REVISTA
DESAFIOS
ISSN: 2359-3652
V.13,n.3, maio/2026–DOI:10.20873/vol13n3pibic202529

IN SILICO STRUCTURING OF SUBSTRATE-DRIVEN MICROBIOMES FOR BIOENERGY OPTIMIZATION IN ANAEROBIC DIGESTERS

ESTRUTURAÇÃO IN SILICO DE MICROBIOMAS DIRECIONADOS POR SUBSTRATO PARA OTIMIZAÇÃO DE BIOENERGIA EM BIODIGESTORES ANAERÓBIOS

ESTRUCTURACIÓN IN SILICO DE MICROBIOMAS IMPULSADOS POR SUSTRATO PARA LA OPTIMIZACIÓN DE BIOENERGÍA EN BIODIGESTORES ANAEROBIOS

João Vítor Aires Teixeira

E-mail: joao.aires1@mail.uft.edu.br

Elki Cristina de Souza

E-mail: elkicsouza@mail.uft.edu.br

V.13,n.3, maio/2026–DOI:10.20873/vol13n3pibic202529

ABSTRACT:
The valorization of agro-industrial residues for bioenergy production is a crucial strategy for the circular economy in hubs such as the Brazilian Cerrado. This study performed an in silico meta-analysis of 564 16S rRNA amplicon sequencing samples from 16 studies to decipher the relationship between biomass type and the functional structure of the microbiome in anaerobic digesters, aiming to support the design of optimized systems. Using a robust bioinformatics pipeline, it was demonstrated that substrate type is the main selective factor of the microbial community, with PERMANOVA confirming significant structuring across all beta diversity metrics ($p = 0.001$), particularly via phylogenetic metrics (weighted UniFrac pseudo-F = 10.14). Distinct taxonomic signatures were revealed across substrate categories: *Hydrothermae* dominated algal biomass digesters (11.94%), *MBA03* prevailed in lignocellulosic substrates (14.28%), *Thermovirga* characterized vinasse systems (8.95%), and *Methanosarcina* reached exceptional abundance in food waste (21.76%), compared to a global mean of 4.95%. Despite this high specialization, the fraction of unclassified sequences reached 26.96% globally, pointing to a still largely unresolved microbial diversity in these ecosystems. It is concluded that this computational prospecting provides a predictive roadmap for the engineering of microbial consortia, enabling a transition toward targeted biodigester management. The study provides essential knowledge for the rational design of biodigesters and accelerates the development of sustainable bioenergy solutions for the region.

KEYWORDS: anaerobic digestion; metagenomics; bioinformatics; microbiome; bioenergy.

RESUMO:
A valorização de resíduos agroindustriais para a produção de bioenergia é uma estratégia crucial para a economia circular em polos como o Cerrado brasileiro. Este trabalho realizou uma meta-análise in silico de 564 amostras de sequenciamento de amplicons de RNAr 16S provenientes de 16 estudos, visando decifrar a relação entre o tipo de biomassa e a estrutura funcional do microbioma em digestores anaeróbicos, com o objetivo de subsidiar o design de sistemas otimizados. Utilizando um pipeline de bioinformática robusto, demonstrou-se que o tipo de substrato é o principal fator seletivo da comunidade microbiana, com PERMANOVA confirmando estruturação significativa em todas as métricas de diversidade beta ($p = 0,001$), especialmente via métricas filogenéticas (UniFrac ponderado pseudo-F = 10,14). Assinaturas taxonômicas distintas foram reveladas entre as categorias de substrato: *Hydrothermae* dominou os digestores de biomassa algal (11,94%), *MBA03* prevaleceu em substratos lignocelulósicos (14,28%), *Thermovirga* caracterizou sistemas de vinhaça (8,95%), e *Methanosarcina* atingiu abundância excepcional em resíduos alimentares (21,76%), frente a uma média global de 4,95%. Apesar desta alta especialização, a fração de sequências não classificadas atingiu 26,96% globalmente, apontando para um conjunto microbiano ainda amplamente inexplorado nesses ecossistemas. Conclui-se que esta prospecção computacional oferece um roteiro preditivo para a engenharia de consórcios microbianos, viabilizando a transição para um gerenciamento direcionado de biodigestores. O estudo fornece conhecimento fundamental para o design racional de biodigestores e acelera o desenvolvimento de soluções de bioenergia sustentável para a região.

PALAVRAS-CHAVE: digestão anaeróbia; metagenômica; bioinformática; microbioma; bioenergia.

RESUMEN:
La valorización de residuos agroindustriales para la producción de bioenergía es una estrategia crucial para la economía circular en polos como el Cerrado brasileño. Este estudio realizó un metanálisis in silico de 564 muestras de secuenciación de amplicones de ARNr 16S provenientes de 16 estudios, con el objetivo de descifrar la relación entre el tipo de biomasa y la estructura funcional del microbioma en digestores anaeróbicos, para apoyar el diseño de sistemas optimizados. Mediante un robusto pipeline de bioinformática, se demostró que el tipo de sustrato es el principal factor selectivo de la comunidad microbiana, con PERMANOVA confirmando una estructuración significativa en todas las métricas de diversidad beta ($p = 0{,}001$), especialmente en métricas filogenéticas (UniFrac ponderado pseudo-F = 10,14). Se revelaron firmas taxonómicas distintas entre las categorías de sustrato: *Hydrothermae* dominó los digestores de biomasa algal (11,94%), *MBA03* prevaleció en sustratos lignocelulósicos (14,28%), *Thermovirga* caracterizó los sistemas de vinaza (8,95%), y *Methanosarcina* alcanzó una abundancia excepcional en residuos alimentarios (21,76%), frente a una media global de 4,95%. A pesar de esta alta especialización, la fracción de secuencias no clasificadas alcanzó un 26,96% a nivel global, señalando una materia oscura microbiana aún en gran parte inexplorada en estos ecosistemas. Se concluye que esta prospección computacional ofrece una hoja de ruta predictiva para la ingeniería de consorcios microbianos, permitiendo una transición hacia una gestión dirigida de los biodigestores. El estudio proporciona conocimiento esencial para el diseño racional de biodigestores y acelera el desarrollo de soluciones de bioenergía sostenible para la región.

PALABRAS CLAVE: digestión anaerobia; metagenómica; bioinformática; microbioma; bioenergía.

INTRODUCTION

In a global scenario of energy transition and growing urgency for sustainable waste treatment solutions, the valorization of residual biomass emerges as a fundamental strategy for promoting the circular economy. In particular, in the context of prominent agro-industrial hubs like the Brazilian Cerrado, the vast generation of agribusiness waste represents both an environmental challenge and a strategic opportunity for bioenergy production. The utilization of these liabilities for biogas generation directly aligns with the Sustainable Development Goals (SDG 7 and 12), positioning the region as a potential center of innovation in renewable energies and contributing to local energy security and the mitigation of climate impacts (Barros et al., 2023; Obaideen et al., 2022; Santos et al., 2025).

The optimization of these bioenergy systems relies fundamentally on understanding and managing the complex anaerobic digestion process, which is driven by a highly specialized syntrophic microbial consortium. In this environment, hydrolytic (*Clostridium*, *Ruminococcus*) and acidogenic (*Proteiniphilum*) bacteria convert complex polymers into volatile fatty acids (VFAs) such as propionate and butyrate. The frequently rate-limiting step is the subsequent oxidation of VFAs to acetate, H_2, and CO_2 by specialized acetogenic bacteria (*Syntrophomonas*, *Syntrophobacter*). This reaction is thermodynamically feasible only under continuous H_2 consumption by hydrogenotrophic methanogenic archaea (*Methanoculleus*, *Methanobacterium*). Overcoming these thermodynamic bottlenecks and preventing acidification are crucial steps to maximize energy recovery and ensure stable, high-yield biogas production via the activity of methanogens such as *Methanothrix* and *Methanosarcina* (Fitzgerald et al., 2019; Mutungwazi; Awosusi; Matambo, 2023; Ostos; Flórez-Pardo; Camargo, 2024; Santos et al., 2025).

Furthermore, the efficiency of such advanced systems is intrinsically dependent on the composition of the microbial community, the study of which has been revolutionized by high-throughput sequencing (HTS) of the 16S rRNA gene. The massive application of this technology has consistently corroborated that the chemical composition of the biomass used as substrate or feedstock acts as the primary selective driver of the microbiome (Derakhshani; Tun; Khafipour, 2016; Rahman et al., 2021). The present work aims to establish a comprehensive and detailed panorama of the complex relationships governing biogas production, investigating the interaction between different types of biomass used as substrate and the taxonomic and functional composition of microbial communities. The expectation is that the generated results will provide a solid foundation for the formulation of new hypotheses and for the strategic planning of future applications in precise sustainable energy solutions (De Clercq et al., 2016; Zhang, 2025).

METHODS

Studies were identified through a systematic search of the NCBI Sequence Read Archive (SRA) (available at https://www.ncbi.nlm.nih.gov/sra) for datasets related to terms such as “anaerobic digestion”, “methanogenic bioreactors”, and “biogas”. Datasets generated using any sequencing platform were considered, with no restrictions regarding sequencing depth or targeted 16S rRNA amplicon hypervariable regions sequencing. Studies were included when raw amplicon sequencing data from anaerobic digester systems were publicly available and accompanied by sufficient metadata describing the reactor and feedstock context. Amplicon sequencing data were retrieved using fasterq-dump (SRA Toolkit v3.1.1) (SRA Toolkit v3.1.1) (Leinonen; Sugawara; Shumway, 2010). Sequences were imported into QIIME 2 (Bokulich et al., 2018; Bolyen et al., 2019) as demultiplexed artifacts. Feature/ASV inference was performed using DADA2 algorithm (Callahan et al., 2016).

Taxonomic classification was conducted using a Naive Bayes classifier pre-trained on the SILVA v138 99% database, applying a standard bootstrap confidence threshold of >0.7 (Bokulich et al., 2018; Quast et al., 2012). Sequences assigned as mitochondria or chloroplasts were subsequently removed to eliminate eukaryotic contaminants. A rooted phylogenetic tree was constructed by performing multiple alignment of representative sequences via MAFFT v7.475 (Katoh; Standley, 2013), masking ambiguous regions, and inferring an unrooted tree with FastTree v2.1.10 under the GTR+G substitution model (Price; Dehal; Arkin, 2010), followed by midpoint rooting. Intermicrobial co-occurrence interactions at the genus level were investigated by constructing a correlation network using the SCNIC plugin (Shaffer et al., 2023) within QIIME 2. Pairwise SparCC correlations were calculated with 500 bootstrap iterations to estimate statistical significance. The co-occurrence network was assembled retaining only edges supported by a bootstrapped p-value below 0.05, and co-occurrence modules were subsequently detected using the SCNIC method with a minimum SparCC correlation coefficient $r \geq 0.3$.

To equalize sampling effort and eliminate sequencing depth biases, the feature table was rarefied to a uniform depth of 2,220 reads per sample. This threshold maximized sample retention (n=564 samples from 16 studies), maintaining >95% of the originally included studies. Alpha diversity (Observed Features, Shannon, Pielou's Evenness, Faith's Phylogenetic Diversity) and beta diversity (Bray-Curtis, Jaccard, weighted/unweighted UniFrac) analyses were computed on this rarefied table. Statistical significance was evaluated using PERMANOVA (999 permutations) for beta diversity, and Kruskal-Wallis with Benjamini-Hochberg correction for alpha diversity, followed by Dunn's post-hoc test with Holm correction. Ordination matrices via Principal Coordinate Analysis (PCoA) with 95% confidence ellipsoids and alpha diversity violin plots were generated in the R v4.3.0 environment using phyloseq v1.44.0 (McMurdie; Holmes, 2013) and ggplot2 v3.4.2 (Wickham, 2016).

To mitigate the uneven sample distribution across substrate groups, taxonomic abundances were adjusted using a three-level hierarchical equalized weighting method. Read counts were first converted to relative frequencies via Total Sum Scaling (TSS), then averaged at the study level, and subsequently averaged across categories to generate an unbiased global profile, the equalized mean. Estimate stability was assessed using stratified bootstrap resampling (1,000 iterations, n=12 samples per category) to calculate the Bootstrap Mean and 95% confidence intervals (2.5th and 97.5th percentiles). Finally, to differentiate ubiquitous taxa from substrate-restricted specialists, complementary metrics were computed, including prevalence across categories and samples, conditional abundance and category-normalized abundance.

Row clustering applied Ward’s linkage on Euclidean distances derived from Centered Log-Ratio (CLR) transformed abundances (0.001% pseudocount), equivalent to the Aitchison distance (Aitchison, 1982; Austin; Korem, 2024), while column clustering utilized Bray-Curtis dissimilarity. Genera were mapped against the MiDAS Field Guide for metabolic interpretation (Dueholm et al., 2024). Taxonomic normalizations and visualizations were generated in Python v3.13 (matplotlib v3.8.4, seaborn v0.12.2). All personalized scripts, statistics sheets, and QIIME 2 artifacts are available at https://github.com/jvtarss/biogas-2025.

RESULTS AND DISCUSSION

The meta-analysis compiled 564 samples from 16 studies, grouped into six main biomass categories as presented in Table 1.

Table 1 - Distribution and origin of samples used in the meta-analysis.

Substrate category	Study identification (sample count)	Total samples	Percentage
Manure	(Ahlberg-Eliasson et al., 2021) (129), (Mazzurco Miritana et al., 2020) (35), (Chiariotti; Crisà, 2018) (26), (Shaw et al., 2017) (20), (Li et al., 2022) (6)	216	38.30%
Mixed (Manure + Silage)	(Liu et al., 2017a) (111), (Ahlberg-Eliasson et al., 2018) (56), (Hupfauf et al., 2020) (45)	212	37.60%
Lignocellulosic substrate	(Ge et al., 2022) (30), (Fitzgerald et al., 2019) (18), (Wojcieszak et al., 2017) (12)	60	10.60%
Food waste	(Svensson et al., 2018) (36), (Londono, 2019) (9)	45	8.00%
Algal biomass	(Viruela et al., 2016) (10), (Serna-García et al., 2020) (9)	19	3.40%
Vinasse	(Dias et al., 2016) (7), (Li et al., 2022) (5)	12	2.10%
TOTAL	16 distinct studies	564	100%

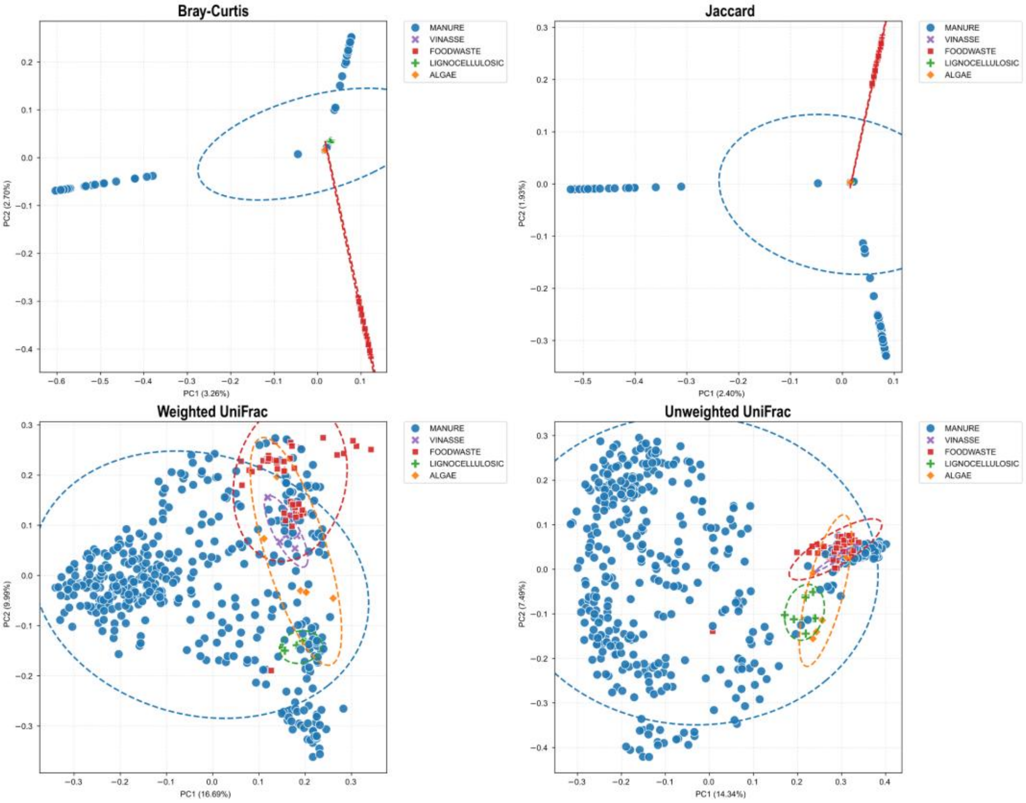
Source: Authors (2026).

The dataset is markedly uneven in its distribution, with manure constituting the largest single category with 216 samples from 5 studies (38.30%), followed by mixed manure-silage substrates with 212 samples from 3 studies (37.60%). Together, these two animal-derived categories account for approximately 76% of the total dataset, reflecting both the practical relevance of livestock waste in industrial biogas production and a probable publication bias, as manure-based systems offer greater operational stability and are more extensively studied. Lignocellulosic substrates represent the third largest category with 60 samples from 3 studies (10.60%), followed by food waste with 45 samples from 2 studies (8.00%), algal biomass with 19 samples from 2 studies (3.40%), and vinasse with 12 samples from 2 studies (2.10%).

The mixed manure-silage category specifically comprises samples from reactors co-digesting cattle or swine manure with grass, clover, or whole-crop barley silages, constituting a compositionally intermediate substrate whose microbial community structure reflects both enteric and lignocellulosic selection pressures simultaneously. The methodological heterogeneity across studies, including differences in amplified 16S rRNA hypervariable regions, sequencing platforms, and bioinformatic pipelines, represents an inherent limitation of any meta-analytical approach, though substrate type consistently emerged as the dominant structuring factor overriding inter-study technical variability (Giuliano et al., 2023; Lin et al., 2018; Šafarič et al., 2018; Wojcieszak et al., 2017).

Figure 1 presents the ordination of microbial beta diversity by Principal Coordinate Analysis (PCoA).

Figure 1 - Ordination of microbial beta diversity by Principal Coordinate Analysis (PCoA) using (a) Jaccard dissimilarity and (b) weighted UniFrac distance methods, alongside PERMANOVA statistical significance values for each method.

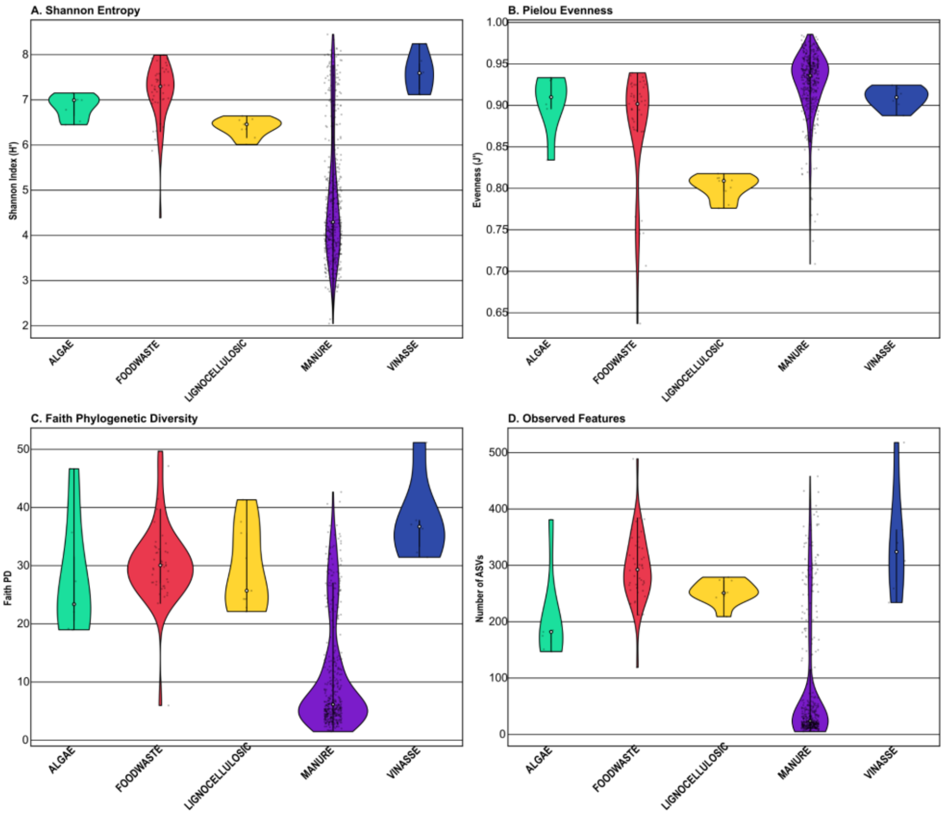


Source: Authors (2026).

Beta diversity analysis (Figure 1) revealed that substrate type is a determining factor in the structuring of anaerobic digester microbial communities, with PERMANOVA confirming statistically significant structuring across all four metrics (p = 0.001). Phylogenetic metrics demonstrated greater discriminatory power than non-phylogenetic ones: weighted UniFrac yielded the highest pseudo-F value (10.14), followed by unweighted UniFrac (pseudo-F = 9.28), Bray-Curtis (pseudo-F = 4.54), and Jaccard (pseudo-F = 3.19). This hierarchy indicates that community differences between substrate categories are expressed more strongly at the level of phylogenetic lineage abundances than at simple compositional presence-absence, suggesting that substrate biochemistry selects not merely for distinct taxa but for the differential amplification of evolutionarily related functional lineages. The variance explained by the first two principal axes was substantially higher for phylogenetic metrics (weighted UniFrac: PC1 = 16.69%, PC2 = 9.99%; unweighted UniFrac: PC1 = 14.34%, PC2 = 7.49%) compared to non-phylogenetic ones (Bray-Curtis: PC1 = 3.26%, PC2 = 2.70%; Jaccard: PC1 = 2.40%, PC2 = 1.93%), reflecting the high compositional dimensionality characteristic of meta-analyses integrating studies from different geographical origins and operating conditions. The more pronounced separation of food waste, lignocellulosic, and algal categories from the main manure cluster in the UniFrac ordinations reflects fundamental differences in substrate biochemical composition, while the substantial overlap between manure and mixed manure-silage co-digestions in ordination space suggests that the silage co-substrate does not drastically alter the phylogenetic backbone of communities where enteric microorganisms adapted to the gastrointestinal tract dominate (Liu et al., 2017), a pattern consistent with the convergence of community structure over time documented in mixed substrate reactors regardless of inoculum origin.

Figure 2 presents violin plots for alpha diversity indices.

Figure 2 - Violin plots for alpha diversity metrics: (a) Shannon entropy, (b) Pielou's evenness, (c) Faith's phylogenetic diversity, and (d) observed ASVs.



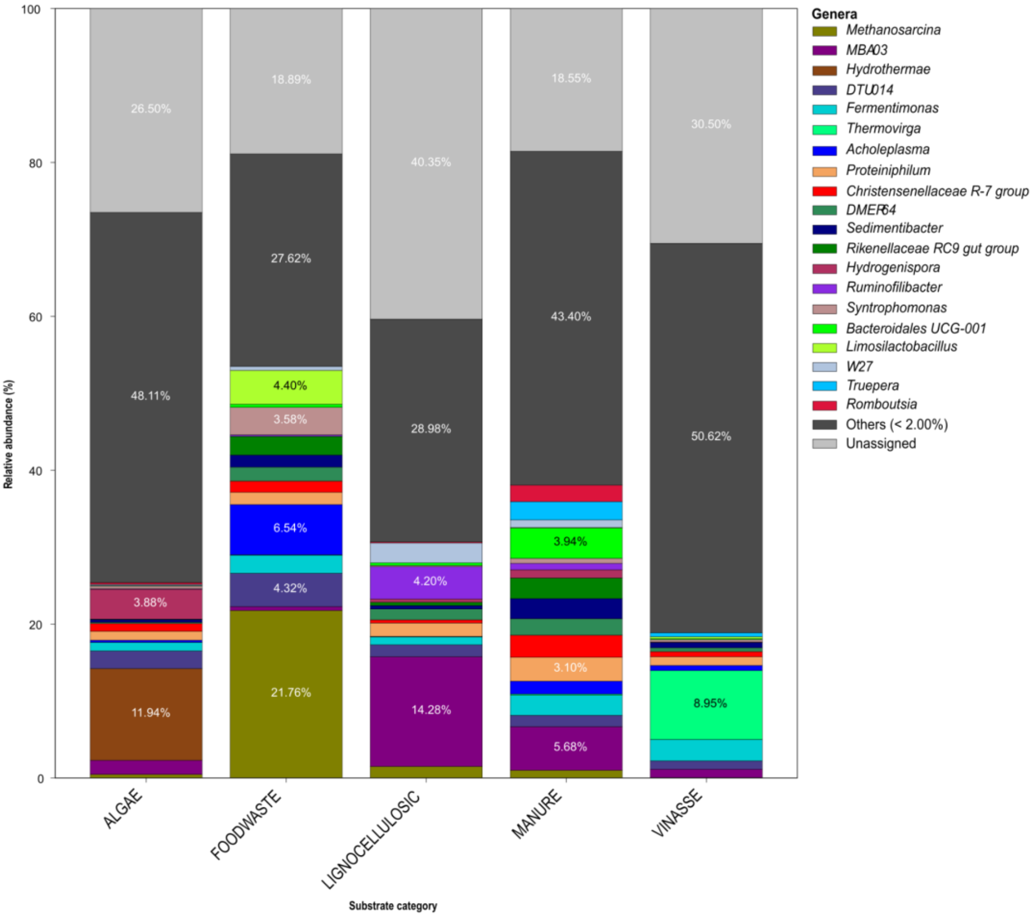
Source: Authors (2026).

Alpha diversity analyses (Figure 2) corroborated the strong substrate-driven selective pressure across all evaluated indices, with Kruskal-Wallis tests yielding H values above 350 and p values below 10^{-68} for Shannon entropy, observed ASVs, and Faith's phylogenetic diversity. The manure category presented the most distinctive pattern, with a bimodal distribution encompassing a low-diversity cluster (median ~30 ASVs, Shannon ~3.5) alongside high-diversity outliers, an internal heterogeneity reflecting the diversity of operating conditions among the five contributing studies, including mono-digestion, poultry co-digestion, and distinct thermal regimes (Ahlberg-Eliasson et al., 2021). This ecological asymmetry between high richness and low evenness in manure is consistent with rumen-inoculated communities harboring high genus richness dominated by few highly abundant taxa (Chiariotti; Crisà, 2018), and with the longer stabilization periods required by low-diversity manure inocula (Liu et al., 2017).

Foodwaste and vinasse samples showed the highest median values for Shannon diversity (~7.5–8.0), ASV richness (~280–320), and Faith's phylogenetic diversity (~30–40), while lignocellulosic substrates exhibited an intermediate profile with narrow distributions, suggesting greater homogeneity in processing conditions across the three included studies. Algal biomass showed diversity values comparable to food waste for phylogenetic diversity and ASV richness but with slightly lower Shannon entropy, likely reflecting the dominance of few abundant taxa such as *Hydrothermae* (11.94%), as discussed in subsequent sections. Pielou's evenness showed comparatively less variation between categories (H = 214.0), with medians between 0.80 and 0.95 across all groups except manure, which spanned the full range from dominance-structured to equitably distributed assemblages (0.65–0.98).

Figure 3 shows the taxonomic composition of microbial communities across the different biomass categories through a stacked bar chart of mean relative abundances, highlighting the dominant taxa associated with each substrate type.

Figure 3 - Stacked bar chart for mean relative abundances of taxa across groups, alongside a listing of genera shared between different biomass groups.



Source: Authors (2026).

Specialized microbiome and dominance of *Hydrothermae* in algae samples

The microbiome associated with microalgal biodigestion presents a markedly distinct taxonomic structure, with *Hydrothermae* emerging as the predominant genus (11.94% normalized abundance), followed by unassigned sequences (26.50%) and *Hydrogenispora* (3.88%). This composition reflects the biochemical specificities of *Scenedesmus* and *Chlorella* cultivated in wastewater effluents through membrane photobioreactors: a recalcitrant cell wall composed of cellulose, hemicellulose, and in some species sporopollenin, alongside high concentrations of polyunsaturated lipids and photosynthetic pigments like chlorophylls and carotenoids, and the absence of the terrestrial lignocellulosic architecture typical of plant-based substrates. The efficient hydrolysis of these structural components requires specialized enzymatic machinery that appears to characterize Hydrothermae and few other taxa adapted to this highly selective niche (Valcarcel; Sanz; Vázquez, 2020). The extremely low prevalence of Hydrothermae (present in only 20% of categories, exclusively in algae samples) combined with its exceptional conditional abundance (11.94% where present) demonstrates strict ecophysiological specialization, confirmed by sensitivity analysis showing that removal of the algae category reduces its global equalized mean to near zero (Gris et al., 2014; Serna-García et al., 2020; Viruela et al., 2016).

Methanogenic dominance and highly efficient syntrophic consortium in food waste samples

Foodwaste samples induce the most asymmetrically methanogenic taxonomic profile among all categories, with *Methanosarcina* reaching an exceptional abundance of 21.76% (4.4 times the global equalized mean of 4.95%). This dominance is directly attributable to the easily biodegradable nature of foodwaste substrates rich in simple carbohydrates, soluble proteins, and emulsified lipids that generate continuous acetate availability, as confirmed by Svensson et al. (2018); the work demonstrated that the frequently fed digester (10 times daily) produced 20% more methane (6.3 L/L/day) than the once-daily fed system, with *Methanosarcina* reaching 9.3 - 9.9% archaeal relative abundance during the first 10 -15 days of operation. Besides, the complementary abundance of *Syntrophomonas* (3.58%, the highest across all categories) sustains efficient beta-oxidation of long-chain fatty acids that accumulated to 7.1 g/kg in the less stably operated daily-fed digester alongside propionic acid peaks of 2,500 mg/L, consistent with beta-oxidation inhibition triggered by acetate accumulation up to 1,500 mg/L and pH fluctuations of 0.50 units within 24-hour feeding intervals.

Metabolic challenges and diverse hydrolytic community in lignocellulosic and mixed substrates

Lignocellulosic substrates encompassing corn stover, whole-crop barley silage, and agricultural residues, represented by 60 samples from 3 studies, impose the greatest biochemical challenge of any category due to the structural recalcitrance of lignin and cellulose crystallinity, reflected in the highest percentage of unassigned sequences (40.35%) and the minimum *Methanosarcina* abundance (1.49%) across all groups, with *MBA03* emerging as the dominant genus (14.28%, approximately 3 times its global average) and suggesting a primary role in complex polysaccharide degradation or syntrophic electron transfer via conductive minerals when H_2 transfer kinetics are limited by diffusion barriers in dense lignocellulosic matrices.

Ge et al. (2022) demonstrated that DLC pretreatment of corn stover with 4% $Ca(OH)_2$ and 2% urea increased cumulative methane yield by 59.27% (224.30 vs. 140.86 mL/g VS untreated) by reducing scum formation, increasing solid substrate utilization, and enhancing dissolved organic biodegradability, which specifically enriched *Ruminofilibacter* (4.20%, 73.33% prevalence) and *Methanobacterium* while establishing a co-occurrence network with more positive interspecies correlations. Such reports indicate that alkaline pretreatment creates physical and chemical conditions favoring robust syntrophic partnerships between aromatic compound-degrading bacteria and hydrogenotrophic methanogens (Li et al., 2019). Liu et al. (2017), using grass-manure mixtures with three different inocula in the mixed manure-silage category, demonstrated that although inoculum source significantly influenced community structure during the first hydraulic retention time, community structures and methane yields converged across reactors over time, confirming that substrate and operating parameters are the primary long-term drivers of community assembly and that cellulose-degrading bacteria originate predominantly from the substrate rather than the inoculum.

Nevertheless, the mixed manure-silage category with 212 samples, composed entirely of co-digestion systems from studies by Liu et al. (2017), Ahlberg-Eliasson et al. (2018), and Hupfauf et al. (2020), occupies an ecologically intermediate position between pure manure and pure lignocellulosic substrates, retaining the gut microbiome legacy of animal-derived inputs while enriching hydrolytic guilds capable of processing structural carbohydrates, consistent with the beta diversity ordination showing near-complete overlap between this category and pure manure in weighted UniFrac space (Kucharska et al., 2018; Martín et al., 2022).

Residual gut microbiome and functional versatility in manure samples

Manure represents the most heterogeneous category numerically, with 216 samples from 5 studies. Its taxonomic profile is dominated by *MBA03* (5.68%) and *Bacteroidales UCG-001* (3.94%), both members of *Bacteroidota* associated with ruminant and swine gastrointestinal tracts, alongside *Christensenellaceae R-7 group* (2.88%), known for its correlation with livestock feed conversion efficiency, confirming the well-documented functional transfer from host gut microbiome to anaerobic digester that was directly evidenced by Chiariotti and Crisà (2018), who showed that buffalo rumen inoculum outperformed sludge inoculum in bio-hydrogen yield (120.8 vs. 65.4 mL H_2/g VS) with *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Tenericutes* accounting for 88.2% of sequences, a compositional profile continuous with the manure category of the present meta-analysis.

Ahlberg-Eliasson et al. (2021) demonstrated that mesophilic-to-thermophilic transitions (37-42°C to 52°C) in cattle manure digesters were feasible with predictable community shifts toward typical thermophilic assemblages, but that co-digestion with nitrogen-rich poultry manure introduced ammonia inhibition risks (0.4 -0.7 g NH_3/L) with fatty acid accumulation (2 g/L) and reduced specific methane production. Meanwhile, Hupfauf et al. (2020) showed that operation at 45°C offered the highest tolerance to organic loading (6 g VS/L/day without VFA accumulation) with the archaeome shifting from acetotrophic dominance at 37°C and 45°C to hydrogenotrophic dominance at 55°C, reflecting a thermodynamically driven reorganization of syntrophic consortia. Mazzurco Miritana et al. (2020) added an important dimension by demonstrating that sulfamethoxazole residues from animal husbandry, routinely present in manure, paradoxically increased cumulative CH_4 production approximately five times relative to controls (2,030.6 vs. 386.8 mL/L) by stimulating H_2 production during acidogenesis. This process shifts methanogenesis from the acetoclastic route (*Methanosarcina* present at 9.3% in controls but absent under antibiotic conditions) toward the hydrogenotrophic route dominated by *Methanobrevibacter* (26%) and *Methanosphaera* (17%), an indication that antibiotic contamination in manure-based digesters can substantially alter methanogenic pathway distribution in ways not captured by purely substrate-driven ecological models (Koniuszewska et al., 2021; Wang et al., 2022).

Extreme environment and thermophilic-acidotolerant specialization in vinasse samples

Vinasse, as a liquid residue of ethanol distillation characterized by extreme physicochemical conditions including pH 3.5 -5.0, high organic load (50 -150 g COD/L), and toxic concentrations of phenolic compounds and melanoidins, selects a rather ecologically distinctive microbiomes in the entire dataset despite comprising only 12 samples from 2 studies (Li et al., 2022; Dias et al., 2016), with *Thermovirga* dominating at 8.95% (vs. 1.84% globally).

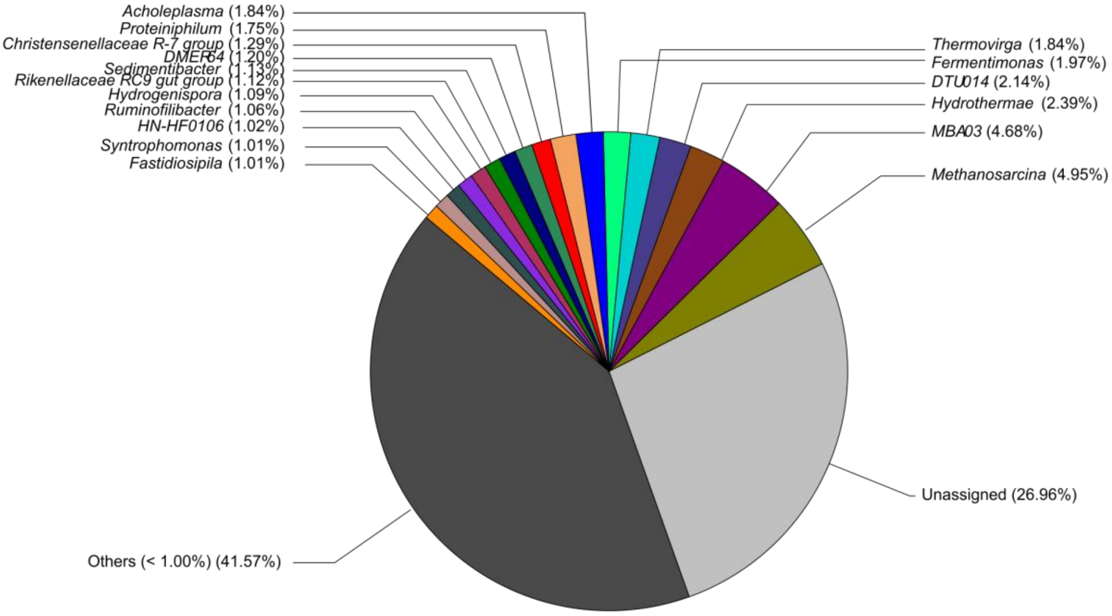
Besides, the sensitivity analysis confirms near-exclusive functional association with this substrate (global abundance collapses to 0.06% upon vinasse removal), reflecting the central role of this strict anaerobic proteinaceous substrate fermenter belonging to *Synergistota* in H_2 and acetate production under thermotolerant and acidotolerant conditions (Dahle; Birkeland, 2006).

This is complemented by *Fermentimonas* (2.78%) and *Proteiniphilum* (1.13%) for residual carbohydrate and protein degradation, *Sphaerochaeta* (1.39%, 100% sample prevalence) for consumption of residual sugars and ethanol compatible with vinasse composition, and *Thiopseudomonas* (1.28%, 83.33% prevalence) for sulfur cycling involving sulfite and sulfate inherent to vinasse, a process confirmed as metabolically central by Dias et al. (2016) through metagenomic analysis of a vinasse biogas desulfurization system. The absence of Methanosarcina from the top-ranked genera and the dominance instead of *Methanotrix* (ultimately mentioned as *Methanosaeta*, its heterotypic synonym, validly published under the ICNP and considered as illegitimate name) (Rosenberg et al., 2014) (0.77%), an oligotrophic aceticlastic archaeon that suggests adaptation to low acetate concentrations and inhibition by phenolic compounds under C:N:P imbalance (C:N ratios frequently exceeding 100:1). This scenario requires precise operational management including nutrient supplementation and rigorous pH control through alkalinization (Borges et al., 2026).

The consolidated genus-level analysis across all 564 samples revealed a global taxonomic profile dominated by the unassigned fraction at 26.958% (bootstrapped 95% CI: 24.532 - 28.912%). Its 100% prevalence and bootstrap mean (26.658%) tracking closely to the equalized estimate indicate this pattern is not driven by any single substrate group. Among classified genera, *Methanosarcina* (4.950%; CI: 1.313 -6.925%) and *MBA03* (4.676%; CI: 3.607 -6.973%) ranked highest. Both exhibited 100% category prevalence with conditional abundances identical to their equalized means, confirming ubiquitous distribution across all substrate types rather than niche concentration. Other top-ranked taxa, including *DTU014* (2.144%, *Firmicutes*), *Fermentimonas* (1.969%), *Thermovirga* (1.841%), *Acholeplasma* (1.837%), and *Proteiniphilum* (1.746%), also displayed 100% prevalence, representing low-to-moderate-abundance constituents of a shared functional core rather than substrate-specific specialists.

Figure 4 illustrates the conditional abundance for genera with an equalized mean above 0.35%, highlighting the disproportionate contribution of substrate-restricted specialists to local community structure relative to their global signal.

Figure 4 - Pie chart of conditional abundance (%) for the main genera identified across all samples, reflecting the mean relative abundance of each genus within the substrate categories where it was detected.

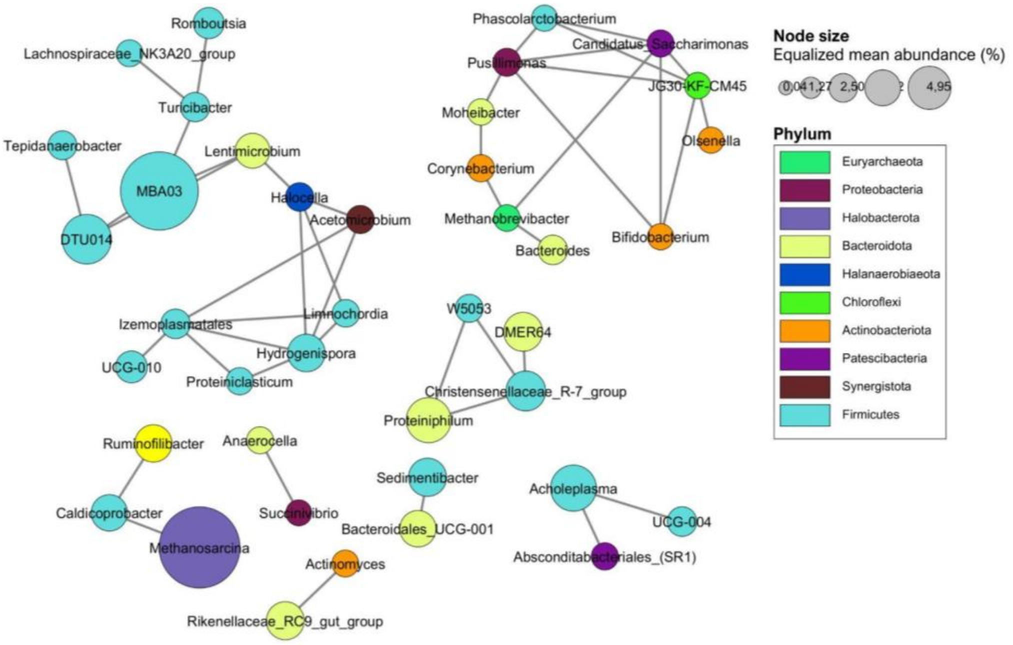


Source: Authors (2026).

The conditional abundance metric corrects for the diluting effect of absence across categories, revealing the true ecological weight of each genus where it occurs. As shown in Figure 4, *Hydrothermae* presents a striking divergence: its equalized mean of 2.387% (CI: 0.000 - 4.378%) masks a conditional abundance of 11.936%, a direct consequence of its detection exclusively in algal biomass (20% prevalence). Conversely, genera with 100% prevalence present identical equalized and conditional values, indicating their global averages reflect local abundance without dilution. Taxa like *Limosilactobacillus* (prevalence 60%), *Truepera* (60%), W27 (80%), Rikenellaceae RC9 gut group (80%), Romboutsia (80%), and *Clostridium sensu stricto 1* (80%) display moderate prevalence gaps, indicating preferential but not exclusive functional roles within specific substrates.

Figure 5 presents the genus-level co-occurrence network constructed from SparCC correlations.

Figure 5 - Genus-level co-occurrence network of anaerobic digester microbiomes. Edges represent statistically significant positive SparCC correlations (bootstrapped $p < 0.05$; $r \geq 0.3$), with edge width proportional to correlation strength. Node size reflects equalized mean abundance; node color indicates phylum-level affiliation.



Source: Authors (2026).

The network topology reveals two structurally distinct sub-graphs. The left sub-graph is anchored by *MBA03* (4.68%, *Firmicutes*) as the highest-degree hub, directly connected to *Lentimicrobium* (*Bacteroidota*), *DTU014* (*Firmicutes*), and bridging toward a secondary cluster formed by *Hydrogenispora*, *Limnochordia*, *Acetomicrobium* (*Synergistota*), and Proteiniclasticum, with *Acetomicrobium* and *Hydrogenispora* sharing one of the strongest edges in the network ($r = 0.438$), suggesting a putative syntrophic partnership consistent with H_2-producing acetate catabolism and interspecies electron transfer in lignocellulosic and mixed substrate environments where these genera show highest prevalence.

A peripheral gut-associated cluster within this sub-graph groups *Turicibacter*, *Romboutsia*, and *Lachnospiraceae NK3A20* group, all *Firmicutes* of gastrointestinal origin, reinforcing the gut microbiome legacy documented in manure-dominated samples (Ćesić et al., 2023; Mutungwazi; Awosusi; Matambo, 2023). *Methanosarcina* (4.95%, *Halobacterota*), the most abundant classified genus globally, is notably peripheral, forming edges only with *Caldicoprobacter* ($r = 0.320$) and *Ruminofilibacter* ($r = 0.338$), indicating that its dominance across substrate categories is functionally independent rather than syntrophically constrained, consistent with its metabolic versatility across acetoclastic, hydrogenotrophic, and methylotrophic methanogenic pathways described in preceding sections. The overall low network density and prevalence of small, loosely connected modules reflect the meta-analytical nature of the dataset, where inter-study variability attenuates correlation magnitudes and ensures that retained edges represent only the most reproducible intermicrobial associations across the full diversity of substrates and operating conditions compiled in this study.

FINAL CONSIDERATIONS

This study demonstrated that substrate biochemistry acts as the primary ecological filter structuring microbial communities in anaerobic digesters, a conclusion supported by statistically significant community differentiation across all four beta diversity metrics (PERMANOVA, $p = 0.001$) and by the distinct taxonomic signatures identified per substrate category. Each biomass type was found to select a characteristic methanogenic strategy, while the universal detection of a conserved low-abundance functional scaffold including especially the *Methanosarcina* genus across categories, suggesting that a shared syntrophic backbone underpins process stability regardless of feedstock.

Specially, translating these findings into practical applications for agro-industrial contexts such as the Brazilian Cerrado requires targeted experimental validation of the patterns identified here, particularly the functional characterization of highly abundant but metabolically unresolved genera such as *MBA03* and *DTU014*, and the empirical confirmation of predicted syntrophic partnerships identified in the co-occurrence network.

Acknowledgements

To the Electrochemistry Laboratory of Federal University of Tocantins (UFT) Gurupi Campus team for their support and collaboration. We also acknowledge the financial support provided by the Tocantins Research Support Foundation (FAPT). In addition, we acknowledge the support of the Google Cloud Research Credits Program, which provided Google Cloud Platform (GCP) computational resources used for the execution of analytical scripts.

REFERENCES

AHLBERG-ELIASSON, K. *et al.* Forage types and origin of manure in codigestion affect methane yield and microbial community structure. **Grass and Forage Science**, v. 73, n. 3, p. 740–757, set. 2018.

AHLBERG-ELIASSON, Karin *et al.* Anaerobic Digestion of Animal Manure and Influence of Organic Loading Rate and Temperature on Process Performance, Microbiology, and Methane Emission From Digestates. **Frontiers in Energy Research**, v. 9, p. 740314, 20 dez. 2021.

AITCHISON, J. The Statistical Analysis of Compositional Data. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 44, n. 2, p. 139–160, 1 jan. 1982.

AUSTIN, George I.; KOREM, Tal. **Compositional transformations can reasonably introduce phenotype-associated values into sparse features**. Bioinformatics, , 21 fev. 2024. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/2024.02.19.581060>. Acesso em: 10 mar. 2026

BARROS, Murillo Vetroni *et al.* Going in Circles: Key Aspects for Circular Economy Contributions to Agro-industrial Cooperatives. **Circular Economy and Sustainability**, v. 3, n. 2, p. 861–880, jun. 2023.

BOKULICH, Nicholas A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. **Microbiome**, v. 6, n. 1, p. 90, dez. 2018.

BOLYEN, Evan *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. **Nature Biotechnology**, v. 37, n. 8, p. 852–857, ago. 2019.

CALLAHAN, Benjamin J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. **Nature Methods**, v. 13, n. 7, p. 581–583, jul. 2016.

ĆESIĆ, Diana *et al.* Association of Gut Lachnospiraceae and Chronic Spontaneous Urticaria. **Life**, v. 13, n. 6, p. 1280, 30 maio 2023.

CHIARIOTTI, Antonella; CRISÀ, Alessandra. Bio-Hydrogen Production From Buffalo Waste With Rumen Inoculum and Metagenomic Characterization of Bacterial and Archaeal Community. **Frontiers in Sustainable Food Systems**, v. 2, p. 13, 2 maio 2018.

DAHLE, Håkon; BIRKELAND, Nils-Kåre. Thermovirga lienii gen. nov., sp. nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. **International Journal of Systematic and Evolutionary Microbiology**, v. 56, n. 7, p. 1539–1545, 1 jul. 2006.

DE CLERCQ, Djavan *et al.* Biomethane production potential from restaurant food waste in megacities and project level-bottlenecks: A case study in Beijing. **Renewable and Sustainable Energy Reviews**, v. 59, p. 1676–1685, jun. 2016.

DERAKHSHANI, Hooman; TUN, Hein Min; KHAFIPOUR, Ehsan. An extended single-index multiplexed 16S rRNA sequencing for microbial community analysis on MiSeq illumina platforms. **Journal of Basic Microbiology**, v. 56, n. 3, p. 321–326, mar. 2016.

DIAS, Marcela França *et al.* Metagenomic analysis of a desulphurisation system used to treat biogas from vinasse methanisation. **Bioresource Technology**, v. 205, p. 58–66, abr. 2016.

DO VALE BORGES, André *et al.* Acidic-alkaline shocks in vinasse fermentation shape methanogenesis and sulfate reduction dynamics. **Science of The Total Environment**, v. 1019, p. 181556, mar. 2026.

DUEHOLM, Morten Kam Dahl *et al.* MiDAS 5: Global diversity of bacteria and archaea in anaerobic digesters. **Nature Communications**, v. 15, n. 1, p. 5361, 25 jun. 2024.

FITZGERALD, Jamie A. *et al.* High-resolution 16S biogas upgrading communities: contrasting in situ and ex situ setups. **Access Microbiology**, v. 1, n. 1A, 1 mar. 2019.

GE, Mianshen *et al.* Densification pretreatment triggers efficient methanogenic performance and robust microbial community during anaerobic digestion of corn stover. **Bioresource Technology**, v. 362, p. 127762, out. 2022.

GIULIANO, A. *et al.* Evaluation of the controlled hydrodynamic cavitation as gas mass transfer system for ex-situ biological hydrogen methanation. **Chemical Engineering Journal**, v. 471, p. 144475, set. 2023.

GRIS, Barbara *et al.* Cultivation of Scenedesmus obliquus in Photobioreactors: Effects of Light Intensities and Light–Dark Cycles on Growth, Productivity, and Biochemical Composition. **Applied Biochemistry and Biotechnology**, v. 172, n. 5, p. 2377–2389, mar. 2014.

HUPFAUF, Sebastian *et al.* Biomethanation at 45 °C offers high process efficiency and supports hygienisation. **Bioresource Technology**, v. 300, p. 122671, mar. 2020.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

KONIUSZEWSKA, Izabela *et al.* The Effect of Antibiotics on Mesophilic Anaerobic Digestion Process of Cattle Manure. **Energies**, v. 14, n. 4, p. 1125, 20 fev. 2021.

KUCHARSKA, Karolina *et al.* Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes. **Molecules**, v. 23, n. 11, p. 2937, 10 nov. 2018.

LEINONEN, R.; SUGAWARA, H.; SHUMWAY, M. The Sequence Read Archive. **Nucleic Acids Research**, v. 39, n. Database, p. D19–D21, nov. 2010.

LI, Naihui *et al.* The role of the biogas slurry microbial communities in suppressing fusarium wilt of cucumber. **Waste Management**, v. 151, p. 142–153, set. 2022.

LI, Zhipeng *et al.* Heterogeneous development of methanogens and the correlation with bacteria in the rumen and cecum of sika deer (Cervus nippon) during early life suggest different ecology relevance. **BMC Microbiology**, v. 19, n. 1, p. 129, dez. 2019.

LIN, Richen *et al.* Graphene Facilitates Biomethane Production from Protein-Derived Glycine in Anaerobic Digestion. **iScience**, v. 10, p. 158–170, dez. 2018.

LIU, Tong *et al.* Importance of inoculum source and initial community structure for biogas production from agricultural substrates. **Bioresource Technology**, v. 245, p. 768–777, dez. 2017a.

LIU, Tong *et al.* Importance of inoculum source and initial community structure for biogas production from agricultural substrates. **Bioresource Technology**, v. 245, p. 768–777, dez. 2017b.

LONDONO, Jorge Enrique Gonzalez. **Characterization of fast anaerobic digestion in a novel reactor design with immobilized biofilms**. Tese de doutorado—Kongens Lyngby, Dinamarca: Technical University of Denmark, 2019.

MARTÍN, Carlos *et al.* Hydrothermal Pretreatment of Lignocellulosic Feedstocks to Facilitate Biochemical Conversion. **Frontiers in Bioengineering and Biotechnology**, v. 10, p. 846592, 16 fev. 2022.

MAZZURCO MIRITANA, Valentina *et al.* Effects of Sulfamethoxazole on the Microbial Community Dynamics During the Anaerobic Digestion Process. **Frontiers in Microbiology**, v. 11, p. 537783, 16 set. 2020.

MCMURDIE, Paul J.; HOLMES, Susan. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **PLoS ONE**, v. 8, n. 4, p. e61217, 22 abr. 2013.

MUTUNGWAZI, Asheal; AWOSUSI, Ayotunde; MATAMBO, Tonderayi S. Comparative functional microbiome profiling of various animal manures during their anaerobic digestion in biogas production processes. **Biomass and Bioenergy**, v. 170, p. 106728, mar. 2023.

OBAIDEEN, Khaled *et al.* Biogas role in achievement of the sustainable development goals: Evaluation, Challenges, and Guidelines. **Journal of the Taiwan Institute of Chemical Engineers**, v. 131, p. 104207, fev. 2022.

OSTOS, Iván; FLÓREZ-PARDO, Luz Marina; CAMARGO, Carolina. A metagenomic approach to demystify the anaerobic digestion black box and achieve higher biogas yield: a review. **Frontiers in Microbiology**, v. 15, p. 1437098, 11 out. 2024.

PRICE, Morgan N.; DEHAL, Paramvir S.; ARKIN, Adam P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. **PLoS ONE**, v. 5, n. 3, p. e9490, 10 mar. 2010.

QUAST, Christian *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. **Nucleic Acids Research**, v. 41, n. D1, p. D590–D596, 27 nov. 2012.

RAHMAN, M. Shaminur *et al.* Microbiome signature and diversity regulates the level of energy production under anaerobic condition. **Scientific Reports**, v. 11, n. 1, p. 19777, 5 out. 2021.

ROSENBERG, Eugene *et al.* (ORGS.). **The Prokaryotes: Other Major Lineages of Bacteria and The Archaea**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

ŠAFARIČ, Luka *et al.* Dynamics of a Perturbed Microbial Community during Thermophilic Anaerobic Digestion of Chemically Defined Soluble Organic Compounds. **Microorganisms**, v. 6, n. 4, p. 105, 11 out. 2018.

SANTOS, Maele *et al.* Biogas, hydrogen, green ammonia and electricity production from canned peach processing residues: Aspects of the circular economy for the Brazilian agroindustry. **International Journal of Hydrogen Energy**, v. 105, p. 45–55, mar. 2025.

SERNA-GARCÍA, R. *et al.* Co-digestion of harvested microalgae and primary sludge in a mesophilic anaerobic membrane bioreactor (AnMBR): Methane potential and microbial diversity. **Bioresource Technology**, v. 298, p. 122521, fev. 2020.

SHAFFER, Michael *et al.* SCNIC: Sparse correlation network investigation for compositional data. **Molecular Ecology Resources**, v. 23, n. 1, p. 312–325, jan. 2023.

SHAW, Grace Tzun-Wen *et al.* Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. **PLOS ONE**, v. 12, n. 7, p. e0181395, 21 jul. 2017.

SVENSSON, Kine *et al.* Feeding frequency influences process performance and microbial community composition in anaerobic digesters treating steam exploded food waste. **Bioresource Technology**, v. 269, p. 276–284, dez. 2018.

VIRUELA, Alexandre *et al.* Water resource recovery by means of microalgae cultivation in outdoor photobioreactors using the effluent from an anaerobic membrane bioreactor fed with pre-treated sewage. **Bioresource Technology**, v. 218, p. 447–454, out. 2016.

WANG, Chengxian *et al.* A review of the effects of antibiotics on the anaerobic digestion of swine waste. **Current Opinion in Environmental Science & Health**, v. 25, p. 100312, fev. 2022.

WICKHAM, Hadley. **ggplot2: Elegant Graphics for Data Analysis**. New York: Springer-Verlag, 2016.

WOJCIESZAK, Martyna *et al.* Adaptation of Methanogenic Inocula to Anaerobic Digestion of Maize Silage. **Frontiers in Microbiology**, v. 8, p. 1881, 28 set. 2017.

ZHANG, Xiaoyu. Highly effective batch effect correction method for RNA-seq count data. **Computational and Structural Biotechnology Journal**, v. 27, p. 58–64, 2025.