversities.

The *Chloroflexi* were categorized into 4 classes (S18A Fig) and 5 orders (S19A Fig) and comprised of 2–7% of the total microbial communities identified (Table 1). The communities belonged to *Chloroflexi* (1–4% of the identified communities) and *Dehalococcoidetes* ($\leq$1–3% of the identified 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 the class *Chloroflexi* (S20 Fig). The two orders comprised of 1–2% and <1% of the total identified microbiomes, respectively. Other identified classes of the phylum were *Thermomicrobia* and *Ktedonobacteria*; all of which comprised of <1% of the total identified communities. The *Thermomicrobia* were classified into orders *Sphaerobacterales* and *Thermomicrobiales* (S21A Fig). All the identified *Ktedonobacteria* belonged to the order *Ktedonoacteriales* (also representing <1% of the identified communities). We identified two treatments (reactor 1 and 7) that revealed low β-diversity of their local *Chloroflexi* communities (S18B Fig). Four treatments (reactor 4, 5, 10 and 12), out of the remaining ten were revealed to also cluster on the lower right quandrant of the PCoA plot (S18B and S18B Fig).

Similarly, the *Chloroflexi* communities of reactors 1 and 5 and those of reactors 3 and 7 were found to cluster on the upper right and lower right quandrants of the PCoA plot (S20B Fig). The *Chloroflexi* class communities of reactor 1 and 10 were found to cluster partially with the cluster of microbes from reactor 3 and 7. We also identified partial clustering of *Chloroflexi*

from reactor 8, 11 and 12 indicating low β-diversities among the communities in these reactors (S20B Fig). The *Chloroflexi* from the rest of the treatments revealed high β-diversities. In contrast, the *Thermomicrobia* communities were significantly dissimilar among the treatments as indicated in S21B Fig.

The *Cyanobacteria* comprised of 1–3% of the total microbial composition and majority in this group were unclassified. Only <1% of the total population were assigned to the class *Gloeobacteria* (S22A Fig). Unlike other phylotypes, only the *Cyanobacteria* communities of reactors 3, 7 and 10 were found to exhibit significantly high β-diversities ($P \leq 0.05$), (S22B Fig) 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) included members of the following orders; *Chroococcales*, *Nostocales*, *Oscillatoriales* and *Prochlorales* (S23A Fig). Overall, majority of the local *Cyanobacteria* communities were found to reveal low β-diversities. The *Cyanobacteria* communities in reactors 4, 6, 9 and 12, however exhibited high β-diversities (S23B Fig). Similarly the identified four phylotypes at the order level (derived from unclassified reads, S24A Fig) were also revealed to express high local compositional dissimilarities among the twelve biogas treatments (S24B Fig).

We also identified *Acidobacteria* which comprised of 1–2% of the total identified microbial communities. Majority of these populations were from the order *Acidobacteriales* (comprising <1–2% of the total identified microbial communities). The remaining species were affiliated to *Solibacteres*, particularly of the order *Solibacterales* (<1% of the total identified microbial communities) (S25A and S26A Figs). The *Acidobacteria* communities of five (reactors 1, 2, 3, 8, 12) out of twelve treatments were revealed to exhibit low β-diversities. *Acidobacteria* communities of reactor 2 and 8 clustered on the upper left quadrant of the PCoA plot (at order level; S25B and S26B Figs), indicating similarity between the two local communities. *Deinococcus-Thermus*, another phylum, had the orders *Deinococcales* and *Thermales* species (S27A Fig), the two belonging to the class *Deinococci*. The relative abundances of the two orders were ≤1% of the total identified microbial populations. We observed low β-divesity between the *Deinococci* species of reactor 11 and 12 while the diversities of its communities in reactors 3 and 6 and those of reactors 7 and 9 were found to be significantly similar (S27B Fig). All other local *Deinococcus-Thermus* communities expressed high β-diversities.

The identified *Verrucomicrobia* communities were classified into 3 classes that comprised of 1–2% of the total identified microbial communities. Among the identified classes were *Opitutae*, *Verrucomicrobiae* and *Spartobacteria* (S28A Fig). The affiliates of *Verrucomicrobiales* were the most abundant followed by those of the *Puniceicoccales* and *methylacidiphilales* species; all representing <1% of the total identified communities (S29B Fig). Notably, we found out that all the *Opitutae* and *Spartobacteria* reads were unclassified at a lower taxonomic level. Only the *Verrucomicrobia* communities of reactors 3 and 7 were found to cluster partially on the lower right quadrant of the PCoA plot, (at the class level, S28B Fig). In contrast, several local communities including the ones identified in reactors 1 and 10, and reactors 3, 6, 7 and 11 were found to cluster partially (indicative of low β-divesities) within the plots (at order level; S29B Fig). However, their communities in reactor 5 and 8 were found to cluster significantly on the upper right quadrant of the PCoA plot (S29B Fig), implying high similarity between the *Verrucomicrobia* species in the two treatments. Other treatments were found to reveal high β-diversities. Among the identified *Tenericutes*, were *Acholeplasmatales*, *Mycoplasmatales* and *Entomoplasmatales*, all comprising of <1% of the total identified microbial communities and belonged to the class *Mollicutes* (S30A Fig). Interesting, member species of this class were obsolutely absent in reactor 5 only, contrary to other phyla species, probably due to natural selection pressure. Our data also revealed significantly high similarities of the identified

Tenericutes species inhabiting reactors 4 and 9 and reactors 6 and 8 (S30B Fig). The local communities of the other treatments were observed to be significantly dissimilar.

Our results also identified *Planctomycetacia* affiliates which comprised of 1–3% of the total identified microbial communities. These affiliates were all assigned to the order *Plantomyceta-cetales* (*Planctomycetacia*). *Chlorobia* were identified in our study (comprising of 1–2% of the total identified communities). These belonged to the order *Chlorobiales*. *Spirochaetes* were also identified (<1–4% of the total identified microbial composition). All the *Spirochaetes* identified in this study belonged to the order *Spirochaetales*. *Thermotogae* of the order *Thermotogales* were also identified in the study. Other less abundant communities identified were *Synergistales* (*Synergistates*, *Synergistia*), *Lentisphaerales* (*Lentisphaerae*, *Lentisphaerae*), *Fusobacterales* (*FusoBacteria*, *FusoBacteria*), and *Aquificales* (*Aquificae*, *Aquificae*), all of which comprised of <1% of the total identified communities. *DeferriBacteriales*, *Chlamydiales*, *Chrysiogenales*, *Nitrospirales*, *Fibrobacteriales*, *Elusimicrobiales*, *Dictyoglomales*, *Gemmatimonadales* were other groups of microbes identified in this study. The *Candidatus poribacteria* affiliates were not accurately classified at the lower taxonomic ranks (Table 1).

The identified *Archaeal* biomes

We analyzed the identified *Archaea* domain and established that their species comprised of 6–12% of the total microbial populations. Out of the identified communities, we obtained 5 phyla (S31A Fig), 9 classes (S32A Fig) and 16 orders (S33A Fig). Among the communities, those identified in reactor 4 and 12 were found to cluster on the lower left quadrant of the PCoA plot, indicating their high similarity. We also revealed low β-diversities of the identified *Archaeal* communities, of reactor 3 and 6 and also those detected in reactor 7 and 9 (Phylum level; S31B Fig). However, at a lower taxonomic level, only those identified in reactor 3 and 7 were observed to have low β-diversity (Class and order level; S32B and S33B Figs). All the other reactor treatments were observed to contain highly dissimilar communities of *Archaea* at different taxonomic ranks.

A majority of *Archaea* species identified in this study were classified as *Euryarchaeota* and comprised of 5–11% of the total microbial communities. The *Euryarchaeota* communities were classified into 8 classes (S34A Fig) and 10 orders (S35A Fig). Interestingly, we observed significant dissimilarities (high β-diversities) of *Euryarchaeotes* among the reactor treatments at the two taxonomic levels of class and order (S34B and S35B Figs). The class *Methanomicrobia* was the most abundant *Euryarchaeota* phylotype and comprised of 4–9% of the total identified microbial populations. Further, we established that the majority of the members of the class were affiliates of *Methanomicrobiales* order (representing 2–4% of the identified total communities), followed by those assigned to *Methanosarcinales* and *Methanocellales* (S36A Fig), which comprised of 1–3% and ≤1% of the identified total microbial communities, respectively. The local composition of *Methanomicrobia* species were highly dissimilar among the reactor treatments except for those from reactor 1 and 7 that were found to cluster. The communities of the two treatments revealed significant similarity (significant threshold, $P\leq$ 0.05; S36B Fig). 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 comprise of <1% of the detected microbial communities.

Other non-methanogenic *Euryarchaeotes* identified in our study were the affiliates of *Thermococcales* (*Thermococci*), *Archaeglobales* (*Archaeglobus*), *Halobacterales* (*Halobacteria*), and *Thermoplasmatales* (*Thermoplasmata*). Among these phylotypes of *Euryarchaeotes*, affiliates of *Thermococcales* were dominant while the *Thermoplasmatales* were the least abundant.

Crenarchaeota and *Thaurmarchaeota* phyla were among the other identified non-methano-genic species. Between the two phyla, *Crenarchaeotes* were the most diverse and were classified into 4 orders (S37A Fig). Within the identified *Crenarchaeota*, the following orders were iden-tified; *Desulfurococcales*, *Thermoproteales*, *Sulfolobales* and *Acidilobales* (in order of abun-dance). All these orders are members of *Thermoprotei* class which comprised of <1% of the identified microbial communities. We noted that half of the studied treatments (reactor 6 and 9, reactor 7 and 10 and reactor 2 and 11) had *Crenarchaeota* communities that had low β-diversities, while the rest of the communities, in the other six treatments had significantly high β-diversities (significant threshold, $P \leq 0.05$; S37B Fig). *Thaurmarchaeota*, another non-metha-nogenic phylum was classified into *Nitrosopumilales* and *Cenarchaeales* orders (S38A Fig) both of which exhibited high β-diversities (S38B Fig). Other identified phyla were *Korarch-aeota* and *Nanoarchaeota*. The classification of these two phyla at the lower taxonomic ranks was inaccurate and the sequence reads from the two were therefore treated as unclassified.

The identified *Fungal* biomes

Our study also identified the presence of *Fungal* species. The *Fungal* communities were classi-fied into 5 phyla (S39A Fig), 14 classes (S40A Fig) and 23 orders (S41A Fig). Their relative abundances were however low and they comprised of <1% of the total microbial populations among the reactor treatments. Significantly high β-diversities were noted for the identified local communities of *Fungi* at the phyla and class levels (S39B and S40B Figs). However, at the lower taxonomic levels, low β-diversities were the common feature particularly for the com-munities identified in reactor 2 and 10, reactor 3 and 6, reactor 4 and 9, and reactor 8 and 11 (S41B Fig).

The affiliates of *Ascomycota* were the most abundant guild (representing <1 of the total communities) within the *Eukaryotes* and were classified into 7 *Ascomycota* classes (S42A Fig) that included *Eurotiomycetes*, *Saccharomycetes* and *Sordariomycetes*, respectively. Other identi-fied classes were *Schizosaccharomycetes*, *Dothideomycetes*, *Leotiomycetes* and *Pezizomycetes* all of which comprised of <1% of the total microbial communities in the respective treatments. We also further classified the identified *Ascomycota* species into 11 orders (S43A Fig). Interest-ingly we observed significantly low β-diversity between *Ascomycota* communities in reactor 7 and 8. Other identified *Ascomycota* communities were observed to exhibit high β-diversities at the lower taxonomic levels (class and order level; S42B and S43B Figs). Among the classes, *Sor-dariomycetes* were the richest within the phylum and their species were classified into 4 orders; the *Hypocreales*, *Sordariales*, *Magnaporthales* and *Phyllachorales* (S44A Fig). The *Sordariomy-cetes*' communities were found to exhibit significantly high β-diversities (Statistical threshold, $P \leq 0.05$) (S44B Fig) among the studied treatments. Despite their higher abundance compared to *Sordariomycetes* phylotypes, the *Eurotiomycetes*' another class of *Ascomycota* were classified into two orders; *Eurotiales* and *Onygenales* (S45A Fig). The two *Eurotiomycetes* orders were shown to comprise of <1% of the identified total microbial populations. Like other fungal phy-lotypes, only the *Eurotiomycetes*' communities of reactor 1 and 3 were found to cluster on the PCoA plot, which was indicative of high community similarity. The other reactor communities were revealed to have significantly dissimilar *Eurotiomycetes* populations (significant thresh-old, $P \leq 0.05$; S45B Fig). Other identified fungal species were the affiliates of *Saccharomycetales*, *Schizosaccharomycetales*, *Pleosporales*, *Helotiales* and *Pezizales* phylotypes, all comprising <1% of the total microbial communities.

Among the identified *Fungi*, *Basidiomycota* species were the second most abundant (repre-senting <1% of the total communties). The species were classified into 4 classes and 5 orders (S46A and S47A Figs, respectively). Among the classes, *Agaricomycetes* dominated the

treatments, followed by *Tremellomycetes*, *Ustilaginomycetes*, and *Exobasidiomycetes* species (S46A Fig) respectively. The affiliates of the class *Agaricomycetes* were further classified as *Agaricules* and *Polyporales* species (S48A Fig). Other *Basidiomycota* species were affiliated to the following orders; *Tremellales* (*Tremellomycetes*), *Ustilaginales* (*Ustilaginomycetes*) and *Malasseziales* (*Exobasidiomycetes*). We observed two treatments (reactor 11 and 12) that contained *Agaricomycete*'s communities with low β-diversity (S48B Fig) while the local *Basidiomycota* communities in the other reactor treatments had significant, high β-diversities (S46B and S47B Figs).

The *Chytridiomycotes* were among the identified rare species. This was classified into *Chytridiomycetes* (more abundant) and *Monoblepharidomycetes* (S49 Fig). Among the *Chytridiomycetes* were *Spizellomycetales* (identified in reactor 9 and 10), *Rhizophydiales* and *Cladochytriales*, the latter two dectected only in reactor 9 (S50 Fig). All the affiliates of *Monoblepharidomycetes* were classified as *Monoblepharidales* species and were also detected in reactor 9 only. Other rare phylotypes identified in our study included *Blastocladiales*, (*Blastocladiomycetes*, *Blastocladiomycota*, detected in reactor 9), *Entomophthorales*, *Mortierellales*, and *Mucorales*, all derived from unclassified *Fungal* reads (also identified in reactor 9, 10 and 11 respectively, S51 Fig). *Microsporidia* species were not accurately classified at the lower taxonomic levels and were regarded as unclassified though they comprised of <1% of the total microbial populations.

The microbiomes α-diversity and biogas production

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 representative of this maximum. To answer these questions, we conducted a rarefaction analysis through statistical resampling and plotted the total number of distinct species annotations against 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). This finding indicated that 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 947–1116 genera across the treatments, comfirming our maximum sampling effort was attained for further diversity studies. The study also revealed variabilities of reactor performance i.e biogas productivity (Table 3) among treatments, the phenotype that we linked to the observed α-diversity variability, among the sites.

The microbiomes β-diversity and biogas production

To understand the influence of environmental variations and other variables on the local microbial communities and thus biogas production, we conducted comparative community diversity studies on the twelve biogas reactor treatments. 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 (S52A Fig), 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 the lower right or upper left quadrants of the PCoA plot (S52B and S52C Fig). We also observed low β-diversity between the communities of reactor 4 and 9 at the lower taxonomic levels (phylum, class and order) with species clustering partially on the lower left and upper right quadrants of the PCoA (S52B and S52C Fig), respectively. Other treatments that revealed low β-diversity
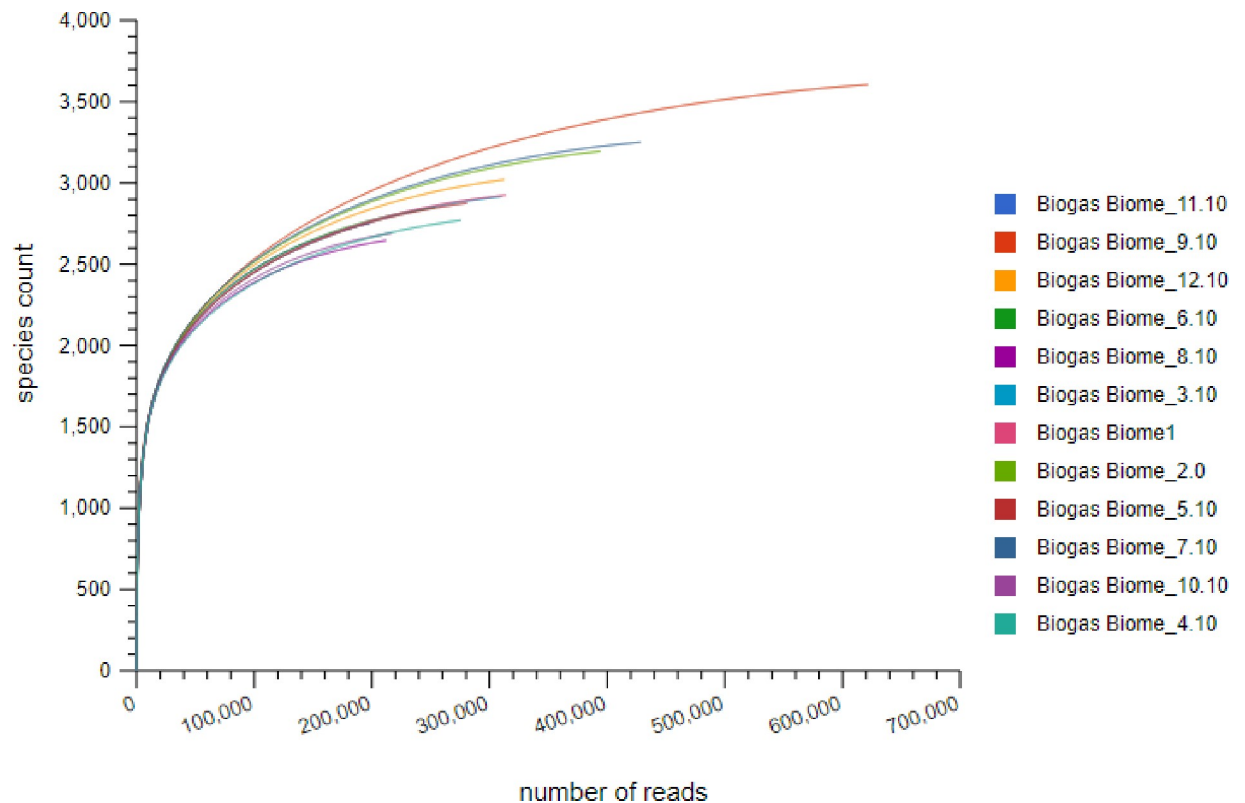
Fig 4. Rarefaction plot showing a curve of annotated species richness. The curve is a plot of the total number of distinct species annotations as a function of the number of sequences sampled.

between their communities were reactor 8 and 11 and reactor 3 and 6. Communities from these reactors were found to cluster partially in the lower right and lower left quadrants of the plot (Fig 5). Similarly, we also noted significant variability of biogas production among the treatments (Table 3), an indication of the observed β-diversities at different taxonomic levels and within the taxa (stated above).

Discussion

Annotations against the 16 biological databases

The aim of our study was to use next generation sequencing technologies and different bioinformatics tools to investigate the biogas microbiomes from 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 read hits annotations, the findings that corroborated with the previous reports [24]. Unlike other databases, SEED integrated several evidences that provided annotation benefits over the other 15 utilized databases. PATRIC database had fewer annotations 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 manual 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
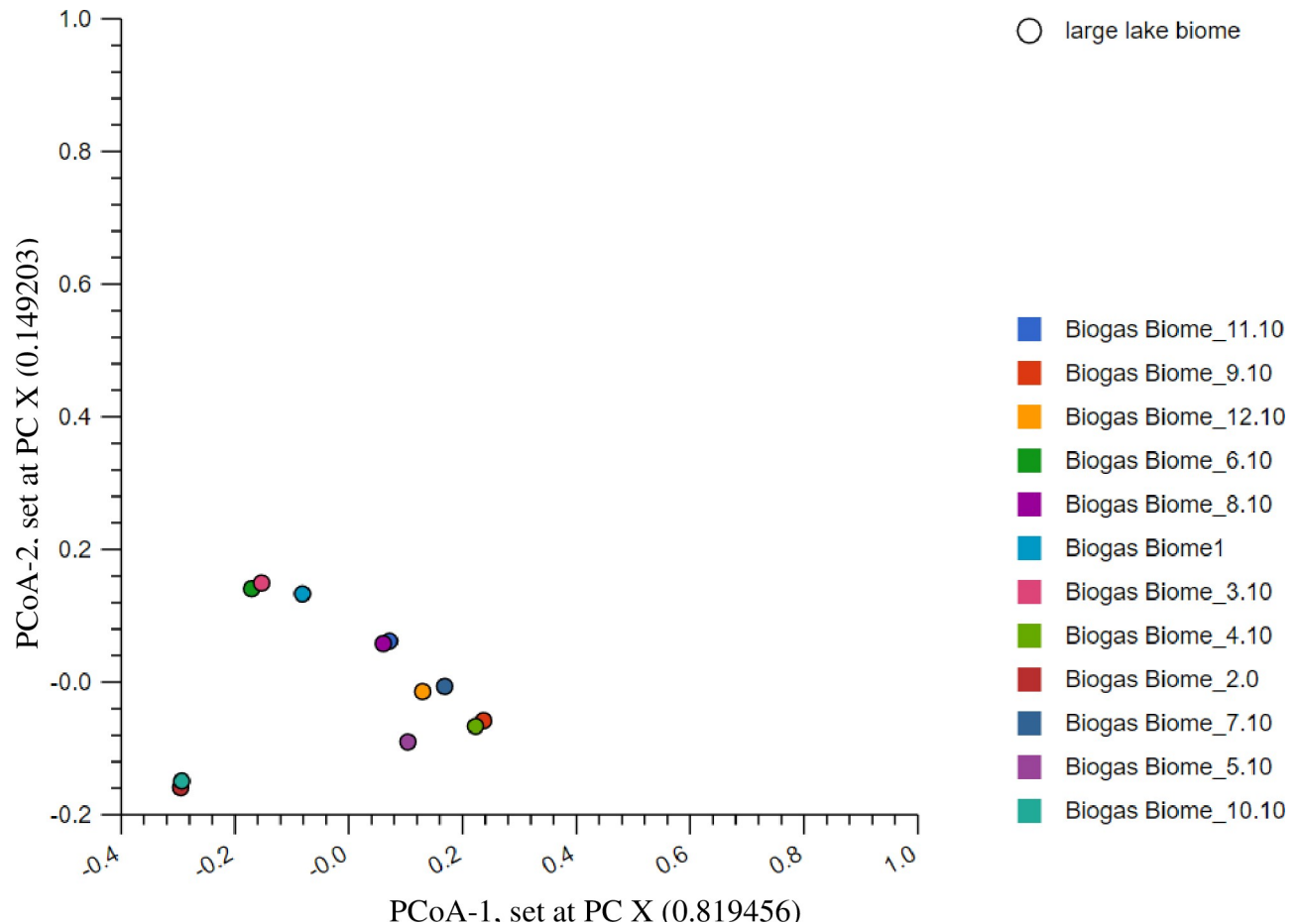
Fig 5. The PCoA analysis revealing β-diversities of the twelve treatments at the order level. The reads compositions in reactor 2 and 10 were partially similar, located in the lower left quadrant, those in reactor 8 and 11 located in the upper right quadrant, reactor 4 and 9 in the lower right quadrant, while those in reactor 3 and 6 were located in the lower left quadrant of the plot.

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 coverage of 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 reconstruct 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 sequence databases. However, it's 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.

Annotations against the archived taxonomic profiles

Our taxonomic annotations approach generated more detection, compared to the other annotations conducted elsewhere [25]. 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.* [26]. Furthermore, while most of the studies used 19–35% of the total generated sequences [25], our approach annotated and used 39.75–46.70% of our total generated sequences. All these findings indicate that 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 [27]. Among the domains, *Bacterial* species dominated in abundance, comprising 87–93% of the total microbial populations, followed by *Archaea* and *Eukaryotes*, while *Viruses* comprised the least abundant population. The observed high *Bacterial* abundances, suggest their crucial metabolic roles in biomass conversion and other reactions within the reactor systems [27]. As expected, we found low abundance among the *Eukaryotes* (*Fungi*) which was indicative of the presence of low recalcitract biomasses concentration. Generally, the *Fungal* species are involved in several biochemical reactions of biogas production including hydrolysis, fermentation and syntrophic reactions. The identified affiliates of *Archaea* were majorly consumers of smaller substrates that were generated by the *Bacterial* and *Fungal* species. The *Archaea* species are able to use different methanogenic routes to convert the substrates into $CH_{4(g)}$. Nevertheless, the main roles of the identified and less abundant group of the virus were unclear, though the species could have been active in degrading other microbial cells in the AD systems. However, their detection was largely attributed to the viral pathogenicity on both *Bacteria* and *Archaea* species. High β-diversities among the local communities (at the domain level) were an indication of environmental influences at the different ecological location.

The identified *Bacterial* biomes

More than 90 representative orders of the *Bacteria* domain were revealed in our datasets. The first step of the AD process involves hydrolytic reactions that convert large macromolecules into smaller substrates [28]. Several *Bacterial* communities are capable of hydrolysis, even on high lignocellulosic plant biomasses. The identified species that participate in these reactions were annotated to the *Firmicutes*, *Bacteroidetes*, *Chloroflexi* to mention but a few and their species are known to harbor cellulosomes, that secrete extracellular enzymes, to degrade the substrate. Previous reports have indicated that species belonging to *Bacteroidales*, *Bifidobacteriales*, *Clostridiales Lactobacillales*, *Thermotogales* and many other identified *Bacterial* species, catalyze the first step of AD [28]. The step involves rate limiting reactions, largely regulated by the substrate composition. The nature of the chemical composition of the substrate, generally depends on the environmental variables (Tables 2 and 3), which influence

microbial species abundances and β-diversity differently. All these affects reactor stability and performance. As expected, the class *Clostridia*, that comprised *Clostridiales*, *Thermoanaerobacteriales* and *Halanaerobiales* orders, was the most dominant among the identified phylotypes. The guild converts large macromolecules into organic acids, alcohols, $CO_{2(g)}$ and $H_{2(g)}$ [29–35]. In this case, the species use a substrate-level phosphorylation mechanism to conserve the energy. Acetogenic and homoacetogenic species belonging to the same guild such as *Moorella thermoacetica* and *Thermoanaerobacter kivui* are able to use the reductive acetyl-CoA pathway to reduce $CO_{2(g)}$ into $CO_{(g)}$, concomitantly generating reducing equivalents [36, 37] and vice versa. As a consequence, the acetogens and homoacetogens use proton or sodium dependent mechanisms to conserve their energy [37]. Above all, the *Clostridia* guilds are able to catalyze other biochemical reactions such as reductive dechlorination of various compounds that leads to production of sulfide from either thiosulfate or sulfite. The reactions consume pyruvate and lactate as the carbon source. In this case, the member species are able to use [NiFe]-hydrogenases to generate $H_{2(g)}$ [35]. In our systems, we also identified *Halanaerobiales* species that have a capacity to tolerate high salt concentrations [38]. Generally the reactor treatments comprised high salts concentrations (treated cow dug) and the *Halanaerobiales* must have regulated the toxic cations from the ionized salts. All these roles of *Clostridia* suggest their metabolic versatility, and these 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 roles of *Clostridia*. However, the observed low β-diversities (S13B Fig) of the *Clostridia* communities in reactor 7, 11 and 12 were attributed to similar treatment perturbations and the identified environmental variables (Tables 2 and 3).

The *Bacillales* and *Lactobacillales* species identified in our study, all belonged to the class *Bacilli*. These species are able to use mixed fermentation routes to metabolize carbohydrates and other biological macromolecules [39, 40]. Formate dehydrogenase complex activities are the main mechanisms of energy conservation [40] in these microbial guilds. Generally, *Bacilli* co-exist in symbiotic relationship with *Clostridia* species in lactate and acetate metabolism [41], a mechanism that is termed as lactate cross-feeding. High β-diversities among the observed treatments (S13B Fig) were solely attributed to cow-dung's substrate composition and environmental variables (Tables 2 and 3). However, the low β-diversities observed between or among the treatments (S14B Fig) that formed the clusters were largely attributed to similarities of reactor's operation conditions. All these implied that the respective organisms were in the same succession stage at the time of sampling. The *Erysipelotrichi* another identified *Firmicutes* class, utilize the syntrophic acetogenesis route to metabolize lipids into acetate, $H_{2(g)}$ and $CO_{2(g)}$ [25]. Its species, together with other *Firmicutes* species like *Smithella propionica* use non-randomizing pathway [42] to metabolize propionate into acetate. Among other identified *Firmicutes*, were *Syntrophomonas wolfei* (*Clostridiales*, *Clostridia*) that use 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 $H_{2(g)}$ [43, 44]. Notably, these syntrophic species are consumers of 30% of the generated electrons, which bypasses the rate limiting steps of the volatile fatty acids accumulation. All these reactions, contributes to reactor stabilities. In general, we associated the observed low *Firmicutes* abundances and diversities (richness, S11A and S12A Figs) with genetic versitility that conferred capacity to encode an assortment of proteins; which metabolize a variety of biomasses into biogas. Furthermore, the observed high β-diversities (S11B and S12B Figs) of the *Firmicutes* support our concepts on the genetic divergence among the species and we postulate that the trait was mainly influenced by the size of the treatment and the variability of the environmental conditions (Tables 2 and 3).

The *Synergistete*s species identified in this study have been established to ferment amino acids into lactate, acetate, $H_{2(g)}$ and $CO_{2(g)}$ [45, 46]; though they have no ability to metabolize carbohydrates and fatty acids. One such identified example is *Dethiosulfovibrio peptidovorans*. The affiliates of genus *Anaerobaculum* which belong to *Synergistete*s catabolize peptides into acetate and short-chain fatty acids while their close relatives, *Aminobacterium*, are capable of fermenting amino acids into acetate, propionate and $H_{2(g)}$ [45, 46]. Other amino acid metabolizers identified in this study were the affiliates of *Fusobacteria* [39, 47] which are able to catabolize 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 are known to use substrate-level phosphorylation to conserve their energy [48] from amino acids metabolism. Our findings contradict the metabolism of the amino acid by the *Clostridia* species, whose energy conservation mechanisms remain unclear and have been postulated to occur through chemiosmosis [32]. However, the two phyla described here have been found to generate energy through syntrophic association with methanogens; otherwise the generated protons would increase the systems' alkalinity and lead to inhibited steady-state of the biochemical reactions. Based on the low abundance of the two phyla and richness among the treatments, we postulate low amino acid affinities among their identified species, which would suggest another selection pressure in the AD systems [32].

The identified affiliates of *Bacteroidales*, *Flavobacterales*, *Cytophagales* and *Sphigobacteriales*, all belong to *Bacteroidetes*, and are majorly known to ferment carbohydrates and proteins into mixed products (acetate, lactate, ethanol, and volatile organic acids), concomitantly releasing $H_{2(g)}$ [49]. Moreover, their species are also capable of metabolizing aliphatic, aromatic and chloronated compounds [50] and were observed to co-exist with methanogens possibly to increase energy extraction from indigestible plant materials. The member species were highly varied among the treatments and we hypothesize that the cattle breeds and the operating conditions in the treatments were the key drivers of the observed β-diversities. Interestingly, we also observed consistency in the relative abundance (2% of total reads) of the *Cytophagale*s guild among the sites, which was attributable to long periods of low nutrient limitation [51]. Although, the metabolic activity of the identified *Chlorobi* overlap with that of *Bacteroidetes* [52], their species use reverse tri-carboxylic acid (rTCA) cycle to fix $CO_{2(g)}$ and $N_{2(g)}$ in anaerobic conditions. We hypothesize that these rTCA reactions may have been utilized by the *Chlorobi* to oxidize reduced sulfur compounds [53, 54]. In our results, the *Bacteroidetes* species however revealed a higher selection advantage over *Chlorobi* due to the observed low *Chlorobi* abudunces and richness (i.e with only a single order, the *Chlorobale*).

The identified *Proteobacteria* guilds consisted of metabolically versatile species and were revealed to comprise the highest number of species (richness) compared to other identified phyla in the treatments. The guilds comprised of species with capacity to metabolize proteins [55] carbohydrates and lipids [56] and we associated their low β-diversity (S4B and S5B Figs) with low cow-dug's substrate composition heterogeneity and differences in environmental conditions (Tables 2 and 3). The *γ-Proteobacteria* was revealed in our study as the most diverse class of the *Proteobacteria* and the class comprised of 14 orders (S7 Fig). The *Enterobacteriales* included *Escherichia coli*, which are facultative anaerobes able to ferment carbohydrates into lactate, succinate, ethanol, acetate, $H_{2(g)}$ and $CO_{2(g))}$ [57]. Unlike other syntrophic species, the affiliates of *Enterobacteriales* have been found to use methyl-citrate cycle [58] to metabolize propionate into pyruvate, a precursor for butanol, isopropanol and other mixed acid fermentation products. *Burkholderiales*, another *γ-Proteobacteria* phylotype identified in this study, produces succinate from carbohydrates concomitantly oxidizing $H_{2(g)}$, S^0 and Fe^0 into their respective oxidized forms to conserve energy [59, 60]. We suggest that the observed high

abundances and β-diversities of the γ-*Proteobacteria* communities (S7 Fig, Table 3) could be due to low environmental temperature.

In contrast, the affiliates of genus *Sutterella*, among other species of β-*Proteobacteria* identified in this study are known saccharolytic and nitrate reducers [60]. The species also consume propionate, butyrate and acetate syntrophically [56, 60]. We suggest that the high abundance and diversity of β-*Proteobacteria* (S8 Fig) in the treatments was favoured by the agroecological zones and prevailing environmental temperatures (Tables 2 and 3) of the reactors. The *Caulobacterales*, an order of α-*Proteobacteria* were among the identified low abundant species in our study. These species have previously been revealed to survive in low nutrient environments [61]. *Rhodospirillales* (α-*Proteobacteria*), such as *Rhodospirillum rubrum*, have been previously determined to participate in homoacetogenesis and convert $CO_{(g)}$ and formate into acetyl-CoA, $CO_{2(g)}$ and $H_{2(g)}$ [62]. The produced acetyl-CoA is further converted into acetate, the precursor for methanogenesis. Like members of β-*Proteobacteria*, the α-*Proteobacteria* species identified in this study were also influenced by the agroeclogical zones (Tables 2 and 3) and we postulate that the cow-dug's substrate composition was the main cause of the observed β-diversities and abundance variations (S9 Fig) among the treatments. The *Campylobacterales* and *Nautiliales* species which are affiliates of ε-*Proteobacteria* were also identified in our study. These species metabolize proteins in the AD systems and have also been revealed to express an additional role of antibiotic resisistance in biofilms and granules [56].

Among the identified δ-*Proteobacteria* species, *Pseudomonadales*, the gemmules producing anaerobes [63] were the most abundant and have been found to co-exist with algae [63]. The reported syntrophic traits of the *Pseudomonadales* order were revealed to enable the species to withstand unfavourable environmental conditions, which could have offered the advantage to the species against the selection pressure; over other δ-*Proteobacteria* species. Generally, the δ-*Proteobacteria* species are syntrophic oxidizers [64–66], with the ability to convert volatile fatty acids into acetate, formate, $CO_{2(g)}$ and $H_{2(g)}$. The *Syntrophobacteriales* species identified in this study e.g *Syntrophus aciditrophus* are 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 [64, 65] while *Syntrophobacter wolinii* [66] another member species, is known to use β-oxidation pathway to metabolize propionate into acetate. Other species belonging to the genus *Syntrophaceticus* identified in this study have been reported to oxidize the produced acetate into $CO_{2(g)}$ and $H_{2(g)}$ [67]. Interestingly, for these species to effeciently oxidize the respective macromolecules, they have been revealed to use an electron transport mechanism that creates positive redox potential by reversing electrons that were previously generated in the syntrophic reactions. In contrast, the identified affiliates of *Desulfarculales*, *Desulfobacteriales* and *Desulfovibrionales* phylotypes are sulfate reducers. However, in the absence of electron acceptors such as sulfate, the species are able to oxidize organic compounds, (e.g lactate, acetate and ethanol) into $CO_{2(g)}$ and $H_{2(g)}$ syntrophically [68]. The *Desulfuromonadales* species have been revealed to use Fe^{3+} compounds to oxidize ethanol into $CO_{2(g)}$ and $H_{2(g)}$, but in strict co-operation with methanogens [56]. Lastly, the *Myxococcales* and *Chlamydiae* species have been identified as fermenters of carbohydrates [69] and proteins [70], which are ultimately converted into organic acids. All these δ-*Proteobacterial* roles contribute to the systems' stability by providing electrons, which act as reducing equivalents in the AD systems. We also postulate that differences in cow-dug's substrate composition were the main drivers of the observed variations in the δ-*Proteobacteria* abundances and β-diversities among the treatments (S6B Fig). The other low abundance species identified in our study were assigned to the *Mariprofundale*'s order, the affiliates of *Z-Proteobacteria* class. The affiliates co-exist with *Nitrosopumilales* to oxidize Fe^0 [71] into Fe^{2+} or Fe^{3+} ions. Moreover, the two phylotypes species have previously been revealed to convert nitrogen substrates into $NO_{3\ (aq)}^-$ in their syntrophy [72,

73]. In this study, through their syntrophy, we hypothesize the completion of redox reactions in the treatments through electrons release. In summary, all affiliates of *Proteobacteria* identified in this study are capable of reversing oxidative damage to methionine [56], a trait that explains why their species are more (richness) compared to other phyla species in our treatments. Moreover, *Proteobacteria* species are able to express shorter diversification time-scales over other microbes, which lead to large population sizes, high growth rates, and strong selective regimes of the species. All these factors facilitate rapid *Proteobacteria* adaptation through either mutation or recombination events.

We also identified *Solibacterales* and *Acidobacterales* species that were affiliated to the phyla *Acidobacteria*. Members of these species ferment polysaccharides and monosaccharides into acetate and propionate [74]. The species are also able to fix $CO_{2(g)}$ anaplerotically [75], ultimately replenishing carbon atoms in the AD systems. Furthermore, the affiliates are also able to catabolize inorganic and organic nitrogen substrates as their N-sources [76]. The detection of these species in our study is attributed to the high amount of $NH_4^+{}_{(g)}$ in our AD systems. We postulate that reactor perturbations rather than environmental variables (Tables 2 and 3) were the main influencers of the genetic composition of the *Acidobacteria* in our study (S25B and S26B Figs).

Generally, the *Mollicute*'s species (*Tenericutes* phylum) detected in our study are facultative anaerobes. These were all absent in reactor 5, probably due to the nature of the reactor's substrates or feeds fed to the cattle breeds. The species are fermenters of sugars, and they convert the substrates into organic acids such as lactic acid, saturated fatty acids and acetate [77, 78]. All the identified species are affiliates of *Acholeplasmatales*, *Mycoplasmatales* and *Entomoplasmatales* which have previously been shown to exhibit reduced respiratory systems, with incomplete TCA cycle, that lack quinones and cytochromes. The metabolisms of *Tenericutes* generally yield low amount of ATP and large quantities of acidic metabolic end products [77]. It is probable that the produced acidic metabolites negatively affected the performance of other microbes in the systems in our study; which led to poor reactor performance. We hypothesize that the occurrence of these species in our systems was driven by the agroecological zones and soil types, which directly affected the nature of the cow-dug's substrates fed into our treatments (Tables 2 and 3 and S30 Fig).

Like other *Bacterial* species, the affiliates of *Actinobacteria* (S17A Fig) identified in this study, are able to metabolize a wide variety of substrates [79–81] to produce acetic acid, lactic acid, $H_{2(g)}$ and other products. Among the *Actinobacteria*, *Coriobacteriales* metabolize formate and $H_{2(g)}$, and concomitantly reduce $NO_3^-{}_{(aq)}$ into $NH_{3(g)}$ [79]. *Slackia Heliotrinireducens* was one of the *Coriobacteriales* species identified in this study. Species of the *Bifidobacteriale* were also identified in our study. These species are able to metabolize oligosaccharides and release lactic acid [80]. Other species belonging to *Actinomycetales* syntrophically metabolize propionate into acetate [81]. The metabolism of $NH_{4(aq)}$ by *Actinobacteria* is known to protect methanogenic *Archaea* from other oxidizing agents. However, their species have been reported to be more sensitive to nutrients perturbations [81] and due to this, reactor perturbations particularly the feeding regimes (S17B Fig) may have contributed to the observed species abundances and diversities among our treatments in this study. The low abundance of *Plactomycete*s in this study has also been observed elsewhere [82]. *Plactomycete*s have been found to metabolize sulphate-containing polysaccharides [82], recalcitrant hydrocarbons and other organic compounds [83]. However, when in syntrophic association with *Chloroflexi* (S18A–S21A Figs), the species are also able to oxidize butyrate [84] into smaller substrates.

The low abundance of communities of *Thermotogae* in this study is similar to the findings of previous studies [85]. The carbohydrate metabolism roles of the *Thermotogae* were previously determined based on the gene content of *Defluviitoga tunisiensis* [86]. The species have

been reported to convert the macromolecule into acetate, ethanol, $CO_{2(g)}$, and $H_{2(g)}$. In most cases, the species utilize the Embden-Meyerhof Parnas (EMP) pathway to generate pyruvate, which is either converted into acetyl-CoA, the precursor for acetate and ATP formation or reduced to ethanol, while at the same time oxidizing electron carriers. The $H_{2(g)}$ produced in these reactions is coupled to the oxidation of NADH and reduction of ferredoxin [86]. The reactions have been found to favour proton production which reduces the P_{H2} in the AD systems. However, for these reactions to proceed favourably *Thermotogales* have been revealed to depend on *Clostridiales* species [87] which implies that the abundance and richness of *Thermotogales* is regulated by the *Clostridiales* species. The *Spirochaete*s were detected in low abundance in our study as was also found out by Wang and his co-wokers [88]. The identified *Spirochaete*s in this study are homoacetogens and acetogens which have previously been revealed to convert $H_{2(g)}$, $CO_{2(g)}$ and $CO_{(g)}$ into acetate [88, 89]. These species have also been revealed to co-exist with methanogens to provide an alternative metabolic route, termed as the butyrate oxidation pathway for $H_{2(g)}$ consumption [89]. Similarly, the *Spirochaetes* species have also been found to co-exist with the *Thermotogale's* species and to jointly ferment tetrasaccharides and cellobiose into organic acids [87, 90]. These syntrophies protect *Clostridiale's* species from substrate inhibition mechanisms [87]. Based on our findings and those reported elsewhere, we suggest that numerous microorganisms identified in our AD systems co-operate through the use of an assortment of mechanisms to metabolize the substrates. Due to this cooperation, we further postulate the existence of a holobiontic mechanism in our treatments, which could be regulating symbionts and other syntrophic species through ecological selection, via co-evolution to minimize conflict between or among the symbiotic species [91].

The *Verrucomicrobia* communities identified in this study were affiliated to *Methylacidiphiles*, *Puniceicoccales* and *Verrucomicrobiales* orders (S28A and S29A Figs). These communities have been established to metabolize cellulose and other C_1-containing compounds [92]. The affiliates of *Cyanobacteria* were classified as *Chroococcales*, *Oscillatoriales* and *Nostocales* species (S23A and S24A Figs) which are mainly involved in $H_{2(g)}$ and $N_{2(g)}$ metabolism [93–95] in addition to carbohydrate metabolism. Research has shown that the conversion of N-containing substrates into $NH_{3(g)}$ is an endergonic reaction that requires ATP [93–95]. The reaction has also been found to oxidize $H_{2(g)}$ into obligate $H^+_{(aq)}$ ions, circumventing high P_{H2} in the AD systems [93–95]. Therefore, the recycling of $H_{2(g)}$ through the oxyhydrogen route, also termed as the Knallgas reaction by the *Cyanobacteria* species, has been revealed to provide ATP and protect strict anaerobes and other enzymes from oxidizing agents, by providing reducing equivalents and removing toxic $O_{2(g)}$ species [96, 97]. In our case, we attributed the variation in abundance of *Cyanobacteria* species among the treatments to the variability of $H_{2(g)}$ concentration in their respective treatments.

Another identified phylotype in our study was the *Candidatus poribacterial* species. These species ferment carbohydrate and proteins, including urea [98] into organic acids. The species have been revealed to use the Wood-Ljungdahl pathway also known as $CO_{2(g)}$ reductive pathway to fix $CO_{2(g)}$ generated from carbohydrates and other macromolecules. The species have also been previously reported to utilize the oxidative deamination pathway [99] to ferment amino acids, and the randomized methyl-malonyl-CoA pathway to metabolize propionate. All these reactions produce $H_{2(g)}$ and we postulate commensal symbiontism between *C. poribacteria* and acetogenic species in our treatments. Like other *Bacteria* species, *Fibrobacter* members such as *Fibrobacter succinogenes* identified in our study were likely to have been involved in the metabolism of cellulose [100] and other macromolecules in our AD systems.

Although it's not a common tradition to explain the existence of metal-reducing organisms in the AD systems, it is worth noting that the electron sink ability of any system depends on these organisms. One of such phylotypes identified in our study were the affiliates of

Gloeobacterales order (S22A and S23A Figs). The member species have previously been reported to transfer electrons directly from methanogens to syntrophic *Bacteria* [101]. These kinds of reactions are thought to enhance clean biogas production through the inhibition of sulfate reducing *Bacteria*. Other identified metal reducing species in our study were the members of *Clostridiales* that have previously been reported to utilize Fe^{3+} Co^{3+} and Cr^{6+} as electron acceptor [101]. Other identified *Clostridiales*' species have also been reported to use the acrylate pathway to reduce arsenate and thiosulfate concomitantly fermenting organic molecules [102]. The affiliates of *Nitrospirales* (*Nitrospira* phylum) identified in this study, have been reported to reduce sulfate using $H_{2(g)}$ and thereby promoting acetate oxidation [103, 104]. We hypothesize that all these metal reducing reactions contributed to the reactor stability through the provision of reducing equivalents in the systems.

The identified *Archaeal* biomes

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 $CH_{4(g)}$ [105]. The *Archaea* domain constituted six (6) methanogenic and ten (10) non-methanogenic guilds identified in our study, with the two groups playing equally important roles in the AD systems. The methanogenic species are consumers of *Bacterial* products and convert them into $CH_{4(g)}$, the biogas. The methanogenic and non-methanogenic species identified in this study are affiliates of *Euryarchaeotes*, which utilize hydrogenotrophic, acetoclastic, methylotrophic and methyl-reducing pathways to metabolize the substrates of *Bacterial* species into $CH_{4(g)}$. Generally, the species of *Euryarchaeota* are known to indirectly stabilize the AD systems by consuming $H_{2(g)}$ [106] and we suggest that the varied $H_{2(g)}$ concentrations among the treatments, contributed to the observed variances in *Euryarchaeota*'s abundances and β-diversities among the local communities. However, environmental (Tables 2 and 3) and reactor perturbations may also have influenced the community abundances. Among the methanogens, *Methanosarcinales*, the affiliates of *Methanomicrobia* class, were the second most abundant group, consisting of 1–3% of the total communities identified. The affiliates of *Methanosarcinales* use 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 is also consumed by the same species to maintain P_{H2} [107]. However, some species within the guild, are known to lower the accumulated P_{H2} through the reassimilation of acetate, the process that transfers $H_{2(g)}$ to acetogens. Under certain conditions, the species of *Methanosarcinales* (e.g *Methanosarcina acetivorans C2A* and *Methanosarcina barkeri*) and *Methanobacteriales* (*Methanothermobacter thermoautotrophicus*) have been reported to express metabolic features similar to acetogenic organisms. Previously, these species have been reported to use acetyl-CoA pathway to metabolize $CO_{(g)}$ into acetate and formate, under different $P_{CO(g)}$ [108]. These reactions provides alternative route for ATP generation from $CO_{2(g)}$ reduction by the generated $H_{2(g)}$. Normally, the *Methanosarcinales* couple methanogenesis to ion transport, also termed as chemiosmotic coupling or chemiosmosis, to establish an electrochemical gradient across the cell membrane [108]; because the $CO_{2(g)}$ reduction reaction is normally unfavourable. The *Methanosarcinales* have however been reported to express significant metabolic plasticity.

Among the *Archaea* in our study, *Methanomicrobiales*, affiliates of the *Methanomicrobia* class (S36A Fig) were the most abundant within the class, comprising of 2–4% of the identified total microbial populations. The affiliates of *Methanomicrobiales* use two methanogenesis pathways; hydrogenotrophic and methylotrophic routes, to convert formate, $H_{2(g)}/CO_{2(g)}$, 2-propanol/$CO2_{(g)}$ and secondary alcohols [109] into $CH_{4(g)}$. The *Methanomicrobiales*

consume isopropanol to yield acetone and the reducing equivalent, NADPH [110]. However, the species are known to lack the $CO_{2(g)}$ reductive pathway, and hence uncouple methanogenesis from $CO_{2(g)}$ fixation, which implies that the species uses electron birfucation to conserve energy. Due to these characteristics, the *Methanomicrobiales'* species are able to express high acetate requirement phenotype for carbon assimilation [111]. In costrast, the identified *Methanocellale*'s species, which are affiliates of the same class, solely use the acetoclastic route to convert $CO_{2(g)}$ into $CH_{4(g)}$ [112]. Majority of the *Methanocellales'* species have been reported to reduce $CO_{2(g)}$ with $H_{2(g)}$, except in rare cases where formate and secondary alcohols are used as alternative electron donors. Similarly, all the identified *Methanobacteriales'* species, the affiliates of *Methanobacteria* class, use the $CO_{2(g)}$ reductive route to convert $CO_{2(g)}$ into $CH_{4(g)}$ with formate and $H_{2(g)}$ acting as electron donors. We are prompted to postulate that formate metabolism particularly in acetoclastic methanogenesis was triggered by the low P_{H2} and that 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 *Methanococci* class (S34A and S35A Figs). The members of the two classes have previously been reported to metabolize $H_{2(g)}$ and $CO_{2(g)}$ into $CH_{4(g)}$ under thermophilic and saline conditions [113, 114], which indicate that our treatments had high amounts of salts and that the temperatures were high enough to support the species.

In contrast, the non-methanogenic group of *Archaea* generally reduce metal elements and their compounds [101, 102], to provide reducing equivalents in the AD systems. The affiliates of *Halobacteriales* (*Halobacteria*, *Euryarchaeota)* were among the detected non-methanogenic species in this study. The member species of this group of non-methanogens use the arginine deiminase pathway [115, 116] or other metabolic routes to catabolize amino acid substrates. The *Halobacteriales* conserve energy when the intermediate metabolites; ornithine and carbamoyl-phosphate, are further converted into $CO_{2(g)}$ and $NH_{3(g)}$. The significant differences in relative abundances of the *Halobacteriales* species among the treatments can be attributed to the concentration of the alternative electron acceptors and salts concentrations in the systems. However, other deterministic factors such as competition and niche differentiation could also have led to the observed variations in *Halobacteriales* abundance among the treatments.

Our study also identified *Thermoplasmatales*, the affiliates of *Thermoplasmata* class. These groups of *Archaea* utilize sulfur for energy consevation [117]. However, their close relatives, *Methanomassiliicoccales*, formerly *Methanoplasmatales* have been found to use the methylreducing pathway to convert methanol into $CH_{4(g)}$. These relatives conserve energy through electron bifurcation [118] and we postulate that the affiliates of *Thermoplasmatales* are also involved in $CH_{4(g)}$ production, only that few or none of the biochemical assays have previously been conducted to test their biogas production abilities. We also identified *Thermococcales'* species, of the class *Thermococci*. The affiliates of this class of *Archaea* have been reported to have short generation time and to withstoond nutrient stress for longer durations [119]. Like for *Bacterial* species, and in the absence of sulfur, *Thermococcales*, have been reported to be involved in both acidogenesis and acetogenesis reactions. Species of this group metabolize proteins and carbohydrates into organic acids, $CO_{2(g)}$, and $H_{2(g)}$ [120, 121]. Other affiliates of *Thermococcales'* are involved in homoacetogenic reactions through which they oxidize $CO_{(g)}$ into $CO_{2(g)}$, and catabolize pyruvate into $H_{2(g)}$ and other products [122, 123]. We also identified *Archaeoglobales*, the affiliates of *Archaeoglobus* class which use sulfate, as the reducing agent, to oxidize lactate, long fatty acids and acetate into $CO_{2(g)}$ [124]. Some species e.g *Geoglobus ahangari*, however, reduce Fe^{3+} compounds with $H_{2(g)}$ [125]. The identified *Archaeoglobales* and other metal reducing species identified in our study indicate the importance of these microbes in redox balancing reactions which contribute to reactor stability.

The *Sulfolobales* and *Desulfurococcales*, all belonging to *Thermoprotei* class and *Crenarchaeota* phylum were also identified in this study (S37A Fig). Species of these groups metabolize carbohydrates and amino acids [126, 127] into pyruvate and $CO_{2(g)}$. They also ferment pyruvate into lactate and $CO_{2(g)}$. The $CO_{2(g)}$ generated from the above reactions either enters into the $CO_{2(g)}$ reductive pathway, the route that is catalyzed by acetogens/homoacetogens to produce $CH_{4(g)}$ or is fixed through the 3-hydroxypropionate/4-hydroxybutyrate or dicarboxylate/4-hydroxybutyrate cycles by the identified *Thermoprotei* species to generate the energy [128]. Among other *Thermoprotei* affiliates identified in this study were *Thermoproteales* that use sulfur and its compounds as electron acceptors, to ferment carbohydrates, organic acids and alcohols [129]. In these reactions, the species consume $CO_{2(g)}$ and $H_{2(g)}$. Though the species have been previously reported to grow at low rate, they also ferment and oxidize amino acids and propionate, respectively [130]. Other affiliates like *T. neutrophilus* are known to assimilate acetate and succinate in the presence of $CO_{2(g)}$ and $H_{2(g)}$ [120, 131]. This species use $H_{2(g)}$ to reduce sulfur. Unlike most of the other *Archaeal* species, *Thermoproteales* use oxidative and reductive TCA routes to conserve energy. *Acidilobales* another *Thermoprotei* order identified in this study uses S^0 and thiosulfate to metabolize protein and carbohydrate substrates [132]. The species use either EMP or the Entner-Doudoroff (E-D) [133] pathways to convert the substrates into acetate and $CO_{2(g)}$ [134]. The affiliates of *Acidilobales* are also able to oxidize triacylglycerides and long-chain fatty acids [134] into acetate and $CO_{2(g)}$. Substrate-level and oxidative phosphorylation are the major routes of energy conservation in *Acidilobales*; and the two mechanisms occur when S^0 and acetyl-CoA concentrations are high. Due to S^0 respiration in *Acidilobales* species, we postulate that their member species close the anaerobic carbon cycle via complete mineralization of organic substrates and the variations of these substrates together with other deterministic and non-deterministic (Tables 2 and 3) factors contributed to the observed variation in abundance among the treatments and the high β-diversity (S37B Fig). We also hypothesize that soil type was another factor that exerted natural selection on the *Acidilobales* through the alteration of the cow-dung's substrates chemical composition.

The identified *Thaurmarchaeota* communities were classified into *Cenarchaeles* [135] and *Nitrosopumilales* [136] phylotypes (S38 Fig). The member species of these groups use the modified 3-hydroxypropionate/4-hydroxybutyrate pathway to fix inorganic carbon to generate energy. The communities are also able to metabolize urea and its products into, $NH_{3(g)}$ [137, 138]. The *Nanoarchaeota* (S31A Fig) species exploit *Crenarchaeota* mechanisms [139, 140] to metabolize amino acids and complex sugars [141, 142] while the *Korarchaeotes* are sole amino acid fermenters. The energy metabolism in the last two phyla is still unclear and further biochemical studies are needed to unravel their full metabolic activities.

The identified *Fungal* biomes

Like the *Bacterial* species described in this study, the identified *Fungi* (S39A–S41A Figs) have the ability to convert macromolecular substrates into acetate, $CO_{2(g)}$ and $H_{2(g)}$ [143]. Among the identified *Fungal* species, *Ascomycota* were the richest and most abundant guild, comprising of <1% of the identified communities in the treatments. Among the *Ascomycota* species, a majority were affiliates of *Saccharomycetales* (*Saccharomycetes*, *Ascomycota*), that generally use the EMP pathway to ferment glucose and other sugars into pyruvate. One of such identified species is *Saccharomyces cerevisiae* [144] that is known to ferment carbohydrates into mixed fermentation products. Other *Saccharomycetes* species e.g *Hansenula polymorpha* yeast ferment monosaccharides, fatty acids and amino acids into acetate, ethanol, $CO_{2(g)}$ and $H_{2(g)}$ [145]. Furthermore, the species are also able to oxidize methanol into formaldehyde and H_2O_2 $_{(g)}$ [146]. The two products act as the sources for formate and $H_{2(g),}$ the precursors for

methanogenesis. The observed variations in abundance of the *Saccharomycetes* species may be attributed to their metabolic plasticity, deterministic factors and other environmental variables. The other *Ascomycota* affiliates belonging to *Pezazales* (*Pezizomycetes*), *Pleosporales* (*Dothideomycetes*), and *Eurotilaes* (*Eurotiomycetes*) identified in this study use mixed fermentation pathways to ferment carbohydrates [147, 148]. The metabolisms of these *Ascomycota* species have however been observed to vary within and among the phylotypes, despite the presence of a common enzyme for the pyruvate metabolism [147, 148]. For example, among the *Eurotiale*'s order; the species *Paecilomyces variotii* expresses low metabolic capacity on plant-derived compounds compared to other *Eurotiale*'s members [149, 150]. This seems to indicate the genetic diversity of *Eurotiomycetes*' species, seen in the variations of abundance of the species among the treatments in this study. Our data also suggests that the communities of *Eurotiomycetes* were influenced by the environmental temperature and soil type (Tables 2 and 3 and S45A Fig). Like the identified *Fusobacteria* species and other identified amino acid metabolizers, *Dothideomycetes*, another amino acid consuming phylotype, metabolize lipids and proteins into volatile organic acids [150]. On the other hand, the affiliates of *Pezizomycetes* use sulfate as an electron sink to consume amino acids molecules [151, 152]. We attribute the community variation observed within the above groups to the composition of the cow-dung's substrates and natural selection pressure differences among the treatments.

The *Sordariomycetes*' affiliates identified in this study (S44A Fig) produce enzymes that express an assortment of degradative mechanisms and other fermetative roles. The *Hypocreales*' affiliates are such an example that produces powerful hydrolytic enzymes [153] that catabolize complex biomass in their vicinity. *Fusarium* sp. a member of *Hypocreales*, metabolizes $NH_{3(g)}$, lipids and proteins [154, 155] into their respective smaller molecules. The species has been found to use NAD(P)H to reduce $NO_3^-{}_{(aq)}$ into $NH_4^+{}_{(aq)}$. The production of $NH_4^+{}_{(aq)}$ by *Hypocreales* is normally coupled to the substrate-level phosphorylation and ethanol oxidation; the processes that produce acetate [154, 155]. The two mechanisms ameliorate the $NH_4^+{}_{(aq)}$ toxicity to micro-organisms. Other *Hypocreales* species such as *Hirsutella thompsonii* confer wide metabolic activities that involve chitinases, lipases, proteases and carbohydrate degrading enzymes [156]. *Monilia sp.* [157] and *Neurospora crassa* [158] were among the identified *Sordariomycetes* and they convert cellulose into ethanol. *Magnaporthales* such as *Pyricularia penniseti* convert isoamylalcohol into isovaleraldehyde while other *Magnaporthales*' species convert fumaryl-acetoacetate into fumarate and acetoacetate [159, 160]. Other roles played by the *Sordariomycete*'s species include the metabolism of aromatic compounds through which, phenolics and aryl alcohols/aldehydes are converted into acetate, formate and ethanol [161, 162]. Our data suggests that *Sordariomycetes* community's abundances were influenced by the cow-dug's substrates and the environmental variables (Tables 2 and 3) of the treatments.

In contrast to the *Ascomycota* species, affiliates of *Basidiomycota* establish susceptible hyphal networks [163] that may have been constrained by the reactor pertubartions in our study. This may explain why the *Ascomycota* species dominated over the *Basidiomycota*'s species in this study. Nonetheless, like *Ascomycota* species, the affiliates of *Malasseziales*, *Ustilaginales*, *Agaricules* and *Tremellales* species, all belonging to *Basidiomycota*, are able to metabolize carbohydrates [148, 164], urea [164] and $NO_3^-{}_{(aq)}$ [164] in the systems. Other members like *Polyporales* metabolize woody substrates and their respective products [165]. All these functions clearly indicate that *Basidiomycota's* species played diverse metabolic roles in our AD systems in this study. Interestingly, we also identified *Chytridiomycota* affiliates in reactor 9 and 10. These affiliates are capable of catabolizing pollen, chitin, keratin, and cellulose [166, 167], which may be indicative of the presence of these substrates in the two treatments. The detected *Spizellomycetales*, *Cladochytriales* and *Rhizophydiales* species in the two

treatments, further revealed the presence of highly decayed plant tissue with high cellulose content. The identified *Monoblepharidomycetes* members are close relatives of *Cladochytriales* and *Spizellomycetales* [168] species and were detected only in reactor 9 and we postulate that these groups perform similar roles in the AD systems.

Generally, *Chytridiomycota* species are carbohydrate metabolizers as previously revealed by the presence of glucose transporters and invertase activity in their systems [166, 167]. The species use electron bifurcation mechanisms to conserve the energy and the mechanisms occur in the hydrogenosome [169]. The *Blastocladiales*, the close relatives of *Chytridiomycotes* and the affiliates of *Blastocladomycota*, class *Blastocladomycetes* were only detected in reactor 9 in our study. Their member species metabolize cellulose into acetate, $CO_{2(g)}$ and $H_{2(g)}$. Interestingly, *Blastocladiales* have previously been reported to use the Malate fermentation pathway to generate and conserve energy [170]. *Mortierellales* formerly classified under *Mucoromycotina* [171] and latter assigned to *Mortierellomycotina* phylum [172] are converters of organic compounds into poly-unsaturated fatty acids [173] while their relatives *Entomophthorales* and *Mucorales* metabolize decaying organic matter. However, the affiliates of the three orders were only detected in reactor 9, 10 and 11, respectively.

Like similar studies [174, 175], our data revealed that *Fungi* co-occurred with the species of *Bacteria* and methanogens. In these syntrophy, we postulate that 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 increase metabolism of the dry matter [176] in the treatments. We also suggest that the *Fungal* syntrophy is more complex than simple cross-feeding mechanisms, because the species shift more oxidized products into a more reduced form. Through these mechanisms, the precursors, lactate and ethanol are converted into acetate and formate depending on the prevailing conditions. The involved interactions between *Fungi* and other domains' species have been determined to be very crucial to an extent of influencing the isolation and culturing of some *Fungal* species [176, 177]. The findings indicate why *Fungal* species should be jointly considered with other domains' species in the context of biogas production.

The microbiomes diversity and biogas production

To establish whether environmental conditions and perturbations influenced the microbial populations, we compared the collected phenotypic traits with the genotypic dataset. The phenotypic datasets revealed an insight on the cause of the observed communities' β-diversities and biogas production variations. We attempted to use the principles of the four ecological mechanisms that include selection, drift, speciation, and dispersal to better understand the cause of the observed variance in microbial composition among the treatments and thus the cause of the biogas production variability among the sites. While selection is deterministic and ecological drift is stochastic, dispersal and speciation may have contributed to the two ecological parameters. One of the challenges of examining communities' in any given niche is the difficulty of estimating 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, we suggest that selection and drift were the likely two main factors that contributed to variation of community abundances among the treatments and the observed high β-diversities in our study. The two factors were most likely influenced by the substrate inputs (Table 3) and disturbances from environmental variables (Tables 2 and 3). Environmental conditions directly affect the chemical composition of the substrate fed to the cattle breeds. As a result, their wastes, the cowdung is fed to

the reactor which ultimately influences the microbial composition and biogas productivity. 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. All these could have contributed to the observed variation of biogas production among the studied treatments. We postulate that environmental temperatures combined with 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. The findings increased the number of the total species in the few treatments, where the rare species were identified as opposed to where they were absent. Nonetheless, we never observed significant differences in reactor performances between the two categories of the treatments. Consequently, we linked the observations with the low abundances of the rare *Fungal* species identified in the few respective reactors, which implies the importance of the species abundance in biogas production. The composition of the cow-dung's substrates fed to these treatments could largely be influenced by cattle breeds and several environmental factors including agroecological zones, soil type, rainfall, environmental temperatures to mention but a few, and its chemical compositions could have stimulated the growth of specific species, favouring colonization and speciation in the respective treatments. Similarly the same phenomenon could have happened to other microbial communities. In this case, the available resources could have reduced competition among the species, weakening the niche selection abilities. Generally, these ecological factors normally strengthen the ecological drift and prority effects [178] that result to unpredictable site to site variation of microbial abundances and high β-diversities, even under similar environmental conditions. We used the same arguments to explain our observations on the variations of the relative abundances among treatments and high β-diversities, particularly when the conditions were similar (Tables 2 and 3). We also extended the same argument to the biogas production capacities among our treatments.

Besides the cow-dung's substrate effects, we hypothesize that the intensity and duration of the disturbances [179], led to high rate of species mortality, low growth rates and species extinction. These kinds of disturbances caused decrease of stochasticity and to account for their effects on our findings, we adopted the concepts of neutral theory [179]. Generally their associated caveats reduce ecological drifts that result to low competition abilities among the low abundance species. Due to these facts, we postulate that the persistence and intensities of these disturbances in our specific treatments eliminated weak and less adapted microbial populations. 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 (S51 Fig) was an indication of unfavorable disturbances that could have originated from either cow-dung's substrates or other reactor perturbations. All these may have led to more predictable site-to-site variations and we used these concepts to explain the observed low β-diversity among the treatments (e.g in reactor 2 and 10, reactor 3 and 6 etc). However, to understand better the links between the environmental conditions and the microbial communities in these AD systems, we suggest the utilization of the 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 variance 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 be used to ask and test whether dispersal, diversification, environmental selection and ecological drift were the vital elements 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.

Conclusions and future prospects

Our results provide an insight into the biogas microbiomes composition, richness, abundances and variations within and among the taxa and the twelve reactor treatments. Interestingly, microbiomes structure and richness were similar among the twelve treatments, except for 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. The Cow-dung's substrates 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 phylotype. All the identified species abundances among the treatments and the community's β-diversities in the majority of the treatments significantly varied. All these contributed to our conclusion that environmental and spatial variables largely influenced biogas microbiome populations and thus the amount of biogas produced. The discoveries highlight the importance of identifying the involved natural processes that give rise to such identified variations. Based on the results obtained in this study, we suggest further studies on species responses on the natural conditions for optimal biogas production and microbial community managements. We also suggest the application of metacommunity theory among other ecological theories in developing management strategies for the AD systems.

Materials and methods

Ethics considerations

Before embarking on the research, the study protocols were approved by the Institutional Research and Ethics committee (Ref. No. IREC/2017/52) of Moi University. Permission for conducting research was granted under the permit No. NACOSTI/P/17/86022/18890 by the National Commission for Science, Technology and Innovation, Kenya. The procedures were also reviewed and approved through the University of Eldoret [Ref. No. UoE/B/DVASA/REG/97] Research and Innovation Committee and the Institute for Biotechnology Research Board, JKUAT [Ref. No. JKU/BR411-1663/2014]. The consents to sample the biogas reactor treatments (installed in privately owned premises, Table 2), were sought orally from the reactors' owners. The collected phenotypic informations for the studied reactor treatments were specifically used for the purpose of the research and the owners are updated on the research findings. All other experiments were conducted in secured environments and in compliance with the biosafety and biosecurity requirements.

Experimental design

Randomized complete block design was utilized as previously described [180], 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 treatment size and the farmers' routine activities. The sludge sampling was conducted following the previously described procedure of Boaro *et al.* [181]. The treatments' gas volumes (M^3) were recorded using their installed gas metres at the farmers' premises. The treatments'

performances (M^3) were simulated through computing and integrating the reactor volumes (M^3) and the farmers actual collected gas (M^3). Geographic co-ordinates (Table 2) and agro-ecological zones (Table 3) were generated by Global Positioning System (etrex, Summit^{HC}, Garmin) and ArcGIS® software (ver10.1) [182], respectively. A pre-tested questionnaire was utilized according to Oksenberg *et al.* [183] protocol to collect the phenotypic dataset. Small scale farming and zero grazing of Freshian breeds were the main activities among the sampled farmers. However, the owners of reactor 3 and 11 (Meru and Kiambu blocks, respectively), reared Jersey breeds of cattle. Samplings were conducted during the warm and cool climatic conditions.

Sample preparation 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 determined as previously described by Campanaro *et al.* [184]. The dsDNA was normalized and library preparation performed using Nextera^{TM} 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). Quantification for the libraries was conducted by Qubit^{TM} 2.0 (Invitrogen, Life Technologies, Carlsbad, CA., USA) and Bioanalyzer 2200 (Agilent technologies, USA). To prepare for cluster generation and sequencing, equal volumes (4nM with EBT buffer, 10mM TRIS, 0.1%Tween20) of normalized libraries were pooled, diluted in hybridization buffer (HT1) and denatured according to Nextera^{XT} protocol. PhiX control spikes were also added. *De novo* deep sequencing was conducted by an Illumina MiSeq system as previously described by Caporaso *et al.* [185] but adopting a v3 kit chemistry with 300 cycles (2×150) as previously described [184].

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.* [186]. Trimmomatic software was utilized to trim and filter the adaptor according to Bolger *et al.* [187]. *De novo* assembly of the paired-end contigs were peformed by MetaSpade software (ver3.10) as described previously [188]. To re-trim low-quality regions of the scaffolds, a bit-masked k-difference matching algorithm, termed as Skewer software [189], was utilized. SolexaQA was also used to remove the regions, as described by Cox *et al.* [190] while DRISEE, metagenomic quality assessment software was further used to remove duplicated scaffolds according to Keegan *et al.* [191]. Denoising and normalization of the scaffolds, was conducted using the fastq-mcf [192]. We screened near-exact scaffolds that matched human, mouse and cow genomes following Langmead *et al.* procedure that used a sort read aligner, the Bowtie2 [193]. M5nr [194], a non-redundant tool, FragGeneScan [195] and BLAT [196] tools were utilized to compute datasets against the sixteen (16) publicly available biological databases; RefSeq [197], IMG [198], TrEMBL [199], Subsystems [24], KEGG [200], GenBank [201] SwissProt [202] PATRIC [203], eggNOG [204], KO [205], GO [206], COG [207], RDP [208], LSU [209] SSU [210] and NOG [211], as previously described by Wilke *et al.* [194]. 70% identity cut off value, was set to filter the rRNA genes from a reduced version of M5RNA [194] using SortMeRNA software [212], while the CD-HIT algorithm [213] was used to cluster the identified RNA and the predicted amino acid sequences at 97% and 90% identity, respectively. Subsequently, we assigned our sequences to the archived taxonomic profiles at nucleotide, protein and process levels of annotations. Two ENVO ontologies: large lake and terrestrial biomes were utilized according to Field *et al.* [214]

and Glass *et al*. [215]. Alignment similarity cut off values, 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 [216]. Prior to statistical analysis, the rarefaction curves were reconstructed, datasets was $\log_2$ (x+1) transformed and normalized using an R package, DESeq software [217]. PCoA based on euclidean model [218] was set at $P \leq 0.05$, and conducted against the twelve treatment samples. Results were visualized through MG-RAST pipeline (ver4.01) [219] and Krona radial space filling display [220]. Optimal scripts and/or command lines and restful application programming interface were utilized in our analysis.

Supporting information

S1 Table. The quality control statistics of the obtained scaffolds for the twelve treatments. (PDF)

S1 Fig. Barchart showing unfiltered and filtered sequencing reads (a) and the known and unknown protein genes (b) in our samples. More than 53.07% of the filtered nucleotide reads in our samples contained unknown proteins. (PDF)

S2 Fig. Stacked barchat showing some (21 phyla) of the identified *Bacteria* domain phyla, relative abundances (a) and the PCoA plot (phylum level), based on the Euclidean model (b). The communities' nucleotide composition of reactor 2 and 10 and those of reactors 4 and 9 partially clustered in the upper left and right quadrants of the plot respectively. (PDF)

S3 Fig. The PCoA plot showing the relative abundances variation at the class (a) and order level (b). At the class and order level, the nucleotide composition of reactor 2 and 10, were positioned on the lower left quadrant of the plot, those of reactor 3 and 6, on the upper left quadrant, the communities of reactor 8 and 11, on the upper right quadrant, while the composition of reactor 4 and 9, were postioned on the lower right quadrant of the plot. All clustered partially at both level. (PDF)

S4 Fig. The stacked barchart revealing six classes of *Proteobacteria* communities, relative abundances (a) and their PCoA plots based on the Euclidean model (b). The nucleotide composition for reactor 1, 3 and 6 clustered on the upper right quadrant of the plot. Similarly, the composition of reactor 4, 7 and 12 clustered partially on the lower right, while those of reactor 5 were singly positioned on the upper left quadrant of the plot. (PDF)

S5 Fig. The PCoA plot based on the Euclidean model for all the identified *Proteobacteria* orders. The PCoA plot indicated partial similarities of the nucleotide composition in half of the studied treatments, including reactor 1, 3 and 6, upper right quadrant; reactor 4, 7 and 12, lower right quadrant of the plot at the class level. Similar observations were made at the order level, with an exception of few treatments (reactor 4, 7, 8 and 9 that formed cluster). (PDF)

S6 Fig. Stacked barchat showing seven *δ-Proteobacteria* orders, relative abundances (a) and their PCoA plot revealing nucleotide composition variations, based on the Euclidean model (b). The plot revealed dissimilarity of nucleotide composition among the treatments, with an exception of the composition of reactor 4 and 7 that were found to cluster on the lower left

quadrant of the plot.
(PDF)

S7 Fig. Stacked barchart showing fourteen *γ-Proteobacteria* orders (a) and their PCoA plot based on the Eucleaden model (b). The plot revealed clustering of the reactor 1 and 3 nucleotide composition, which were in close proximity with the composition of reactor 7. The communities of reactor 7 clustered partially with those of reactor 8 and 12. However, the composition of reactor 4 and 9 were in close proximity. Interestingly, all the clustered treatments were positioned on the upper right quadrant of the plot. Moreover, the nucleotide compositions of reactor 5 were singly positioned on the upper left quadrant of the plot.
(PDF)

S8 Fig. Stacked barchat showing seven *β-proteobacteria* orders, relative abundances (a) and their PCoA plot revealing nucleotide composition (dis)similarities among the treatments (b). The plots revealed partial clustering of reactor 7 and 8 nucleotide compositions while the nucleotide composition of reactor 3, 6 and 12 were found to cluster on the lower left quadrant of the plot. The majority of the treatments comprised dissimilar *β-Proteobacteria* nucleotide compositions.
(PDF)

S9 Fig. Stacked barchat showing the seven *α-Proteobacteria* orders, relative abundances (a) and their PCoA plots, revealing the nucleotide composition variation among the reactors based on the Euclidean model (b). The plots revealed close proximity of reactor 1 and 6 nucleotide composition and were positioned on the y-axis. However, the composition of reactor 3 and 11 clustered partially on the upper right quadrant while the nucleotide compositions of the other treatments were found to reveal dissimilarity.
(PDF)

S10 Fig. Stacked barchat showing two *ε-Proteobacteria* orders, relative abundances (a) and their PCoA plot based on Euclidean model (b). The PCoA plot for reactor 4 and 8 clustered on the upper left quadrant of the plot while those of reactor 5 and 9 clustered partially on the upper right quadrant.
(PDF)

S11 Fig. The stacked barchat revealing four *Firmicutes* classes, the relative abundances (a) and their PCoA plots, revealing variance among the twelve reactors based on the Euclidean model (b). The PCoA plots revealed close proximity of reactor 4 and 9 and reactor 1 and 8 nucleotide compositions, while the reactor 1 and 7 composition clustered partially on the plot. However, those of reactor 2, 10 and 12 were found to cluster on the lower left quadrant of the plot while other treatments' compositions were found to reveal dissimilarity, except the composition of reactor 5 that were singly positioned on the upper left quadrant of the plot.
(PDF)

S12 Fig. Stacked barchat showing eight *Firmicute*'s orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The plot revealed partial similarities between the composition of reactor 2 and 10, reactor 3 and 6, and reactor 7 and 12 while the nucleotide composition of reactor 1 and 9 almost clustered.
(PDF)

S13 Fig. Stacked barchat showing the four *Clostridia* orders (a) and their PCoA plots, revealing nucleotide composition variations among the twelve reactors based on the Euclidean model (b). The plots revealed dissimilarities of the nucleotide composition among the majority

of the treatments. However, the composition of reactor 5 and 9 were clustered partially on the upper left quadrant, while the nucleotide of reactor 7, 11 and12 clustered on the lower left quadrant of the plot.
(PDF)

S14 Fig. Stacked barchat showing two *Bacilli* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition varied among the treatments and pairwise partial clustering of the nucleotides was observed among the majority of the treatments, with an exception of reactor 5 and 8 communities that were distinctively positioned within the plot.
(PDF)

S15 Fig. The stacked barchat revealing four *Bacteroidete*s classes, relative abundances (a) and their PCoA plots revealing nucleotide composition variation, based on the Euclidean model (b). The plots revealed partial clustering of reactor 2 and 10 and reactor 8 and 12 communities on the plot. However, the communities of reactor 4 and 9 were found to cluster on the lower right quadrant of the plot.
(PDF)

S16 Fig. Stacked barchat showing the four *Bacteroidete'*s orders, the relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition of reactor 2 and 10 and those of reactor 8 and 12 clustered partially on the lower left quadrant of the plot while the composition of reactor 4 and 9 were revealed to cluster on the lower right quadrant of the plot.
(PDF)

S17 Fig. Stacked barchat showing six *Actinobacteria* orders, the relative abundance (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition of reactor 7 and 8 positioned on the upper right quadrant of the plot were found to cluster partiall, while the nucleotide of reactor 1, 3, 5, 11 and 12 formed a cluster along the y-axis of the plot.
(PDF)

S18 Fig. The stacked barchat showing four *Chloroflexi* classes, relative abundances (a) and PCoA plot revealing variance among the twelve reactors, based on the Euclidean model (b). The plots revealed close proximity of reactor 1 and 7 nucleotide composition, and were positioned on the upper left quadrant of the plot. However, the composition of reactor 4, 5, 10 and 12 were found to cluster on the lower right quadrant of the plot.
(PDF)

S19 Fig. Stacked barchat showing five *Chloroflexi* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition of reactor 4, 5, 10 and 12 clustered, on the lower right quadrant of the plot, while the composition of reactor 1 and 7 were found to reveal close proximity on the upper left quadrant of the plot.
(PDF)

S20 Fig. Stacked barchat showing two *Chloroflexi* class orders, relative abundances (a) and their PCoA plot based on their Euclidean model (b). The nucleotide composition of reactor 1 and 5 (clustered; upper right quadrant) and those of reactor 3 and 7 (clustered; lower right quadrant) were found to similar. Notably, the nucleotide reads of reactor 1 and 10 were revealed to cluster partially with the composition of reactor 3 and 7. Similarly the nucleotides of reactor 8, 11 and 12 clustered partially on the upper left quadrant of the plot.
(PDF)

S21 Fig. Stacked barchat showing the two *Thermomicrobia* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plot revealed dissimilarities of the nucleotide compositions among the twelve studied treatments.
(PDF)

S22 Fig. Stacked barchat showing *Cyanobacteria* class, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The plot revealed dissimilarities of the nucleotide composition among the treatments and only the nucleotides of three treatments (reactor 3, 7 and 10) that were found to be distinct within the plot.
(PDF)

S23 Fig. Stacked barchat showing the five *Cyanobacteria* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plot revealed four out of twelve reactors that were distinctively located in plots. Other treatments formed partial clusters or clustered within the plots.
(PDF)

S24 Fig. Stacked barchat showing the four affiliates of the Unclassified *Cyanobacteria* nucleotide reads, relative abundances (a) and their PCoA plots based on the Euclidean model (b). The plot revealed dissimilarities among the treatments, all distributed in the four plot quadrant.
(PDF)

S25 Fig. The stacked barchat showing the two *Acidobacteria* classes, relative abundances (a) and their PCoA plots for their nucleotide composition based on the Euclidean model (b). The nucleotide composition of reactor 1 and 3 clustered partially on the lower right quadrant of the plot, reactor 10 and 11 compositions were positioned in close proximity on the upper right quadrant of the plot while the nucleotides of reactor 2, 8 and 12 were found to cluster on the upper right quadrant of the plot.
(PDF)

S26 Fig. The stacked barchat showing the two *Acidobacteria* orders, relative abundances (a) and their PCoA plots for their nucleotide composition based on Euclidean model (b). The composition of reactor 1 and 3 clustered partially on the lower right quadrant of the plot, those identified in reactor 10 and 11 were in close proximity, positioned on the upper right quadrant of the plot while the composition of reactor 2, 8 and 12 clustered in the upper right quadrant of the plot.
(PDF)

S27 Fig. Stacked barchat showing two *Deinococcus-Thermus* orders, the relative abundances (a) and the PCoA plots based on the Euclidean model (b). The PCoA plots revealed clustering of the nucleotide composition of reactor 3 and 6, and those in reactor 7 and 9 on the upper left quadrant and lower left quadrant of the plot respectively. However, those identified in reactor 4 were in close proximity with those of reactor 7. The composition of reactor 11 and 12 partially clustered on the lower right quadrant of the plot. The nucleotides of reactor 2 and 10 were the only nucleotide reads located in the upper right quadrant of the plot.
(PDF)

S28 Fig. Stacked barchat showing *Verrucomicrobia* classes, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plot revealed dissimilarities of the nucleotide composition among the treatments, except those identified in reactor 3 and 7 that

clustered partially on the lower right quadrant of the plot.
(PDF)

S29 Fig. Stacked barchat showing three *Verrucomicrobia* orders, relative abundances (a) and their PCoA plot based Euclidean model (b). The composition of reactor 3, 6, 7 and 11 clustered partially, on the lower left quadrant of the plot; those indentified in reactor 1 and 10 were also clustered partially, on the upper right quadrant, while the reads of reactor 5 and 8 were found to cluster on the upper right quadrant of the plot.
(PDF)

S30 Fig. Stacked barchat showing three *Ternericutes* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition in reactor 4 and 9 and those detected in reactor 6 and 8 partially clustered in the upper right quadrant of the plot. Other treatments comprised dissimilar nucleotide composition.
(PDF)

S31 Fig. The stacked barchart showing the four *Archaea* phyla, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The compositions of reactor 4 and 12 clustered, on the lower left quadrant of the plot; while those of reactor 3 and 6 and reactor 7 and 9 clustered partially; on the lower right quadrant and upper left quadrant of the plot. Other treatments were distinctively dissimilar, positioned within the plot.
(PDF)

S32 Fig. Stacked barchat showing nine *Archaea* classes, and the affiliates of the unclassified reads, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide compositions in the respective treatments were distinctively dissimilar.
(PDF)

S33 Fig. Stacked barchat showing sixteen *Archaea* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide of reactor 3 and 7, were clustered partially on the lower left quadrant of the plot. However, the compositions of the other ten treatments were distinctively dissimilar, all distributed in the plot.
(PDF)

S34 Fig. The stacked barchat showing the eight *Eurychaeota* classes, the relative abundances (a) and their PCoA plots revealing dissimilarities of the nucleotide composition based on the Euclidean model (b). The nucleotide composition of reactor 1 and 7, were positioned on the upper left quadrant of the plot while the composition of reactor 6 and 8, were positioned on the lower left quadrant, near the y-axis were in close proximity.
(PDF)

S35 Fig. Stacked barchat showing ten *Euryarchaeota* orders, relative abundances (a) and their PCoA plot based Euclidean model (b). Their nucleotide compositions were found toreveal dissimilarity, all distributed within the four quadrants.
(PDF)

S36 Fig. The stacked barchat showing three *Methanomicrobia* orders, the relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plot revealed partial clustering of reactor 1 and 7 nucleotide composition on the upper left quadrant of the plot. Further the composition of reactor 2 and 10 and those of reactor 4 and 8 were found to be in close proximity on the lower right quadrant of the plot.
(PDF)

S37 Fig. Stacked barchat showing four *Thermoprotei* orders, relative abundances and their PCoA plot based on the Euclidean model (b). However, the nucleotide composition of reactor 3 and 8 were closely positioned in the lower left quadrant of the plot.
(PDF)

S38 Fig. Stacked barchat showing two *Thaurmarchaeota* order, the relative abundances (a) and their PCoA plot based on the Euclidean model (b). The plot revealed partial clustering of the nucleotide composition of reactor 6 and 9, positioned on the lower left quadrant of the plot, reactor 7 and 10, located near the y-axis, and reactor 2 and 11 located on the upper right quadrant of the plot. The nucleotide compositions of reactor 3 were in close proximity to those identified in reactor 7, positioned on the lower right quadrant of the plot. However, the other treatments' nucleotide compositions were found to be dissimilar.
(PDF)

S39 Fig. Stacked barchat showing five fungal phyla, the relative abundances (a) and their PCoA plots based on Euclidean model (b). The plot revealed dissimilarities of the nucleotide composition among the twelve studied treatments and were distributed within the four plot quadrant.
(PDF)

S40 Fig. Stacked barchat showing thirteen fungal classes, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plot revealed nucleotide composition dissimilarities among the twelve studied treatments.
(PDF)

S41 Fig. Stacked barchat showing 23 fungal orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide compositions of reactor 2 and 10, were positioned on the lower left quadrant; reactor 3 and 6, upper left quadrant; reactor 4 and 9, upper right quadrant; and reactor 8 and 11, on the lower right quadrant of the plot. All clustering partially in their respective quadrant.
(PDF)

S42 Fig. The stacked barchat showing the five *Ascomycota* classes, relative abaundances (a) and their PCoA plots based on the Euclidean model (b). The PCoA plots revealed dissimilarity of the nucleotide composition among the twelve treatments. However, the composition of reactor 11 was positioned singly on the lower left quadrant of the plot while those of reactor 10 were located along the y-axis (Negative PCoA 1 and Negative PCoA 2).
(PDF)

S43 Fig. The stacked barchat revealing eleven *Ascomycota* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The model revealed partial clustering of the identified nucleotide compositions of reactor 7 and 8 while those identified in reactor 2 and 8 and reactor 2 and 4 were found to reveal close proximity, located on the upper right quadrant of the plot. The *Ascomycota*'s nucleotide compositions of reactor 11 were singly positioned on the upper left quadrant of the plot.
(PDF)

S44 Fig. Stacked barchat showing three *Sordariomycetes* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The PCoA plots revealed partial clustering of the identified nucleotide reads in reactor 1 and 10 while other treatments revealed dissimilarities of the nucleotide composition.
(PDF)

S45 Fig. Stacked barchat showing the two *Eurotiomycetes* orders, the relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition of reactor 1 and 3 clustered on the x-axis (positive PCoA 1 and Positive PCoA2); those of reactor 2 and 7, were positioned on the lower left quadrant while the communities of reactor 6 and 8, were found in close proximity on the lower right quadrant of the plot. However, the nucleotide compositions of reactor 11 were singly located on the upper right quadrant of the plot. (PDF)

S46 Fig. Stacked barchat showing four *Basidiomycota* classes, the relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition of reactor 4 and 9 were found in close proximity, positioned on the upper right quadrant of the plot. The composition of reactor 3 and 11 were also closely located on the upper right quadrant, near the x-axis. The rest of the communities' nucleotide compositions in other treatments were distributed within the four plot quadrant. (PDF)

S47 Fig. Stacked barchat showing four *Basidiomycota* orders, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide compositions in the twelve treatments were dissimilar and were distributed within the four quadrants. (PDF)

S48 Fig. Stacked barchart showing the proportion of the *Agaricomycetes* order, relative abundances (a) and their PCoA plot based on the Euclidean model (b). The nucleotide composition affiliated to *Agaricomycete*s in reactor 1 and 3 were almost similar, positioned on the upper left quadrant of the plot. Their composition in reactor 7 and 9 were in close proximity, located in the lower left quadrant of the plot. The *Agaricomycete*s nucleotides in reactor 11 and 12 clustered partially on the lower left quadrant of the plot, those of reactor 5 were singly located in the upper right quadrant of the plot. (PDF)

S49 Fig. Stacked barchat showing the two *Chytridiomycota* classes, and their relative abundances in the two treatments. This is an indication of special substrates in the two treatments that were absent in other studied treatnents. (PDF)

S50 Fig. Stacked barchat showing the four *Chytridiomycota* orders, and their relative abundances in the two treatments. The orders are considered rare due to the fact that they were detected in only three treatments. (PDF)

S51 Fig. Stacked barchat showing the three orders affiliated to unclassified fungal nucleotide reads. The orders are considered rare due to the fact that they were detected in only three treatments. (PDF)

S52 Fig. The PCoA analysis revealing β-diversity of the twelve treatmenets at three taxa level: a) Domain, b) Phylum, and c) Class level. At the class level, the reads/nucleotide sequences of reactor 2 and 10 were partially similar, located in the lower left quadrant, those of reactor 8 and 11 were located on the upper right quadrant, reactor 4 and 9 compositions on the lower right quadrant, while those of reactor 3 and 6 were located on the lower left quadrant of the plot. (PDF)

Acknowledgments

Author Contributions

Conceptualization: Samuel Mwangangi Muturi.

Data curation: Samuel Mwangangi Muturi.

Formal analysis: Samuel Mwangangi Muturi.

Funding acquisition: Samuel Mwangangi Muturi.

Investigation: Samuel Mwangangi Muturi.

Methodology: Samuel Mwangangi Muturi, Stephen Obol Opiyo.

Project administration: Samuel Mwangangi Muturi.

Software: Samuel Mwangangi Muturi.

Validation: Samuel Mwangangi Muturi.

Visualization: Samuel Mwangangi Muturi.

Writing – original draft: Samuel Mwangangi Muturi.

Writing – review & editing: Samuel Mwangangi Muturi, Lucy Wangui Muthui, Paul Mwangi Njogu, Justus Mong'are Onguso, Francis Nyamu Wachira, Stephen Obol Opiyo, Roger Pelle.

References

1. Calusinska M, Goux X, Fossépré M, Muller E, Wilmes P Delfosse P. 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. 2018; 11(196): 1–19. https://doi.org/10.1186/s13068-018-1195-8 PMID: 30038663

2. Weiland P. Biogas production: Current state and perspectives. Appl. Microbiol. Biotechnol. 2010; 85 (4): 849–860. https://doi.org/10.1007/s00253-009-2246-7 PMID: 19777226.

3. Verstraete W. The manufacturing microbe. Microbial Biotechnol. 2015; 8(1): 36–37. https://doi.org/10.1111/1751-7915.12183 PMID: 25546061.

4. Austin G, Morris G. Biogas production in Africa, in: Janssen R., Rutz D. (Eds.), Bioenergy Sustain. Dev. Africa, Springer, Dordrecht, 2012.

5. Scarlat N, Dallemand J-F, Fahl F. Biogas: Developments and perspectives in Europe. Renew. Energ. 2018; 129: 457–472. https://doi.org/10.1016/j.renene.2018.03.006.

6. Tiwary A, Williams I, Pant D, Kishore W. Emerging perspectives on environmental burden minimisation initiatives from anaerobic digestion technologies for community scale biomass valorisation. Renew. Sustain. Energ. Rev. 2015; 42: 883–901. https://doi:10.1016/j.rser.2014.10.052.

7. Africa Biogas Partnership Programme, (2017). http://africabiogas.org/

8. Mizrahi-Man O, Davenport E, Gilad Y. Taxonomic classification of *Bacterial* 16S rRNA genes using short sequencing reads: Evaluation of effective study designs. PLoS one. 2013; 8(1): e53608. https://doi.org/10.1371/journal.pone.0053608 PMID: 23308262

9. Campanaro S, Treu L, Kougias G, Zhu X, Angelidaki I. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. Scient. Rep. 2018; 8: 1926. https://doi.org/10.1038/s41598-018-20414-0 PMID: 29386622

10. Fox G, Wisotzkey J, Jurtshuk P. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 1992; 42(1): 166–170. ID: https://doi.org/10.1099/00207713-42-1-166 PMID: 1371061.

11. Locey K, Lennon J. Scaling laws predict global microbial diversity. Proc. Natl. Acad. Sci. USA. 2016; 113: 5970–5975. https://doi.org/10.1073/pnas.1521291113 PMID: 27140646.

12. Locey K, Lennon J. Reply to Willis: Powerful predictions of biodiversity from ecological models and scaling laws. Proc. Natl. Acad. Sci. USA. 2016; 113(35): E5097–E5097. https://doi.org/10.1073/pnas.1609635113 PMID: 27512032.

13. Shoemaker W, Locey K, Lennon J. A macroecological theory of microbial biodiversity. Nat. Ecol. Evol. 2017; 1(5): 107. https://doi.org/10.1038/s41559-017-0107 PMID: 28812691.

14. Dykhuizen D. Santa Rosalia revisited: Why are there so many species of Bacteria? Antonie Van Leeuwenhoek. 1998; 73(1): 25–33. https://doi.org/10.1023/a:1000665216662 PMID: 9602276.

15. Schloss P, Girard R, Martin T, Edwards J, Thrash J. Status of the *Archaeal* and *Bacterial* census: An update. MBio. 2016; 7(3): e00201–16. https://doi.org/10.1128/mBio.00201-16 PMID: 27190214

16. Amann R, Rosselló-Móra R. After all only millions? MBio. 2016; 7(4): e000999–16. https://doi.org/10.1128/mBio.00999-16 PMID: 27381294.

17. Mora C, Tittensor D, Adl S, Simpson A, Worm B. How many species are there on earth and in the Ocean? PLoS Biol. 2011; 9(8): e1001127. https://doi.org/10.1371/journal.pbio.1001127 PMID: 21886479

18. Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K, et al. Initial community evenness favours functionality under selective stress. Nat. 2009; 458(7238): 623–626. https://doi.org/10.1038/nature07840 PMID: 19270679.

19. Carballa M, Regueiro L, Lema J. Microbial management of anaerobic digestion: Exploiting the microbiome-functionality nexus. Curr. Opin. Biotechnol. 2015; 33: 103–11. https://doi.org/10.1016/j.copbio.2015.01.008 PMID: 25682574.

20. Huse S, Huber J, Morrison H, Sogin M, Welch D. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 2007; 8(7): R143. https://doi.org/10.1186/gb-2007-8-7-r143 PMID: 17659080; PMCID: PMC2323236.

21. Gomez-Alvarez V, Teal T, Schmidt T. Systematic artifacts in metagenomes from complex microbial communities. The ISME J. 2009; 3(11): 1314–1317. https://doi.org/10.1038/ismej.2009.72 PMID: 19587772.

22. Hughes R, Inouye B, Johnson M, Underwood N, Vellend M. Ecological consequences of genetic diversity. Ecol. Lett. 2008; 11(6): 609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x PMID: 18400018.

23. Legendre P, De Caceres M. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecol. Lett. 2013; 16(8):951–963. https://doi.org/10.1111/ele.12141 PMID: 23809147.

24. Overbeek R, Begley T, Butler R, Choudhuri J, Diaz N, Chuang H-Y, et al. The Subsystems approach to genome annotation and its use in the project to annotate 1000 Genomes. Nucl. Acids Res. 2005; 33(17):5691–5702. PMCID: PMC1251668; https://doi.org/10.1093/nar/gki866 PMID: 16214803.

25. Luo G, Fotidis I, Angelidaki I. 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. 2016; 9(51): 1–12. https://doi:10.1186/s13068-016-0465-6.

26. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman J, et al. Unlocking the potential of metagenomics through replicated experimental design. Nat. Biotechnol. 2012; 30(6):513–520. https://doi.org/10.1038/nbt.2235 PMID: 22678395.

27. Stolze Y, Zakrzewski M, Maus I, Eikmeyer F, Jaenicke S, Rottmann N, et al. Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. BMC Biotechnol. Biofuel. 2015; 8(14): 1–18. https://doi.org/10.1186/s13068-014-0193-8 PMID: 25688290

28. Lynd L, Weimer P, van Zyl W, Pretorius I. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. 2002; 66(4): 506–577. PMCID: PMC134660. https://doi.org/10.1128/mmbr.66.3.506-577.2002 PMID: 12209002

29. Gold D, Martin J. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. J. Bacteriol. 2007; 189(19): 6787–6795. https://doi.org/10.1128/JB.00882-07 PMID: 17644599

30. Seedorf H, Fricke W, Veith B, Brüggemann H, Liesegang H, Strittmatter A, et al. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. USA. 2008; 105(6): 2128–2133. https://doi.org/10.1073/pnas.0711093105 PMID: 18218779

31. Kaji M, Taniguchi Y, Matsushita O, Katayama S, Miyata S, Morita S, et al. The *hydA* gene encoding the H₂ evolving hydrogenase of *Clostridium perfingens*: Molecular characterization and expression of the gene. FEMS Microbiol. Lett. 1999; 181(2): 329–336. https://doi.org/10.1111/j.1574-6968.1999.tb08863.x PMID: 10585557.

32. Fonknechten N, Chaussonnerie S, Tricot S, Lajus A, Andreesen J, Perchat N, et al. *Clostridium sticklandii*, a specialist in amino acid degradation: Revisiting its metabolism through its genome sequence. BMC Genomic. 2010; 11:555. https://doi.org/10.1186/1471-2164-11-555 PMID: 20937090; PMCID: PMC3091704.

33. Wirth R, Kovács E, Maróti G, Bagi Z, Rákhely G, Kovács KL. Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. BMC Biotechnol. Biofuel. 2012; 5(41):1–16. https://doi.org/10.1186/1754-6834-5-41 PMID: 22673110; PMCID: PMC3395570.

34. Duncan S, Hold G, Harmsen H, Stewart C, Flimt H. Growth requirements and fermentation products of *Fusobacterium prausnitzii* and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov. comb. nov. Int. J. Syst. Evol. Microbiol. 2002; 52(Pt 6): 2141–2146. https://doi.org/10.1099/00207713-52-6-2141 PMID: 12508881.

35. Nokaga H, Keresztes G, Shinoda Y, Ikenaga Y, Abe M, Naito K, et al. Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense Y51* and comparison with *Dehalococcoides ethenogenes*. J. Bacteriol. 2006; 188(6): 2262–2274. PMCID: PMC1428132; https://doi.org/10.1128/JB.188.6.2262-2274.2006 PMID: 16513756.

36. Debarati P, Frank W, Arick T, Bridges S, Burgess S, Yoginder S, et al. Genome sequence of the solvent-producing bacterium *Clostridium carboxidivorans* strain P7. J. Bacteriol. 2010; 192(20): 5554–5555. https://doi.org/10.1128/JB.00877-10 PMID: 20729368; PMCID: PMC2950491.

37. Oelgeschläger E, Rother M. Carbon monoxide-dependent energy metabolism in anaerobic Bacteria and *Archaea*. Arch. Microbiol. 2008; 190(3): 257–269. https://doi.org/10.1007/s00203-008-0382-6 PMID: 18575848.

38. Simankova M, Chernych N, Osipov G, Zavarzin G. *Halocella cellulolytica* gen. nov., sp. nov., a new obligately anaerobic, halophilic, cellulolytic *Bacterium*. Syst. Appl. Microbiol. 1993; 16(3): 385–389. https://doi.org/10.1016/S0723-2020(11)80270-5.

39. Wei C, Kunio O, Shoichi S. Intergeneric protoplast fusion between *Fusobacterium varium* and *Enterococcus faecium* for enhancing dehydrodivanillin degradation. Appl. Environ. Microbiol. 1987; 53(3): 542–548. PMCID: PMC203703; https://doi.org/10.1128/AEM.53.3.542-548.1987 PMID: 3579269.

40. Guerra N, Fajardo P, Fuciños C, Amado R, Alonso E, Torrado A, et al. Modelling the biphasic growth and product formation by *Enterococcus faecium* CECT 410 in realkalized fed-batch fermentations in whey. J. Biomed. Biotechnol. 2010; ID: 290286:1–16. https://doi.org/10.1155/2010/290286 PMID: 20689729

41. Enterococcus faecalis hydrogenase 3. http://biocyc.org/EFAE749511-HMP/NEWIMAGE?object=Hydrogenase-3.

42. de Bok F, Stams A, Dijkema C, Boone D. Pathway of propionate oxidation by a syntrophic culture of *Smithella propionica* and *Methanospirillum hungatei*. Appl. Environ. Microbiol. 2001; 67(4): 1800–1804. PMCID: PMC92800; https://doi.org/10.1128/AEM.67.4.1800-1804.2001 PMID: 11282636.

43. Schmidt A, Muller N, Schink B, Schleheck D. A proteomic view at the biochemistry of syntrophic butyrate oxidation in *Syntrophomonas wolfei*. PLoS one. 2013; 8(2): e56905. https://doi.org/10.1371/journal.pone.0056905 PMID: 23468890

44. Dubbels B, Sayavedra-Soto L, Bottomley P, Arp D. *Thauera butanivorans* sp. nov., a C2- C9 alkane-oxidizing bacterium previously referred to as '*Pseudomonas butanovora*'. Int. J. Syst. Evol. Microbiol. 2009; 59(Pt 7): 1576–1578. https://doi.org/10.1099/ijs.0.000638-0 PMID: 19528200; PMCID: PMC2889399.

45. Baena S, Fardeau M, Labat M, Ollivier B, Thomas P, Garcia J, et al. *Aminobacterium colombiense* gen. nov. sp. nov., an amino acid-degrading anaerobe isolated from anaerobic sludge. Anaerobe. 1998; 4(5): 241–250. https://doi.org/10.1006/anae.1998.0170 PMID: 16887649

46. Chertkov O, Sikorski J, Brambilla E, Lapidus A, Copeland A, Glavina Del Rio T, et al. Complete genome sequence of *Aminobacterium colombiense* type strain (ALA-1). Stand. Genomic. Sci. 2010; 2(3): 280–289. PMCID: PMC3035294; https://doi.org/10.4056/sigs.902116 PMID: 21304712.

47. Ushijima T, Ozaki Y. Potent antagonism of *Escherichia coli*, *Bacteroides ovatus*, *Fusobacterium varium*, and *Enterococcus faecalis*, alone or in combination, for enteropathogens in anaerobic continuous flow cultures. J. Med. Microbiol. 1986; 22(2), 157–163. https://doi.org/10.1099/00222615-22-2-157 PMID: 2875188

48. Kindaichi T, Ito T, Okabe S. Ecophysiological interaction between nitrifying *Bacteria* and heterotrophic *Bacteria* in autotrophic nitrifying biofilms as determined by microautoradiography-fluorescence in situ

hybridization. Appl. Environ. Microbiol. 2004; 70(3): 1641–1650. https://doi.org/10.1128/aem.70.3.1641-1650.2004 PMID: 15006789; PMCID: PMC368389.

49. Bjursel M, Martens E, Gordon J. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J. Biol. Chem. 2006; 281(47): 36269–36279. https://doi.org/10.1074/jbc.M606509200 PMID: 16968696.

50. Li H, Shen T, Wang X, Lin K, Liu Y, Lu S, et al. Biodegradation of perchloroethylene and chlorophenol co-contamination and toxic effect on activated sludge performance. Bioresour. Technol. 2013; 137: 286–93. https://doi.org/10.1016/j.biortech.2013.02.050 PMID: 23587831.

51. Xie G, Bruce D, Challacombe J, Chertkov O, Detter J, Gilna P, et al. Genome sequence of the cellulo-lytic gliding bacterium, *Cytophaga hutchinsonii*. Appl. Environ. Microbiol. 2007; 73(11): 3536–3546. PMCID: PMC1932680; https://doi.org/10.1128/AEM.00225-07 PMID: 17400776.

52. Hiras J, Wu Y-W, Eichorst S, Simmons B, Singer S. Refining the phylum *Chlorobi* by resolving the phy-logeny and metabolic potential of the representative of a deeply branching, uncultivated lineage. The ISME J. 2016; 10(4): 833–845. https://doi.org/10.1038/ismej.2015.158 PMID: 26325358; PMCID: PMC4796922.

53. Wahlund T, Madigan M. Nitrogen-fixation by the thermophilic green sulfur bacterium *Chlorobium tepi-dum*. J. Bacteriol. 1993; 175(2): 474–478. https://doi.org/10.1128/jb.175.2.474-478.1993 PMCID: PMC196162; PMID: 8093448.

54. Wahlund T, Tabita F. The reductive tricarboxylic acid cycle of carbon dioxide assimilation: Initial stud-ies and purification of ATP-citrate lyase from the green sulfur bacterium *Chlorobium tepidum*. J. Bac-teriol. 1997; 179(15): 4859–4867. PMCID: PMC179334; https://doi.org/10.1128/jb.179.15.4859-4867.1997 PMID: 9244275.

55. Ghilarov D, Serebryakova M, Stevenson C, Hearnshaw S, Volkov D, Maxwell A, et al. The origins of specificity in the microcin-processing protease TldD/E. Structure. 2017; 25(10): 1549–1561.e5. PMCID: PMC5810440; https://doi.org/10.1016/j.str.2017.08.006 PMID: 28943336.

56. Moon C, Young W, Maclean P, Cookson A, Bermingham E. Metagenomic insights into the roles of *Proteobacteria* in the gastrointestinal microbiomes of healthy dogs and cats. *MicrobiologyOpen*. 2018; 7(5):e00677. https://doi.org/10.1002/mbo3.677 PMID: 29911322; PMCID: PMC6182564.

57. Han K, Lim HC, Hong J. Acetic acid formation in *Escherichia coli* fermentation. Biotechnol. Bioeng. 1992; 39(6):663–71. https://doi.org/10.1002/bit.260390611 PMID: 18600996.

58. Textor S, Wendisch V, de Graaf A, Mu¨ller U, Linder M, Linder D. Propionate oxidation in *Escherichia coli*: Evidence for operation of a methylcitrate cycle in Bacteria. Arch. Microbiol. 1997; 168(5): 428–436. https://doi.org/10.1007/s002030050518 PMID: 9325432.

59. Hippe H, Hagelstein A, Kramer I, Swiderski J, Stackebrandt E. Phylogenetic analysis of *Formivibrio citricus*, *Propionivibrio dicarboxylicus*, *Anaerobiospirillum thomasii*, *Succinimonas amylolytica* and *Succinivibrio dextrinosolvens* and proposal of Succinivibrionaceae fam. nov. Int. J. Syst. Bacteriol. 1999; 49(Pt 2): 779–782. https://doi.org/10.1099/00207713-49-2-779 PMID: 10319502.

60. Greetham H, Collins M, Gibson G, Giffard C, Falsen E, Lawson P. *Sutterella stercoricanis* sp. nov., isolated from canine faeces. Int. J. Syst. Evol. Microbiol. 2004; 54(Pt 5): 1581–1584. https://doi.org/10.1099/ijs.0.63098-0 PMID: 15388713.

61. Hottes AK, Meewan M, Yang D, Arana N, Romero P, McAdams H, et al. Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. J. Bacteriol. 2004; 186(5): 1448–1461. https://doi.org/10.1128/jb.186.5.1448-1461.2004 PMID: 14973021; PMCID: PMC344409.

62. Müller V, Grüber G. ATP synthases: Structure, function and evolution of unique energy converters. Cell Mol. Life Sci. 2003; 60(3): 474–494. https://doi.org/10.1007/s000180300040 PMID: 12737308.

63. Van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, De Voogd NJ, et al. Global Diversity of Sponges (*Porifera*). PLoS one. 2012; 7(4): e35105. https://doi.org/10.1371/journal.pone.0035105 PMCID: PMC3338747; PMID: 22558119.

64. Mclnernery M, Mackie I, Bryant M. Syntrophic association of butyrate degrading *Bacterium* and *Metha-nosarcina* enriched from bovine rumen fluid. Appl. Environ. Microbiol. 1981; 41(3): 826–828. https://doi.org/10.1128/AEM.41.3.826-828.1981 PMID: 7224635; PMCID: PMC243782.

65. Meher K, Ranade D. Isolation of propionate degrading *Bacterium* in co-culture with a methanogen from a cattle dung biogas plant. J. Biosci. 1993; 18(2): 271–277. https://doi.org/10.1007/BF02703124.

66. Boone D, Bryant M. Propionate-degrading bacterium, *Syntrophobacter wolinii* sp. nov. gen. nov., from methanogenic ecosystems. Appl. Environ. Microbiol. 1980; 40(3): 626–632. https://doi.org/10.1128/AEM.40.3.626-632.1980 PMID: 16345640; PMCID: PMC291629.

67. Mountfort D, Bryant M. Isolation and characterization of an anaerobic syntrophic benzoate-degrading *Bacterium* from sewage sludge. Arch. Microbiol. 1982; 133: 249–256. https://doi.org/10.1007/BF00521285.

68. Spring S, Sorokin D, Verbarg S, Rohde M, Woyke T, Kyrpides N. Sulfate-reducing *Bacteria* that produce exopolymers thrive in the calcifying zone of a Hypersaline *Cyanobacterial* Mat. Front. Microbiol. 2019; 10:862. https://doi.org/10.3389/fmicb.2019.00862 PMID: 31068923; PMCID: PMC6491731.

69. Reichenbach H, Dworkin M. The *Myxobacteria*. In: Starr KP, Stolp H, Truper HG, Belows A, Schlegel HG, Editors.The Prokaryotes: A Handbook on Habitats, Isolation and Identification of Bacteria. New York, Springer, 1991; p. 3415–3487.

70. Biology Dawid W. and global distribution of *Myxobacteria* in soils. FEMS Microbiol. Rev. 2000; 24(4): 403–427. https://doi.org/10.1111/j.1574-6976.2000.tb00548.x PMID: 10978544

71. Vander Roost J, Daae FL, Steen IH, Thorseth IH, Dahle H. Distribution patterns of iron-oxidizing *Zeta*- and *Beta-Proteobacteria* from different environmental settings at the Jan Mayen vent field. Front. Microbiol. 2018; 9: 3008. PMCID: PMC6292416; https://doi.org/10.3389/fmicb.2018.03008 PMID: 30574135.

72. Vander Roost J, Thorseth IH, Dahle H. Microbial analysis of *Z-proteobacteria* and co-colonizers of iron mats in the Troll Wall Vent Field, Arctic Mid-Ocean Ridge. PLoS one. 2017; 12(9): e0185008. https://doi.org/10.1371/journal.pone.0185008 PMID: 28931087

73. Hoshino T, Kuratomi T, Morono Y, Hori T, Oiwane H, Kiyokawa S, et al. Ecophysiology of *Z-Proteobacteria* associated with shallow hydrothermal iron-oxyhydroxide deposits in Nagahama Bay of Satsuma Iwo-Jima, Japan. Front. Microbiol. 2016; 6(11): 1554. https://doi.org/10.1111/j.1574-6976.2000.tb00548.x PMID: 10978544.

74. Pankratov T, Dedysh S. *Granulicella paludicola* gen. nov., sp. nov., *Granulicella pectinivorans* sp. nov., *Granulicella aggregans* sp. nov. and *Granulicella rosea* nov., acidophilic, polymer-degrading *Acidoacteria* from Sphagnum peat bogs. Int. J. Syst. Evol. Microbiol. 2010; 60(Pt 12): 2951–2959. https://doi.org/10.1099/ijs.0.021824-0 PMID: 20118293.

75. Eichorst S, Trojan D, Roux S, Herbold C, Rattei T, Woebken D. Genomic insights into the *Acidobacteria* reveal strategies for their success in terrestrial environments. Environ. Microbiol. 2018; 20(3): 1041–1063. https://doi.org/10.1111/1462-2920.14043 PMID: 29327410; PMCID: PMC5900883.

76. Greening C, Carere C, Rushton-Green R, Harold L, Hards K, Taylor M, et al. Persistence of the dominant soil phylum *Acidobacteria* by trace gas scavenging. Proc. Natl. Acad. Sci. USA. 2015; 112(33): 10497–10502. https://doi.org/10.1073/pnas.1508385112 PMID: 26240343; PMCID: PMC4547274.

77. Lazarev V, Levitskii S, Basovskii Y, Chukin M, Akopian T, Vereshchagin V, et al. Complete genome and proteome of *Acholeplasma laidlawii*. J. Bacteriol. 2011; 193(18): 4943–4953. https://doi.org/10.1128/JB.05059-11 PMID: 21784942; PMCID: PMC3165704.

78. Siewert C, Hess WR, Duduk B, Huettel B, Reinhardt R, Büttner C, et al. Complete genome determination and analysis of *Acholeplasma oculi* strain 19L, highlighting the loss of basic genetic features in the *Acholeplasmataceae*. BMC Genomic. 2014; 15(1): 931. PMCID: PMC4221730; https://doi.org/10.1186/1471-2164-15-931 PMID: 25344468.

79. Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald F, Chater F, et al. Genomics of *Actinobacteria*: Tracing the evolutionary history of an ancient phylum. Microbiol.Mol. Biol Rev. 2007; 71(3): 495–548. PMCID: PMC2168647; https://doi.org/10.1128/MMBR.00005-07 PMID: 17804669.

80. Schell M, Karmrabűntzon M, Snel B, Vilanova D, Berger B, Pessi G, et al. The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. Proc. Natl. Acad. Sci. USA. 2002; 99(22):14422–14427. https://doi.org/10.1073/pnas.212527599 PMID: 12381787; PMCID: PMC137899.

81. Zakrzewski M, Goesmann A, Jaenicke S, Jünemann S, Eikmeyer F, Szczepanowski R, et al. Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. J. Biotechnol. 2012; 158(4): 248–258. https://doi.org/10.1016/j.jbiotec.2012.01.020 PMID: 22342600.

82. Fuerst J, Sagulenko E. Beyond the bacterium: *Planctomycetes* challenge our concepts of microbial structure and function. Nat. Rev. Microbiol. 2011; 9(6): 403–413. https://doi.org/10.1038/nrmicro2578 PMID: 21572457.

83. Kampmann K, Ratering S, Kramer I, Schmidt M, Zerr W, Schnell S. Unexpected stability of *Bacteroidetes* and *Firmicutes* communities in laboratory biogas reactors fed with different defined substrates. Appl. Environ. Microbiol. 2012; 78(7): 2106–2119. https://doi.org/10.1128/AEM.06394-11 PMID: 22247168; PMCID: PMC3302617.

84. Krakat N, Westphal A, Schmidt S, Scherer P. Anaerobic digestion of renewable biomass: Thermophilic temperature governs methanogen population dynamics. Appl. Environ. Microbiol. 2010; 76(6): 1842–1850. https://doi.org/10.1128/AEM.02397-09 PMID: 20097828.

85. Weiss A, Jérôme V, Freitag R, Mayer H. Diversity of the resident microbiota in a thermophilic municipal biogas plant. Appl. Microbiol. Biotechnol. 2008; 81(1): 163–173. https://doi.org/10.1007/s00253-008-1717-6 PMID: 18820906.

86. Maus I, Cibis K, Wibberg D, Winkler A, Stolze Y, König H, et al. Complete genome sequence of the strain *Defluviitoga tunisiensis* L3, isolated from a thermophilic, production-scale biogas plant. J. Biotechnol. 2015; 203: 17–18. https://doi.org/10.1016/j.jbiotec.2015.03.006 PMID: 25801333.

87. Xia Y, Wang Y, Fang H, Jin T, Zhong H, Zhang T. Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. Scient Rep. 2014; 4: 6708. https://doi.org/10.1038/srep06708 PMID: 25330991; PMCID: PMC4204047.

88. Wang W, Xie L, Luo G, Zhou Q, Angelidaki I. Performance and microbial community analysis of the anaerobic reactor with coke oven gas biomethanation and in situ biogas upgrading. Bioresour. Technol. 2013; 146: 234–239. https://doi.org/10.1016/j.biortech.2013.07.049 PMID: 23941705.

89. Siriwongrungson V, Zeng R, Angelidaki I. Homoacetogenesis as the alternative pathway for H_2 sink during thermophilic anaerobic degradation of butyrate under suppressed methanogenesis. Water Res. 2007; 41(18): 4204–4210. https://doi.org/10.1016/j.watres.2007.05.037 PMID: 17618671.

90. Pohlschroeder M, Leschine SB, Canale-Parola E. *Spirochaeta caldaria* sp. nov., a thermophilic bacterium that enhances cellulose degradation by *Clostridium thermocellum*. Arch. Microbiol. 1994; 161: 17–24. https://doi.org/10.1007/BF00248889.

91. Bispo P, Balzter H, Malhi Y, Slik J, dos Santos J, Renno´ C, et al. Drivers of metacommunity structure diverge for common and rare Amazonian tree species. PLoS one. 2017; 12(11): e0188300. https://doi.org/10.1371/journal.pone.0188300 PMID: 29155865.

92. Eichorst S, Kuske C. Identification of cellulose-responsive *Bacterial* and *Fungal* communities in geographically and edaphically different soils by using stable isotope probing. Appl. Environ. Microbiol. 2012; 78(7): 2316–2327. https://doi.org/10.1128/AEM.07313-11 PMCID: PMC330259; PMID: 22287013.

93. Appel J, Schulz R. Hydrogen metabolism in organisms with oxygenic photosynthesis: Hydrogenases as important regulatory devices for a proper redox poising? J. Photochem. Photobiol B: Biol. 1998; B47(1): 1–11.

94. Bergman B, Gallon J, Rai A, Stal L. N_2 fixation by non-heterocystous *Cyanobacteria*. FEMS Microbiol. Rev. 1997; 19(3): 139–185. https://doi.org/10.1111/j.1574-6976.1997.tb00296.x.

95. Lindblad P. *Cyanobacterial* H_2-metabolism: Knowledge and potential/strategies for a photo-biotechnological production of H_2. Biotecnol. Appl. 1999; 16(3): 141–144.

96. Howarth D, Codd A. The uptake and production of molecular hydrogen by unicellular *Cyanobacteria*. J. Gen. Microbiol. 1985; 131(7):1561–1569. https://doi.org/10.1099/00221287-131-7-1561.

97. Weisshaar H, Böger P. Pathways of hydrogen uptake in the *Cyanobacterium Nostoc muscorum*. Arch. Microbiol. 1985; 142: 349–353. https://doi.org/10.1007/BF00491902.

98. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, et al. Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. The ISME J. 2011; 5(1): 61–70. https://doi.org/10.1038/ismej.2010.95 PMID: 20613790; PMCID: PMC3105677.

99. Pelletier E, Kreimeyer A, Bocs S, Rouy Z, Gyapay G, Chouari R, et al. "*Candidatus Cloacamonas acidaminovorans*": Genome sequence reconstruction provides a first glimpse of a new Bacterial division. J. Bacteriol. 2008; 190(7): 2572–2579. https://doi.org/10.1128/JB.01248-07 PMID: 18245282

100. Guiavarch E, Pons A, Christophe G, Creuly C, Dussap CG. Analysis of a continuous culture of *Fibrobacter succinogenes S85* on a standardized glucose medium. Bioprocess Biosyst. Eng. 2010; 33(4): 417–425. https://doi.org/10.1007/s00449-009-0342-9 PMID: 19548008.

101. Ye Q, Roh Y, Carroll S, Blair B, Zhou J, Zhang C, et al. Alkaline anaerobic respiration: Isolation and characterization of a novel alkaliphilic and metal-reducing *Bacterium*. Appl. Environ. Microbiol. 2004; 70(9): 5595–5602. PMCID: PMC520920. https://doi.org/10.1128/AEM.70.9.5595-5602.2004 PMID: 15345448

102. Fisher E, Dawson A, Polshyna G, Lisak J, Crable B, Detter J, et al. Transformation of inorganic and organic arsenic by *Alkaliphilus oremlandii* sp. nov. strain OhILAs. Ann. NY. Acad. Sci. 2008; 1125(1): 230–241. https://doi.org/10.1196/annals.1419.006 PMID: 18378595.

103. Daims H, Lebedeva E, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by *Nitrospira Bacteria*. Nat. 2015; 528(7583): 504–509. https://doi.org/10.1038/nature16461 PMID: 26610024.

104. Koch H, van Kessel M, Lücke S. Complete nitrification: Insights into the ecophysiology of comammox *Nitrospira*. Appl. Microbiol. Biotechnol. 2019; 103(1):177–189. https://doi.org/10.1007/s00253-018-9486-3 PMID: 30415428; PMCID: PMC6311188.

105.   Kotsyurbenko O, Glagolev M, Nozhevnikova A, Conrad R. Competition between homoacetogenic *Bacteria* and methanogenic *Archaea* for hydrogen at low temperature. FEMS Microbiol. Ecol. 2001; 38(2–3): 153–159. https://doi.org/10.1111/j.1574-6941.2001.tb00893.x.

106.   Thauer R, Kaster A, Seedorf H, Buckel W, Hedderich R. Methanogenic *Archaea*: Ecological relevant differences in energy conservation. Nat. Rev. Microbiol. 2008; 6(8): 579–591. https://doi.org/10.1038/nrmicro1931 PMID: 18587410.

107.   Deppenmeier U, Müller V. Life close to the thermodynamic limit: How methanogenic *Archaea* conserve energy. Result and Problem Cell Different. 2008; 45:123–152. https://doi.org/10.1007/400_2006_026 PMID: 17713742.

108.   Rother M, Melchaf W. Anaerobic growth of *Methanosarcina acetivorans C2A* on carbon monoxide: An unusual way of life for a methanogenic archaeon. *Proc. Natl Acad. Sci. USA*. 2004; 101(48): 16929–16934. PMCID: PMC529327; https://doi.org/10.1073/pnas.0407486101 PMID: 15550538.

109.   Anderson I, Sieprawska-Lupa M, Goltsman E, Lapidus A, Copeland A, Glavina Del Rio T, et al. Complete genome sequence of *Methanocorpusculum labreanum* type strain Z. Stand. Genomic Sci. 2009; 1(2):197–203. PMCID: PMC3035222; https://doi.org/10.4056/sigs.35575 PMID: 21304657.

110.   Gilmore S, Henske J, Sexton J, Solomon K, Seppälä S, Yoo J, et al. Genomic analysis of methanogenic *Archaea* reveals a shift towards energy conservation. BMC Genomic. 2017; 18: 639. https://doi.org/10.1186/s12864-017-4036-4 PMID: 28826405

111.   Zellner G, Stackebrandt E, Messner P, Tindall B, Demacario E, Kneifel H, et al. *Methanocorpusculaceae* fam-nov, represented by *Methanocorpusculum parvum*, *Methanocorpusculum sinense* spec-nov and *Methanocorpusculum bavaricum* spec-nov. Arch. Microbiol. 1989; 151: 381–90. https://doi.org/10.1007/BF00416595 PMID: 2742452

112.   Anderson I, Sieprawska-Lupa M, Lapidus A, Nolan M, Copeland A, Glavina Del Rio T, et al. Complete genome sequence of *Methanoculleus marisnigri* Romesser *et al.* 1981 type strain JR1. Stand. Genomic Sci. 2009; 1(2): 189–196. PMCID: PMC3035220; https://doi.org/10.4056/sigs.32535 PMID: 21304656.

113.   Rivera M, Lake J. The phylogeny of *Methanopyrus kandleri*. Int. J.Syst. Bacteriol. 1996; 46(1): 348–351. https://doi.org/10.1099/00207713-46-1-348 PMID: 8573519.

114.   Burggraf S, Fricke H, Neuner A, Kristjansson J, Rouvier P, Mandelco L, et al. *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow submarine hydrothermal system. Syst. Appl. Microbiol. 1990; 13(3): 263–269. https://doi.org/10.1016/s0723-2020(11)80197-9 PMID: 11538305

115.   Ding J, Chen S, Lai M, Liao T. *Haloterrigena mahii* sp. nov., an extremely halophilic archaeon from a solar saltern. Int. J. Syst. Evol. Microbiol. 2017; 67(5): 1333–1338. https://doi.org/10.1099/ijsem.0.001811 PMID: 28109205

116.   Oren A. The order *Halobacteriales*. In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. A Handbook on the Biology of Bacteria. New York, Springer; 2006; p. 113–164.

117.   Huber H, Stetter K. *Thermoplasmatales*. In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. A Handbook on the Biology of Bacteria. New York, Springer; 2006; p. 101–112.

118.   Lang K, Schuldes J, Kling A, Poehlein A, Daniel R, Brunea A. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of "*Candidatus Methanoplasma termitum*". Appl. Environ. Microbiol. 2015; 81(4): 1338–52. https://doi.org/10.1128/AEM.03389-14 PMID: 25501486; PMCID: PMC4309702.

119.   Gorlas A, Alain K, Bienvenu N, Geslin C. *Thermococcus prieurii* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Int. J. Syst. Evol. Microbiol. 2013; 63(Pt 8): 2920–2926. https://doi.org/10.1099/ijs.0.026419-0 PMID: 23355695.

120.   Cossu M, Da-Cunha V, Toffano-Nioche C, Forterre P, Oberto J. Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged *Thermococcales* chromosomes. Biochimie. 2015; 118: 313–321. PMCID: PMC4640148; https://doi.org/10.1016/j.biochi.2015.07.008 PMID: 26166067.

121.   Brasen C, Esser D, Rauch B, Siebers B. Carbohydrate metabolism in *Archaea*: Current insights into unusual enzymes and pathways and their regulation. Microbiol. Mol. Biol. Rev. 2014; 78(1): 89–175. PMCID: PMC3957730; https://doi.org/10.1128/MMBR.00041-13 PMID: 24600042.

122.   Wang X, Gao Z, Xu X, Ruan L. Complete genome sequence of *Thermococcus sp*. Strain 4557, a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent area. *J. Bacteriol*. 2011; 193 (19): 5544–5545. https://doi.org/10.1128/JB.05851-11 PMID: 21914870; PMCID: PMC3187469.

123. Lee H, Kang S, Bae S, Lim J, Cho Y, Kim Y, et al. The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydotrophic metabolism. J. Bacteriol. 2008; 190(22): 7491–7499. https://doi.org/10.1128/JB.00746-08 PMID: 18790866

124. Hartzell P, Reed D. The genus *Archaeoglobus*. In Dworkin M, Falkow S, Rosenberg E, Schleifer K, Stackebrandt E, editors. The Prokaryotes. A Handbook on the Biology of *Bacteria*. New York; Springer, 2006; p. 82–100.

125. Kashefi K, Tor J, Holmes D, Gaw Van Praagh C, Reysenbach A-L, Lovley D. *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. 2002; 52(Pt 3): 719–728. https://doi.org/10.1099/00207713-52-3-719 PMID: 12054231.

126. Dai X, Wang H, Zhang Z, Li K, Zhang X, Mora-López M, et al. 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. 2016; 7(Pt 2): 1902. PMCID: PMC5127849; https://doi.org/10.3389/fmicb.2016.01902 PMID: 27965637.

127. Sakai H, Kurosawa N. Complete genome sequence of the *Sulfodiicoccus acidiphilus* strain HS-1T, the first crenarchaeon that lacks polB3, isolated from an acidic hot spring in Ohwaku-dani, Hakone, Japan. BMC Res. Note. 2019; 12(1): 444. PMCID: PMC6647314; https://doi.org/10.1186/s13104-019-4488-5 PMID: 31331368.

128. Berg I, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea*. *Sci*. 2007; 318(5857): 1782–1786. https://doi.org/10.1126/science.1149976 PMID: 18079405.

129. Siebers B, Tjaden B, Michalke K, Do¨rr C, Ahmed H, Zaparty M, et al. Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. J. Bacteriol. 2004; 186(7): 2179–2194. PMCID: PMC374391; https://doi.org/10.1128/jb.186.7.2179-2194.2004 PMID: 15028704.

130. Zaparty M, Tjaden B, Hensel R, Siebers B. The central carbohydrate metabolism of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: Pathways and insights into their regulation. Arch. Microbiol. 2008; 190(3): 231–245. https://doi.org/10.1007/s00203-008-0375-5 PMID: 18491075.

131. Cossu M, Badel C, Catchpole R, Gadelle D, Marguet E, Barbe V, et al. Flipping chromosomes in deep-sea *Archaea*. *PLoS Genet*. 2017; 13(6): e1006847. https://doi.org/10.1371/journal.pgen.1006847 PMID: 28628615

132. Prokofeva M, Kostrikina N, Kolganova T, Tourova T, Lysenko A, Lebedinsky A, et al. 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. 2009; 59(Pt 12): 3116–3122. https://doi.org/10.1099/ijs.0.010355-0 PMID: 19643887.

133. Mardanov A, Svetlitchnyi V, Beletsky A, Prokofeva M, Bonch-Osmolovskaya E, Ravin N, et al. 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. 2010; 76(16): 5652–5657. https://doi.org/10.1128/AEM.00599-10 PMID: 20581186; PMCID: PMC2918975.

134. Hallam S, Mincer T, Schleper C, Preston C, Roberts K, Richardson P, et al. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine *Crenarchaeota*. PLoS Biol. 2006; 4(4): e95. https://doi.org/10.1371/journal.pbio.0040095 PMID: 16533068

135. Hallam S, Konstantinidis K, Putnam N, Schleper C, Watanabe Y, Sugahara J, et al. Genomic analysis of the uncultivated marine Crenarchaeote *Cenarchaeum symbiosum*. Proc. Natl. Acad. Sci USA. 2006; 103(48): 18296–18301. https://doi.org/10.1073/pnas.0608549103 PMID: 17114289

136. Walker C, de la Torre J, Klotz M, Urakawa H, Pinel N, Arp D, et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine *CrenArchaea*. Proc. Natl. Acad. Sci USA. 2010; 107(19): 8818–8823. https://doi.org/10.1073/pnas.0913533107 PMID: 20421470

137. Hu¨gler M, Huber H, Stetter K, Fuchs G. Autotrophic CO_2 fixation pathways in *Archaea* (*Crenarchaeota*). Arch. Microbiol. 2003; 179(3): 160–173. https://doi.org/10.1007/s00203-002-0512-5 PMID: 12610721.

138. Alonso-Sa´ez L, Waller A, Mende D, Bakker K, Farnelid H, Yager PL, et al. Role for urea in nitrification by polar marine *Archaea*. Proc. Natl. Acad. Sci. USA. 2012; 109(44): 17989–17994. PMCID: PMC3497816; https://doi.org/10.1073/pnas.1201914109 PMID: 23027926.

139. Podar M, Podar M, Makarova K, Graham D, Wolf Y, Koonin E, et al. Insights into *Archaeal* evolution and symbiosis from the genomes of a Nanoarchaeon and its inferred Cren*archaeal* host from Obsidian

Pool, Yellowstone national park. Biol. Direct. 2013; 8: 9. https://doi.org/10.1186/1745-6150-8-9 PMID: 23607440; PMCID: PMC3655853.

140. Munson-McGee J, Field E, Bateson M, Rooney C, Stepanauskas R, Young M. *Nanoarchaeota*, their *sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone national park hot springs. Appl. Environ. Microbiol. 2015; 81(22): 7860–7868. PMCID: PMC4616950; https://doi.org/10.1128/AEM.01539-15 PMID: 26341207.

141. Wurch L, Giannone R, Belisle B, Swift C, Utturkar S, Hettich R, et al. Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. Nat. Commun. 2016; 7:12115. https://doi.org/10.1038/ncomms12115 PMID: 27378076

142. Pinto O, Silva T, Vizzotto C, Santana R, Lopes F, Silva B, et al. Genome-resolved metagenomics analysis provides insights into the ecological role of *Thaumarchaeota* in the Amazon river and its plume. BMC Microbiol. 2020; 20:13. https://doi.org/10.1186/s12866-020-1698-x PMID: 31941452

143. Bernalier A, Fonty G, Bonnemoy F, Gouet P. Degradation and fermentation of cellulose by the rumen anaerobic *Fungi* in axenic cultures or in association with cellulolytic *Bacteria*. Curr. Microbiol. 1992; 25: 143–148. https://doi.org/10.1051/rnd:19920402 PMID: 1418394

144. Costa V, Amorim M, Reis E, Quintanilha A, Moradas-Ferreira P. Mitochondrial superoxide dismutase is essential for ethanol tolerance of *Saccharomyces cerevisiae* in the post-diauxic phase. *Microbiol*. 1997; 143(Pt 5): 1649–1656. https://doi.org/10.1099/00221287-143-5-1649 PMID: 9168613.

145. Viigand K, Tammus K, Alamäe T. 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*. 2005; 5(11): 1019–1028. https://doi.org/10.1016/j.femsyr.2005.06.003 PMID: 16103021.

146. Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: Metabolism, Gene regulation and peroxisome homeostasis. Int. J. Microbiol. 2011; 2011:101298. https://doi.org/10.1155/2011/101298 PMID: 21754936; PMCID: PMC3132611.

147. McHardy AC, Fan L, Yu X. Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. DNA Res. 2017; 24(6): 635–648. https://doi.org/10.1093/dnares/dsx031 PMID: 28992048; PMCID: PMC5726479.

148. Ren X, Liu Y, Tan Y, Huang Y, Liu Z, Jiang X. Sequencing and functional annotation of the whole genome of *Shiraia bambusicola*. Gene, Genome, Genet. 2019; 10(1): 23–35. PMCID: PMC6945017; https://doi.org/10.1534/g3.119.400694 PMID: 31712259.

149. Urquhart A, Mondo S, Mäkelä M, Hane J, Wiebenga A, He G, et al. Genomic and genetic insights into a cosmopolitan fungus, *Paecilomyces variotii* (*Eurotiales*). Front. Microbiol. 2018; 9: 3058. PMCID: PMC6300479; https://doi.org/10.3389/fmicb.2018.03058 PMID: 30619145.

150. Ohm R, Feau N, Henrissat B, Schoch C, Horwitz B, Barry K, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. PLoS Pathog. 2012; 8(12):e1003037. https://doi.org/10.1371/journal.ppat.1003037 PMID: 23236275; PMCID: PMC3516569.

151. Li J, Yu L, Yang J, Dong L, Tian B, Yu Z, et al. New insights into the evolution of subtilisin-like serine protease genes in *Pezizomycotina*. BMC Evol. Biol. 2010; 10: 68. https://doi.org/10.1186/1471-2148-10-68 PMID: 20211028; PMCID: PMC2848655.

152. Murat C, Payen T, Noel B, Kuo A, Morin E, Chen J, et al. *Pezizomycetes* genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. Nat. Ecol. Evol. 2018; 2(12): 1956–1965. https://doi.org/10.1038/s41559-018-0710-4 PMID: 30420746

153. Larriba E, Martín-Nieto J, Lopez-Llorca L. 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. 2012; 58(7): 815–827. https://doi.org/10.1139/w2012-054 PMID: 22690687.

154. Voigt C, Scha¨fer W, Salomon S. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. the Plant J. 2005; 42(3): 364–375. https://doi.org/10.1111/j.1365-313X.2005.02377.x PMID: 15842622.

155. Olivieri F, Eugenia ZM, Oliva C, Covarrubias A, Casalongue C. Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. Eur. J. Plant Pathol. 2002; 108(1): 63–72. https://doi:10.1023/A:1013920929965.

156. Agrawa Y, Khatri I, Subramanian S, Shenoy B. Genome sequence, comparative analysis and evolutionary insights into chitinases of entomopathogenic fungus, *Hirsutella thompsonii*. Genome Biol. Evol. 2015; 7(3): 916–930. PMCID: PMC5322555; https://doi.org/10.1093/gbe/evv037 PMID: 25716828.

157. Yang Y, Liu B, Du X, Li P, Liang B, Cheng X, et al. Complete genome sequence and transcriptomics analyses reveal pigment biosynthesis and regulatory mechanisms in an industrial strain, *Monascus*

purpureus YY-1. Scient. Rep. 2014; 5: 8331. https://doi.org/10.1038/srep08331 PMID: 25660389; PMCID: PMC4321180.

158. Galagan J, Calvo S, Borkovich K, Selker E, Read N, Jaffe D, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. Nat. 2003; 422(6934): 859–868. https://doi.org/10.1038/nature01554 PMID: 12712197.

159. Zheng H, Zhong Z, Shi M, Zhang L, Lin L, Hong Y, et al. Comparative genomic analysis revealed rapid differentiation in the pathogenicity related gene repertoires between *Pyricularia oryzae* and *Pyricularia penniseti* isolated from a *Pennisetum grass*. BMC Genomic. 2018; 19(1): 927. https://doi.org/10.1186/s12864-018-5222-8 PMID: 30545292

160. Zhang N, Cai G, Price D, Crouch J, Gladieux P, Hillman B, et al. Genome wide analysis of the transition to pathogenic lifestyles in *Magnaporthales* fungi. Scient. Rep. 2018; 8(1): 5862. https://doi.org/10.1038/s41598-018-24301-6 PMID: 29651164

161. Bugos R, Sutherland J, Adler J. Phenolic compound utilization by the soft rot fungus *Lecythophora hoffmannii*. Appl. Environ. Microbiol. 1988; 54(7): 1882–1885. https://doi.org/10.1128/AEM.54.7.1882-1885.1988 PMID: 16347701; PMCID: PMC202766.

162. Leonhardt S, Büttner E, Gebauer A, Hofrichter M, Kellner H. Draft genome sequence of the *sordariomycete Lecythophora* (Coniochaeta) *hoffmannii* CBS 245.38. Genome Announc. 2018; 6(7): e01510–17. PMCID: PMC5814504; https://doi.org/10.1128/genomeA.01510-17 PMID: 29449401.

163. Lindahl B, de Boer W, Finlay R. Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. The ISME J. 2010; 4(7): 872–881. https://doi.org/10.1038/ismej.2010.19 PMID: 20220789.

164. Wang Q-M, Theelen B, Groenewald M, Bai F-Y, Boekhout T. *Moniliellomycetes* and *Malasseziomycetes*, two new classes in *Ustilaginomycotina*. Persoonia. 2014; 33(1): 41–47. PMCID: PMC4312936; https://doi.org/10.3767/003158514X682313 PMID: 25737592.

165. Sánchez C. Lignocellulosic residues: Biodegradation and bioconversion by fungi. Biotechnol. Advance. 2009; 27(2): 185–194. https://doi.org/10.1016/j.biotechadv.2008.11.001 PMID: 19100826.

166. James T, Letcher P, Longcore J, Mozley-Standridge S, Porter D, Powell M, et al. A molecular phylogeny of the flagellated fungi, (*Chytridiomycota*) and description of a new phylum (*Blastocladiomycota*). Mycologia. 2006; 98(6): 860–871. https://doi.org/10.3852/mycologia.98.6.860 PMID: 17486963.

167. Bart T, van de Vossenberg L, Warris S, Nguyen H, van Gent-Pelzer M, Joly D, et al. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. Scient. Rep. 2019; 9(1):8672. https://doi.org/10.1038/s41598-019-45128-9.

168. Karov S, Mamanazarova K, Popova O, Aleoshin V, James T, Mamkaeva M, et al. *Monoblepharidomycetes* diversity includes new parasitic and saprotrophic species with highly intronized rDNA. Fungal Biol. 2017; 121(8): 729–741. https://doi.org/10.1016/j.funbio.2017.05.002 PMID: 28705399

169. Van der Giezen M. Hydrogenosomes and mitosomes: Conservation and evolution of functions. J. Euk. Microbiol. 2009; 56(3): 221–231. https://doi.org/10.1111/j.1550-7408.2009.00407.x PMID: 19527349.

170. Embley T, van der Giezen M, Horner D, Dyal P, Foster P. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. Phil. Trans. R. Soc. B: Biol. Sci. 2002; 358(1429): 191–202. PMCID: PMC1693103; https://doi.org/10.1098/rstb.2002.1190 PMID: 12594927.

171. Hibbett D, Binder M, Bischoff J, Blackwell M, Cannon P, Eriksson O, et al. A higher-level phylogenetic classification of the fungi. Mycolog. Res. 2007; 111(Pt 5): 509–547. https://doi.org/10.1016/j.mycres.2007.03.004 PMID: 17572334.

172. Hoffmann K, Voigt K, Kirk P. *Mortierellomycotina* subphyl. nov., based on multi-gene genealogies. Mycotaxon. 2011; 115(1): 353–363. https://doi.org/10.5248/115.353.

173. Wagner L, Stielow B, Hoffmann K, Petkovits T, Papp T, Vágvölgyi C, et al. A comprehensive molecular phylogeny of the *Mortierellales* (*Mortierellomycotina*) based on nuclear ribosomal DNA. Persoonia. 2013; 30(1): 77–93. PMCID: PMC3734968; https://doi.org/10.3767/003158513X666268 PMID: 24027348.

174. Cheng Y, Edwards J, Allison G, Zhu W, Theodorou M. Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. Bioresour Technol. 2009; 100(20): 4821–4828. https://doi.org/10.1016/j.biortech.2009.04.031 PMID: 19467591.

175. Nakashimada Y, Srinivasan K, Murakami M, Nishio N. Direct conversion of cellulose to methane by anaerobic fungus *Neocallimastix frontalis* and defined methanogens. Biotechnol. Lett. 2000; 22(3): 223–227. https://doi.org/10.1023/A:1005666428494.

176. Lee S, Ha J, Cheng K-J. Relative contributions of *Bacteria*, *Protozoa*, and *Fungi* to in vitro degradation of orchard grass cell walls and their interactions. Appl. Environ. Microbiol. 2000; 66(9): 3807–3813. https://doi.org/10.1128/aem.66.9.3807-3813.2000 PMCID: PMC92224; PMID: 10966394.

177. Leis S, Dresch P, Peintner U, Fliegerovà K, Sandbichler A, Insam H, et al. Finding a robust strain for biomethanation: Anaerobic *Fungi* (*Neocallimastigomycota*) from the Alpine ibex (Capra ibex) and their associated methanogens. Anaerobe. 2014; 29: 34–43. https://doi.org/10.1016/j.anaerobe.2013.12.002 PMID: 24384307.

178. Zhou J, Deng Ye, Zhang P, Xue K, Lianga Y, Van Nostrand J, et al. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proc. Natl. Acad. Sci. USA. 2014; 111(9):E836–E845. https://doi.org/10.1073/pnas.1324044111 PMID: 24550501; PMCID: PMC3948316.

179. Ofițeru I, Lunnc M, Curtisa T, Wellsd G, Criddled C, Francise C, et al. Combined niche and neutral effects in a microbial wastewater treatment community. Proc. Natl. Acad. Sci. USA. 2010; 107(35): 15345–15350. https://doi.org/10.1073/pnas.1000604107 PMID: 20705897

180. Grant T. The Randomized Complete Block Design (RCBD). Department of Horticulture and Crop Science, OARDC, The Ohio State University, USA. 2010.

181. Boaro A, Kim Y, Konopka A, Callister S, Ahring B. Integrated 'Omics' analysis for studying the microbial community response to a pH perturbation of a cellulose-degrading bioreactor culture. FEMS Microbiol. Ecol. 2014; 90(3): 802–815. https://doi.org/10.1111/1574-6941.12435 PMID: 25290699.

182. Menon ArcGIS 10.3: The Next Generation of GIS Is Here. 2014; Accessed June, 2019.

183. Oksenberg L, Cannell C, Kalton G. New strategies for pretesting survey questions. J. Off. Statist. 1991; 7(3): 349–365.

184. Campanaro S, Treu L, Kougias P, Francisci D, Valle G, Angelidaki I. Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. BMC Biotechnol. Biofuel. 2016; 9: 26. https://doi.org/10.1186/s13068-016-0441-1 PMID: 26839589; PMCID: PMC4736482.

185. Caporaso J, Lauber C, Walters W, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME J. 2012; 6(8): 1621–1624. https://doi.org/10.1038/ismej.2012.8 PMID: 22402401

186. Zhou Q, Su X, Ning K. Assessment of quality control approaches for metagenomic data analysis. Scient. Rep. 2014; 4: 6957. https://doi.org/10.1038/srep06957 PMID: 25376098; PMCID: PMC4223665.

187. Bolger A, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. BMC Bioinformatic. 2014; 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404; PMCID: PMC4103590.

188. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. metaSPAdes: A new versatile metagenomic assembler. Genome Res. 2017; 27(5): 824–834. PMCID: PMC5411777; https://doi.org/10.1101/gr.213959.116 PMID: 28298430.

189. Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatic. 2014; 15:182. https://doi.org/10.1186/1471-2105-15-182 PMID: 24925680; PMCID: PMC4074385.

190. Cox M, Peterson D, Biggs P. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatic. 2010; 11(1): 485. https://doi.org/10.1186/1471-2105-11-485 PMID: 20875133; PMCID: PMC2956736.

191. Keegan K, Trimble W, Wilkening J, Wilke A, Harrison T, D'Souza M, et al. A platform-independent method for detecting errors in metagenomic sequencing data: DRISEE. PLoS Comput. Biol. 2012; 8 (6):e1002541. PMCID: PMC3369934; https://doi.org/10.1371/journal.pcbi.1002541 PMID: 22685393.

192. Aronesty E. Comparison of sequencing utility programs. The Open Bioinformatic J. 2013; 7:1–8. https://doi:10.2174/1875036201307010001.

193. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10(3): R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID: 19261174; PMCID: PMC2690996.

194. Wilke A, Harrison T, Wilkening J, Field D, Glass E, Kyrpides N, et al. The M5nr: A novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. BMC Bioinformatic. 2012; 13(1): 141. https://doi.org/10.1186/1471-2105-13-141 PMID: 22720753; PMCID: PMC3410781.

195. Rho M, Tang H, Ye Y. FragGeneScan: Predicting genes in short and error-prone reads. Nucl. Acids Res. 2010; 38(20): e191. https://doi.org/10.1093/nar/gkq747 PMID: 20805240; PMCID: PMC2978382.

196. Kent W. BLAT-the BLAST-like alignment tool. Genome Res. 2002; 12(4): 656–664. https://doi.org/10.1101/gr.229202 PMID: 11932250; PMCID: PMC187518.

197. Pruitt K, Tatusova T, Maglott D. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucl. Acids Res. 2007; 35(Database issue): D61–D65. PMCID: PMC1716718; https://doi.org/10.1093/nar/gkl842 PMID: 17130148.

198. Markowitz V, Ivanova N, Szeto E, Palaniappan K, Chu K, Dalevi D, et al. IMG/M: a data management and analysis system for metagenomes. Nucl. Acids Res. 2008; 36(Database issue): D534–D538. PMCID: PMC2238950; https://doi.org/10.1093/nar/gkm869 PMID: 17932063.

199. O'Donovan C, Martin M, Gattiker A, Gasteiger E, Bairoch A, Apweiler R. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Brief. Bioinformatic. 2002; 3(3):275–84. https://doi.org/10.1093/bib/3.3.275 PMID: 12230036.

200. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucl. Acids Res. 2006; 34(Database issue): D354–D357. PMCID: PMC1347464; https://doi.org/10.1093/nar/gkj102 PMID: 16381885.

201. Benson D, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt K, et al. GenBank. Nucl. Acids Res. 2017; 41(Database issue): D36–D42. PMCID: PMC3531190; https://doi.org/10.1093/nar/gks1195 PMID: 23193287.

202. Bairoch A, Apweilar R. The Swiss-PROT protein sequence database and its supplement TrEMBL in 2000. Nucl. Acids Res. 2000; 28(1): 45–48. PMCID: PMC102476; https://doi.org/10.1093/nar/28.1.45 PMID: 10592178.

203. Snyder E, Kampanya N, Lu J, Nordberg E, Karur H, Shukla M, et al. PATRIC: the VBI PathoSystems Resource Integration Center. Nucleic Acids Res. 2007; 35(Database issue): D401–D406. PMCID: PMC1669763; https://doi.org/10.1093/nar/gkl858 PMID: 17142235.

204. Jensen LJ, Julien P, Kuhn M, von-Mering C, Muller J, Doerks T, et al. eggNOG: automated construction and annotation of orthologous groups of genes. Nucl. Acids Res, 2008; 36(Database issue): D250–D254. https://doi.org/10.1093/nar/gkm796 PMID: 17942413.

205. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucl. Acids Res. 2016; 44(D1): D457–D462. https://doi.org/10.1093/nar/gkv1070 PMID: 26476454; PMCID: PMC4702792.

206. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 2000; 25(1): 25–29. https://doi.org/10.1038/75556 PMCID: PMC3037419; NIHMSID: NIHMS269796; PMID: 10802651.

207. Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, Koonin E, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatic. 2003; 4:41. https://doi.org/10.1186/1471-2105-4-41 PMID: 12969510; PMCID: PMC222959.

208. Cole J, Chai B, Marsh T, Farris R, Wang Q, Kulam S, et al. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucl. acids Res. 2003; 31(1): 442–443. PMCID: PMC165486; https://doi.org/10.1093/nar/gkg039 PMID: 12520046.

209. De Rijk P, Caers A, Van de Peer Y, De Wachter R. Database on the structure of large ribosomal subunit RNA. Nucl. Acids Res. 1998; 26(1): 183–186. PMCID: PMC147202; https://doi.org/10.1093/nar/26.1.183 PMID: 9399830.

210. Van de Peer Y, Robbrecht E, de Hoog S, Caers A, De Rijk P, De Wachter R. Database on the structure of small ribosomal subunit RNA. Nucl. Acids Res. 1999; 27(1):179–83. https://doi.org/10.1093/nar/27.1.179 PMID: 9847173; PMCID: PMC148128.

211. Muller J, Szklarczyk D, Julien P, Letunic L, Roth A, Kuhn M, et al. eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. Nucl. Acids Res. 2010; 38(Database issue): D190–D195. PMCID: PMC2808932; https://doi.org/10.1093/nar/gkp951 PMID: 19900971.

212. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. BMC Bioinformatic. 2012; 28(24): 3211–7. https://doi.org/10.1093/bioinformatics/bts611 PMID: 23071270.

213. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatic Appl. Note. 2006; 22(13):1658–9. https://doi.org/10.1093/bioinformatics/btl158 PMID: 16731699.

214. Field D, Sterk P, Kottmann R, Wim De Smet J, Amaral-Zettler L, Cochrane G, et al. Genomic standards consortium projects. Stand. Genomic Sci. 2014; 9(3): 599–601. https://doi.org/10.4056/sigs.5559680 PMID: 25197446; PMCID: PMC4148985.

215. Glass E, Dribinsky Y, Yilmaz P, Levin H, Van Pelt R, Wendel D, et al. MIxS-BE: A MIxS extension defining a minimum information standard for sequence data from the built environment. The ISME J. 2014; 8(1): 1–3. https://doi.org/10.1038/ismej.2013.176 PMID: 24152717

216. Angiuoli S, Matalka M, Gussman A, Galens K, Vangala M, Riley D, et al. CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. BMC Bioinformatic. 2011; 12(1): 356. https://doi.org/10.1186/1471-2105-12-356 PMID: 21878105; PMC: 3228541.

217. Speed Terry. Statistical analysis of gene expression microarray data. Chapman & Hall/CRC Interdisciplinary Statistics) 1st ed. Kindle edition; 2003.

218. Elmore K, Richman M. Euclidean distance as a similarity metric for principal component analysis. Amer. Meteorol. Soc. 2000; 129: 540–549.

219. Meyer F, Paarmann D, D'Souza M, Olson R1, Glass E, Kubal M, et al. The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatic. 2008; 9: 386. https://doi.org/10.1186/1471-2105-9-386 PMID: 18803844; PMCID: PMC2563014.

220. Ondov B, Bergman N, Phillippy A. Interactive metagenomic visualization in a Web browser. BMC Bioinformatic. 2011; 12(1): 385. https://doi.org/10.1186/1471-2105-12-385 PMID: 21961884; PMCID: PMC3190407.